ON CROSS-TECHNOLOGY MUTUALISM IN THE INTERNET OF THINGS: COMMUNICATION AND LOCALIZATION

AN ZHENLIN

PhD

The Hong Kong Polytechnic University
2021

The Hong Kong Polytechnic University

Department of Computing

On Cross-Technology Mutualism in the Internet of Things: Communication and Localization

Zhenlin An

A thesis
submitted in partial fulfilment of the requirements
for the degree of

Doctor of Philosophy

April 2021

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____ (Signed)

Zhenlin An (Name of student)

On Cross-Technology Mutualism in the Internet of Things: Communication and Localization

by

Zhenlin An

Abstract

The Internet of Things (IoT) greatly expands the boundaries of the Internet at present. Numerous wireless technologies have been developed to connect everything around us to the Internet for adapting to different connection needs. However, we unilaterally pursue better performance of each technology on its own but ignore the cooperation and symbiosis between technologies for a long time. This situation has led to a growing divide and severe interference between different technologies and wasted the huge potential for cross-technology connection. This dissertation introduces a novel design paradigm for IoT networks, namely, cross-technology mutualism (CTM). CTM coordinates the heterogeneous IoT networks with minimal cost and explores the huge potential of cross-technology connection. The core idea of CTM is to build a direct communication channel among heterogeneous IoT technologies so that IoT devices can exchange information and sense each other through this channel. In this dissertation, we demonstrate four cross-technology communication and localization systems with the CTM design. First, we present TiFi, a system empowering WiFi devices to identify UHF RFIDs. Next, we present Tagcaster, which demonstrates the activation of the wireless voice for current RFID-based ETC systems. RFID, despite being a major enabler for low-power IoT, has long been used only in specific industrial scenarios due to the lack of convenient human-machine interaction interfaces. The two systems dramatically change the way we interact with RFID and make it great again. Second, beyond communication, we also demonstrate how wireless sensing benefits from CTM by creating two cross-technology sensing platforms, namely, UPS+ and iArk. Current high-precision wireless positioning systems are costly and only designed for one specific protocol. UPS+ and iArk are designed to extend the universality of wireless sensing by CTM design. UPS+ revitalizes the ultrasonic positioning systems for ultrasound-incapable smart devices. iArk is the first general-purpose deep tracking platform across protocols for the IoT, and it supports five mainstream types of IoT devices (i.e., NB-IoT, LoRa, RFID, Sigfox, and Zigbee) and is scalable to other types with minimal effort.

Acknowledgments

My graduate journey at PolyU has been a wonderful and enriching time of my life. First, I would like to thank my advisor, Dr. Lei Yang, who has greatly supported me in study and life. Dr. Yang has a great passion for research and can always come up with interesting ideas. He works incredibly hard for his students and always gives much concrete help. I am very fortunate to have had such a wonderful supervisor guiding me through my graduate career.

I have also had an amazing experience working with Dr. Qiongzheng Lin and Dr. Ping Li throughout my Ph.D. Dr. Lin has very keen observation and quick thinking on the system design. Our daily discussions always enlighten me considerably. I had a memorable experience working with him on multiple challenging systems. Dr. Li has extensive knowledge and a deep understanding of the wireless area. Interacting with him is a truly enjoyable experience.

I would also like to thank my collaborators and co-authors. I appreciate the efforts and suggestions from Prof. Lei Xie, Dr. Chunhui Duan, and Dr. Wei Lou. I also appreciate the help from Dr. Yanni Yang, who helps me a lot with experiments and gives me many constructive suggestions. I also appreciate the lessons from the teachers in our department. They are Prof. Jiannong Cao (my co-supervisor), Dr. Yuanqing Zheng, Dr. Jing Li, Prof. Lei Zhang, Dr. Xiaoming Wu and many others.

I am also thankful to Mr. Fang Wang, Dr. Zhuo Li, Mr. Zhixuan Liang, Dr. Shan Jiang, Mr. Mingjin Zhang, Miss Ningning Hou, Miss Caiyan Cui, Mr. Jinlin Chen, Mr. Zexin Lu, Mr. Hanzhuo Tan, Mr. Xindong Zhang, Dr. Wei Zhang, Mr. Yu Yang, Mr. Ruosong Yang, Mr. Zhiyuan Wen, Dr. Jia Wang and many others for their companion and support. We have great office time. Hong Kong and PolyU went through a very tough time in 2019 and 2020. Your companion makes the difficult time less stressful.

During my graduate time, my roommates also give me much support. They are Dr. Zhewei Liu, Mr. Bo Wu, Mr. Boyang Zhang, Mr. Yeliang Shi, Mr. Tuan

Liu, Mr. Zhe Pan, Mr. Zirong Zhou and Mr. Xiaolei Yin. Our dormitory times are unforgettable.

During my graduate time, I made a lot of friends in Hong Kong. I appreciate them for sharing my happiness and sorrows. They are Mr. Zecheng Li, Dr. Lida Li, Dr. Siqi Ding, Mr. Yunwen Yu, Mr. Ziyang Han, Miss Haiming Cheng and many others.

Special thanks go to my girlfriend, Miss Man Guo, for all her love and support.

Finally, I would like to express my deepest gratitude to my parents. They did the best they could to raise me. I love them so much.

Previously Published Materials[1]

Conference

1. **Zhenlin An**, Qiongzheng Lin, Lei Yang, "Cross-Frequency Communication: Near-Field Identification of UHF RFIDs with WiFi!", in Proc. of **ACM MobiCom**, 2018. (Presented in Chpater 4)
2. **Zhenlin An**, Qiongzheng Lin, Qingrui Pan, Lei Yang, "Activating Wireless Voice for E-Toll Collection Systems with Zero Start-up Cost", in Proc. of **IEEE INFOCOM**, 2020. (Presented in Chapter 5)
3. Qiongzheng Lin†, **Zhenlin An**†, Lei Yang. "Rebooting Ultrasonic Positioning Systems for Ultrasound-incapable Smart Devices", in Proc. of **ACM MobiCom**, 2019. (Presented in Chapter 6)
4. **Zhenlin An**†, Qiongzheng Lin†, Ping Li, Lei Yang. "General-purpose deep tracking platform across protocols for the internet of things", in Proc. of **ACM MobiSys**, 2020. (Presented in Chapter 7)
5. **Zhenlin An**, Qiongzheng Lin, Xiaopeng Zhao, Lei Yang, Dongjiang Zheng, Guiqing Wu, Shan Chang. "One Tag, Two Codes: Identifying Optical Barcodes with NFC", in Proc. of **ACM MobiCom**, 2021.
6. **Zhenlin An**, Qiongzheng Lin, Qingrui Pan, Lei Yang. "Turbocharging Deep Backscatter Through Constructive Power Surges with a Single RF Source", in Proc. of **IEEE INFOCOM**, 2021.
7. **Zhenlin An**, Qiongzheng Lin, Lei Yang, Wei Lou, "Embracing Tag Collisions: Acquiring Bloom Filters across RFIDs in Physical Layer", in Proc. of **IEEE INFOCOM**, 2019.
8. Qiongzheng Lin, Lei Yang, **Zhenlin An**, Yi Guo, Ping Li, "RFCamera: Identifying RFIDs in Pixel Dimensions", in Proc. of **IEEE SECON**, 2020. [**Best Paper Award**]

[1] †: co-primary author

9. Ping Li, **Zhenlin An**, Lei Yang, Panlong Yang, "Towards Physical-Layer Vibration Sensing with RFIDs", in Proc. of **IEEE INFOCOM**, 2019.
10. Lei Yang, Qiongzheng Lin, Chunhui Duan, **Zhenlin An**, "Analog On-tag Hashing: Towards Selective Reading as Hash Primitives in Gen2 RFID Systems", in Proc. of **ACM MobiCom**, 2017.
11. **Zhenlin An**, Lei Yang, Qiongzheng Lin, "Demo: Activating Wireless Voice for E-Toll Collection Systems with Zero Start-up Cost", in Proc. of **ACM MobiCom**, 2019.
12. **Zhenlin An**, Qiongzheng Lin, Lei Yang, "Demo: Near-Field Identification of UHF RFIDs with WiFi!", in Proc. of **ACM MobiCom**, 2018. [**Best Demo Award Runner-up**]

Journal

1. **Zhenlin An**, Lei Yang ,Qiongzheng Lin, "Identifying UHF RFIDs in Range of Readers with WiFi", in IEEE/ACM Transactions on Networking (**IEEE/ACM TON**), 2021. (Presented in Chapter 4)
2. **Zhenlin An**, Qiongzheng Lin, Lei Yang, Yi Guo, "Revitalizing Ultrasonic Positioning Systems for Ultrasound-Incapable Smart Devices," in IEEE Transactions on Mobile Computing (**IEEE TMC**), 2020. (Presented in Chapter 6)
3. **Zhenlin An**, Qiongzheng Lin, Lei Yang, Wei Lou, Lei Xie, "Acquiring Bloom Filters Across Commercial RFIDs in Physical Layer," in IEEE/ACM Transactions on Networking (**IEEE/ACM TON**), 2020.
4. Ping Li, **Zhenlin An**, Lei Yang, Panlong Yang, "Towards Physical-Layer Vibration Sensing with RFIDs," in IEEE Transactions on Mobile Computing (**IEEE TMC**), 2019.
5. Qiongzheng Lin, Lei Yang, Chunhui Duan, **Zhenlin An**, "Tash: Towards Selective Reading as Hash Primitives for Gen2 RFID", ACM/IEEE Transactions on Networking (**IEEE/ACM TON**), 2019.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

"Life did not take over the globe by combat, but by networking."

—Lynn Margulis

The Internet-of-Thing (IoT) is becoming a new global infrastructure that enables the future hyper-connected world where everything is wirelessly connected as a whole. We regard IoT as the leading force for the next-generation technology revolution. The envisioned set of IoT applications includes but is not limited to smart home, wearable technology, the Internet of medical things, vehicle-to-everything communication, and the Industrial IoT. IoT is becoming one of the most ambitious, huge, and complicated ecosystems ever created by humans. The process of building such an IoT ecosystem is proceeding rapidly at present. According to Cisco Internet Report in 2020 [3], a crazy variety of 8.8 billion wireless connected smart devices are accessing mobile networks at this moment, and 13.1 billion global mobile devices and connections will be available by 2023. They are similar to the rapidly growing biota that makes up the IoT ecosystem together.

Diversified wireless networks, as the nervous systems of the IoT world, lay a solid foundation for the IoT ecosystem boom. They are like various axons connecting the smart IoT devices to an intelligent organism. Today, multiple parallel advances in different wireless technologies are building their own nervous systems for specific IoT applications. For example, we deploy RFID tags for

automating supply chain, logistics, and retails [4]. We also build out a smart home by using WiFi and ZigBee to control household appliances wirelessly [5]. In smart agriculture, we create large-scale wireless environmental monitoring networks and remote controlling systems based on LoRa technology [6]. The IoT market is not just one product like a smartphone but hundreds of different products. We are trying our best to create an increasing number of customized networks to meet the increasingly diverse IoT demands.

The great diversity of wireless technologies comes with the high fragmentation of the IoT ecosystem. IoT networks are incompatible with each other. On the one hand, each IoT technical alliance attempts to form a unique wireless network protocol stack to cater to the wireless connectivity needs of specific scenarios. On the other hand, they use technical barriers to build economic moats over their competitors to make profits. As a result, each protocol shapes a closed network. Any two devices with different protocols cannot communicate with each other directly. Therefore, each IoT species lives alone in its territory. For example, the patented LoRa protocol is a proprietary long-range wireless technology owned by Semtech company [7]. Everyone who wants to implement their own LoRa hardware has to obtain the license from Semtech and follow the closed LoRa standard. Directly adjusting the LoRa device to make it compatible with the other wireless technologies is nearly impossible due to such a closed development pattern. Such a divide causes difficulty combining various IoT networks into an intelligent ecosystem.

The wireless environment also becomes increasingly chaotic as wireless technologies diversify and wireless devices increase. When wireless devices work in the same area, they fight for scarce wireless spectrum resources. The situation becomes worse for the IoT ecosystem because most IoT devices work on the narrow ISM frequency band. Different IoT devices cannot communicate with each other and coordinate the spectrum resources because of the lack of direct cross-technology communication (CTC) channel. As a result, strong cross-technology interferences (CTIs) occur in heterogeneous IoT networks [8]. Wireless technolo-

gies become victims of their own success.

To coordinate heterogeneous IoT networks, we built powerful all-in-one intelligent devices, such as smartphones and multifunctional gateways [9], for a long time. As a result, smart devices at present contain excessive wireless transceivers and sensors for connecting to the complex IoT world. However, such a brutal solution suffers from the following several drawbacks. First, wireless transceivers and sensors cause additional space, power, hardware, and deployment costs. Such cost is especially unacceptable for portable IoT devices with strict limitations on internal space and battery capacity. For example, smartphone at present is powerful beyond our minds, but it cannot read UHF RFID tags. This reason is that the UHF RFID reader is bulky and has very high power dissipation [10]. Therefore, we cannot integrate a UHF RFID module into a smartphone. Second, forwarding with the gateway forgoes device-to-device (D2D) communication and causes high communication latency. D2D communication is the foundation capacity of the IoT. It is the key enabler for swarm intelligence, collaborative automation [11], and wireless sensing [12]. The IoT ecosystem is less intelligent, incomplete, and defective without D2D communication.

The profound reason for the abovementioned issues is that we are developing the IoT ecosystem fragmentedly. Each technology is confined to its own small network and application scenarios rather than growing with the others. This isolated development places the IoT ecosystem under difficult circumstances. Therefore, it behooves us to design the development path of each technology from the perspective of the larger IoT ecosystem. Everyone is born to work with others and form a harmonic ecosystem together.

1.1 Cross-Technology Mutualism

We found our design inspiration in the natural ecosystem. In a natural ecosystem, symbiosis describes several types of living arrangements between different species of organisms. Symbiosis has three major categories: including commensalism,

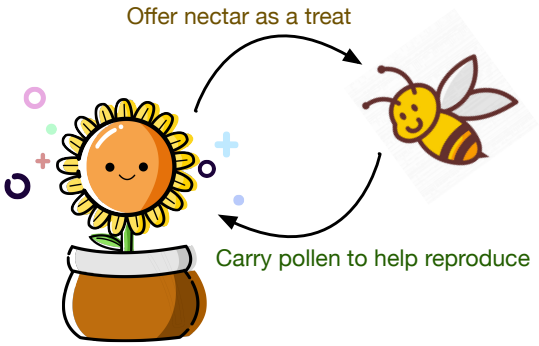


Figure 1-1: Mutualism between Bees and Flowers.

mutualism, and parasitism. Particularly, mutualism is the beneficial interaction between two or more species where each species has a net benefit [13]. Fig. 1-1 shows a classic mutualism example, where bees fly from flower to flower, to gather nectar, which is the food bees depend on. When they land in flowers, the bees get some pollen on their hairy bodies, and the pollen will be carried to other flowers to pollinate the plant when the bees fall on them. Thus, they unknowingly benefit each other during the contract. Another classic example is the relationship between termites and the protists that live in their gut. The protists digest the cellulose contained in the wood to release nutrients for the benefit of the termite. In turn, the protists receive a steady supply of food and live in a protected environment. The protists also have a symbiotic relationship with the bacteria that live in their gut, without which they could not digest cellulose. According to incomplete statistics, more than 48% of land plants on the Earth rely on the mutualism between mycorrhiza and fungi to provide them with inorganic compounds and trace elements; the estimate of tropical forest trees with seed dispersal mutualism with animals ranges from 70 to 80%. Clearly, mutualism plays a key part in ecology.

In this dissertation, we introduce mutualism into the artificial IoT ecosystem. The mutualism in the IoT refers to the interactions between heterogeneous IoT networks where two or more wireless technologies exchange information freely (i.e., without the external helper), benefit from each other (like bees and flowers), collaborate, and further serve single entire intelligence. This paradigm is

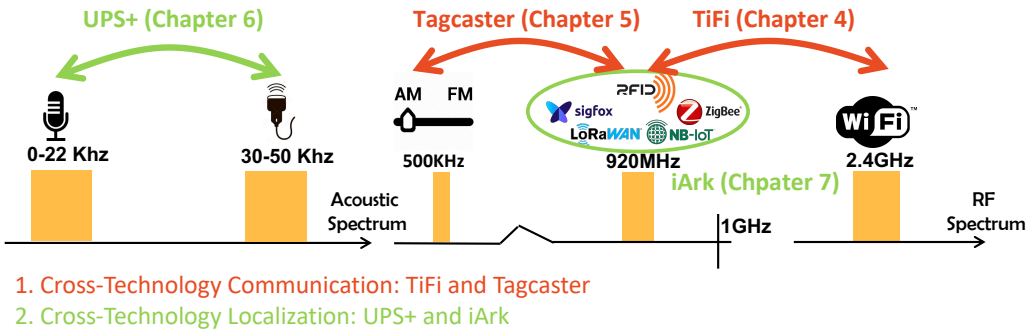


Figure 1-2: Thesis structure

what we called *cross-technology mutualism* (CTM). Specifically, CTM is a novel minimal design paradigm for IoT networks, which builds a cross-technology interaction bridge by extending the existing ability of wireless infrastructures instead of adding new hardware. Any two IoT devices with different protocols can exchange data, sense each other and cooperate with the help of a cross-technology interaction bridge. For example, if we can build such a bridge between RFID and WiFi, then RFID can be read by numerous smart IoT devices with WiFi functions and become simpler to use. From WiFi's perspective, the ability to read RFID tags opens the door to the wide backscatter world. If we enable the CTM between any two heterogeneous IoT networks, then the whole IoT ecosystem can operate together as a hyper-connected intelligent organism. Such harmonic hyper-connectivity is a key to the next generation wireless communication (6G network) [14], which has significant value for IoT network development.

The structure of the rest dissertation is shown in Fig. 1-2. We will describe four cross-technology cooperative systems based on the CTM philosophy to advance two broad goals: (1) creating new connections between heterogeneous networks for cross-technology D2D communication (Chapters 4 and 5) and (2) achieving novel cross-technology localization (Chapters 6 and 7).

1.2 CTM: Novel Communication

The first part of this dissertation introduces how CTM creates new connections between heterogeneous networks and makes them complement each other. Our

work focuses on reviving the UHF RFID technology by CTM design. We first attempt to improve its usability by achieving CTM between UHF RFID and WiFi. The UHF RFID is an automatic identification and data capture technology and was once considered a barcode replacement. However, the use of RFID remains limited to a small number of industrial areas (e.g., logistics, warehouses, and hospitals, etc) to date compared with that of the barcode. RFIDs are not widely accepted in the consumer-oriented market mainly because this technology is not supported by currently available personal mobile devices, unlike the barcode, which can be recognized directly by built-in cameras. Consumers have to use special-purpose RFID readers to query tags, and thus, they cannot benefit from the convenience of their mobile devices. The industry has exerted considerable effort to bridge this gap. For example, ImpinJ Inc. [15] released a special RFID holder into which the user can insert his or her smartphone for RFID scanning. Alien Technology integrates WiFi and (or) Bluetooth modules into readers to temporarily connect with smartphones. These trade-off solutions (additional details are provided in Chapter. 3) aim to promote the integration of RFID technology into smart devices. However, they have achieved minimal progress a few years after their introduction because they either introduce extra hardware costs or increase the deployment complexity. We first present a cross-technology communication technology TiFi to identify UHF RFID tags with WiFi in Chapter 4. If we can enable RF receivers (i.e., WiFi) of smart phones to access RFID tags, then the range of RFID applications will be extremely extended. Accordingly, everyone can enjoy the convenience brought by RFIDs with his or her phones by managing and tracking personal belongings with tiny RFID tags. In this case, the new connection between RFID and mobile devices created by CTM can extend the boundary of RFID and reborn it.

We next guarantee user-friendliness of UHF RFID by achieving CTM between UHF RFID and AM radio. This time, we shifted our attention from the retails to Electronic Toll Collection (ETC). ETC enables the electronic collection of toll payments to allow for near-nonstop toll collection and traffic monitoring. UHF

RFID is the core for the automation in ETC service. However, ETC systems, as a typical technology of the IoT, are designed for machine-to-machine communication only (e.g., identifying cars or monitoring traffic). When drivers pass the ETC station, they cannot directly obtain related information (e.g., charge amount, credit balance, real-time traffic, and road condition). Our idea is to let an AM radio hear the RFID reader (ETC) message directly. Then, AM radio becomes a convenient and user-friendly machine-to-human communication interface to provide informative interaction in ETC stations. We name this technology Tagcaster and give a detailed description of it in Chapter 5. In the two cases, the new connections between RFID and other wireless technologies created by CTM extend the boundary of RFID and reborn it. Notably, the new connections created by CTM are not limited to the RFID. It is conceivable that we can build up more cross-technology connections in other areas, and many more surprising applications will be enabled.

The above two CTM systems have two fundamental challenges: cross-frequency communication (CFC) and cross-protocol communication (CPC). First, we have to break the huge frequency gap between technologies. The UHF RFID system operates at 800-900 MHz, whereas a typical WiFi device works at 2.4 GHz ISM band and an AM radio works at medium frequency (e.g., 500-1700 kHz). We design a series of novel frequency shift mechanisms based on the nonlinearity effects in the RF frontend to build the CFC channels between them. Second, the three IoT technologies use different protocols, including modulation, data encoding scheme, and packet structure. They are similar to three people using three different languages and cannot understand each other. We present software-only emulating technologies to generate packets that can be recognizable by the others for solving the aforementioned problem. Our solutions have been implemented and evaluated in practical systems to demonstrate the actual CTM benefits.

Contributions. Our work makes the following key contributions:

■ *Identifying UHF RFIDs with WiFi*: TiFi (i.e., Tag emulated WiFi) [MOBICOM '18 and TON '20] is the first cross-technology communication system that

allows a 2.4 GHz WiFi receiver (e.g., a mobile phone) to identify UHF RFID tags. TiFi does not require changing current smartphones or tags. Its core has a series of novel signal encoding and modulation techniques that explore the harmonic backscatter as a side channel and use it to communicate with WiFi receivers. We design and implement TiFi with commodity WiFi chipsets (e.g., Broadcom BCM43xx, Murata KM6D28040, and Qualcomm WCN3990). Our comprehensive evaluation shows that TiFi allows WiFi receivers to identify UHF RFID tags within the range of 2 m. TiFi demonstrates that WiFi can extend its reach from high-speed wireless access to a low-power IoT network by CTM design.

■ *Activating Wireless Voice of ETC Systems with Zero Start-up Cost*: Tagcaster [INFOCOM '19] is another cross-technology communication system that first shows that an AM radio receiver can receive UHF RFID signals due to the presence of the nonlinearity effect in the AM receiver. On the basis of this phenomenon, Tagcaster converts an ETC reader to an AM station for broadcasting short messages (e.g., charged fees and traffic forecast) to drivers at toll booths. The key innovation in this work is the engineering of Tagcaster over off-the-shelf ETC systems using shadow carrier and baseband whitening without the need for hardware nor firmware changes. This feature allows zero-cost rapid deployment in the existing ETC infrastructure. Two prototypes of Tagcaster are designed, implemented, and evaluated over four general and five vehicle-mounted AM receivers (e.g., Toyota, Audi, and Jetta). Experiments reveal that Tagcaster can provide good-quality (PESQ$>$ 2) and stable AM broadcasting service with a coverage range of 30 m.

1.3 CTM: Novel Localization

CTM not only can build new connections in the networks but also can bring a thorough revolution to indoor localization technology. Indoor localization is a state-of-the-art technology used to locate people or objects where satellite technologies (i.e., GPS, Beidou) lack precision or fail entirely. It prompts a series of

key mobile applications: indoor navigation (e.g., malls, factories, and airports), augmented reality, location-aware pervasive computing, advertising, and social networking. We can live more comfortably, securely, and more sustainably because of wireless localization systems.

Unfortunately, a high-accurate wireless localization platform is still far from widespread adoption at present. The primary reason is that some localization systems require specific receivers rather than ordinary receivers. One way CTM can improve this situation is to make the existing localization infrastructure compatible with ordinary devices at a minimal cost. For example, the ultrasound positioning system (UPS), which utilizes ultrasonic sound as the ranging media, is highly accurate and economical and is a mature wireless sensing technology. Decades ago, MIT developed a pioneering UPS named Cricket [16] which can already provide centimeter-level accuracy location service. However, UPSs do not receive significant attention at present because they suffer from a severe defect, that is, current smart devices lack ultrasonic sensors and therefore cannot receive ultrasonic beacon signals from UPSs [17–19]. In this case, if we can promote traditional UPSs to become audible to ultrasound-incapable receivers, innumerable smart devices with basic microphone can benefit the UPS. Then, the huge potential of UPS can instantly be unleashed. UPS+ in Chapter 6 introduce the way we achieve this with CTM paradigm. This example shows how CTM renews wireless positioning by cross-technology cooperation. The new connections between heterogeneous IoT networks created by CTM empower cross-technology localization. If we can enable different IoT devices to localize each other, then not only the cost of localization will be reduced, but also the accuracy of localization will be improved.

Another way CTM can help the localization system is to build super-adaptive general-purpose positioning platforms. Current localization platform can only serve one specific wireless protocol device rather than work similar to a general-purpose base station serving various mobile devices simultaneously. This inadequacy drastically limits the application scenarios of wireless positioning tech-

nology. Hundreds of IoT products are around us to date. Building a unique localization system for every technology is nearly impossible. The deep reason is that we lock our minds on the specific protocol and design the localization system case by case. However, although every protocol uses a unique wireless signal (i.e., modulation and bandwidth), our localization theory and rationale, such as Arrival-of-Angle and Time-of-Flight, is generalized. A general-purpose localization platform, such as a base station serving many IoT products, can be built theoretically. If we can realize this possibility, then the development of indoor localization system will be significantly facilitated. For example, one CTM positioning platform that can serve various devices reduces system management and operation costs. This capability in turn will lower the threshold for the use of the positioning system and further expand its application range. We build such a general-purpose localization platform, namely, iArk, and demonstrate its benefits in Chapter 7. The abovementioned discussions only uncover the tip of the iceberg. The huge potential of CTM still needs to be further developed.

Contributions. We made our efforts in the novel generalized localization and our work makes the following key contributions:

■ *Revitalizing Ultrasonic Positioning Systems for Ultrasound-Incapable Smart Devices*: UPS+ [MOBICOM '19 and TMC '20] is the first advanced ultrasound positioning system to be compatible with ultrasound-incapable smart devices. The core concept is to deploy two types of indoor beacon devices, which will advertise ultrasonic beacons at two different ultrasonic frequencies. Their superimposed beacons are downconverted to a low-frequency by exploiting the nonlinearity effect at the microphone of the receiver. This underlying property functions as an implicit ultrasonic downconverter without inflicting harm to the hearing system of humans. We demonstrate UPS+, which is a fully functional UPS prototype, with centimeter-level localization accuracy by using custom-made beacon hardware and well-designed algorithms.

■ *General-Purpose Deep Tracking Platform across Protocols for the Internet of Things*: iArk [MOBISYS '20] is a general-purpose tracking platform for all types

of IoT devices regardless of modulation approaches and wireless protocols. The core of iArk is a novel protocol-free estimation channel algorithm. By the virtue of decoupling from wireless protocols, iArk also allows researchers to concentrate on developing a new tracking algorithm without considering the protocol diversity. To date, the platform can support five mainstream types of IoT devices (i.e., NB-IoT, LoRa, RFID, Sigfox, and Zigbee) and is scalable to other types with minimal effort.

1.4 Organization

The rest of the dissertation is organized as follows. Chapter 2 reviews the evolution of CTM. Chapter 3 presents the fundamentals of cross-frequency communication. Chapter 4 describes TiFi in greater detail and how it empowers WiFi devices to inventory the UHF RFID tags. In Chapter 5, we describe Tagcaster, which is the first CTC technology that bridges the AM radio and UHF RFID system. The two chapters present the creation of a CTM connection and the benefits of CTM on today's networks and real applications. Chapter 6 introduces UPS+, which is the first cross-frequency localization system to revitalize ultrasonic positioning systems for ultrasound-incapable smart devices. In Chapter 7, we describe a general-purpose tracking platform iArk working across protocols. Finally, chapter 8 concludes the whole dissertation and envisions the future work.

Chapter 2

Evolution of CTM

The idea of IoT CTM does not happen overnight but instead goes through a gradual yet fundamental shift. In this chapter, we review the evolution of CTM. The development is composed of four phases, as shown in Fig. 2-1. In the early stage of IoT development, we focus on eliminating *cross-technology interference* (CTI) and achieving cross-technology coexistence. As is well known, the interference among *homogeneous devices* from the same wireless technology can be easily avoided because they could understand their own packages. However, the wireless interference among *heterogeneous devices*, which communicate with different wireless protocols in the same frequency band, is much more severe and disturbing because they are totally blind to each other due to the impassable physical channels. Therefore, the concept of *wireless coexistence* is proposed to tentatively coordinate heterogeneous devices (e.g., WiFi and Zigbee) for reducing extravagant collisions to address the aforementioned issue. In the second phase, the wireless coexistence is further extended to the *in-band cross-technology communication* (CTC), which allows direct communications among heterogeneous devices. The in-band interference is considered an opportunity to exchange information. Then, the low-level coexistence is promoted to the high-level cooperation among the heterogeneous devices.

The start-up of CTM emerges in the third stage. We are still in this phase. Meanwhile, we attempt to further extend the collaboration to the super-heterogeneous

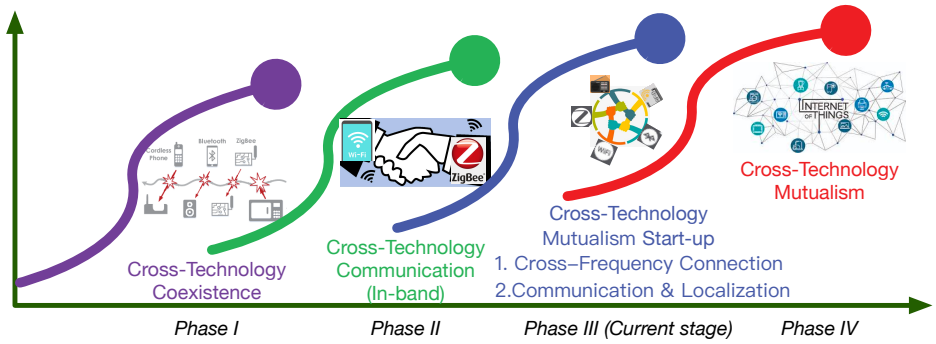


Figure 2-1: Evolution of IoT CTM.

devices, which even operate at different frequency bands and are independent of each other. This paradigm is what we call *cross-frequency communication* (CFC). Additionally, cross-technology cooperation is not about exchanging data but going further to cross-technology localization. Many exciting applications emerge as the boundaries between IoT technologies further dissolves [20, 21]. The ultimate CTM will be our destination. In the envisioned CTM IoT world, the barriers between IoT devices disappear entirely, cross-technology cooperation is unimpeded, and all IoT devices work together as a hyper-connected intelligent organism. The following chapter will provide an in-depth review of the four stages of CTM development.

2.1 Phase I: Wireless Coexistence

In the early stage of IoT, CTIs between heterogeneous IoT networks attracted much attention. The traditional in-protocol media access control schemes, such as carrier sense multiple access (CSMA) and clear channel assessment (CCA), are insufficient for coupling the CTIs [22]. Thus, the research community proposes many new coordination schemes for radios intelligently by eliminating the CTIs to achieve wireless coexistence. We can divide CTI elimination technologies into three categories. Some proposals initially attempt to identify the surrounding CTI and notify the user so that the user can switch off the interference [23]. For example, RFDump [23] introduces a software-defined wireless monitoring

platform to detect the CTIs. Such solutions simply turn off interference and ignore the communication needs of other devices. Even if they have spare capacity in the wireless channel, they cannot fully utilize it. They are too low efficiency to adapt to the demands of wireless technologies and are soon to be abandoned.

The second approach attempts to avoid interference by multiplexing the signals on time domain [24–26], frequency domain [27–33] or space domain [8, 34]. The idea of multiplexing is widely used for sharing scarce channel resources within the same wireless network. It is considered to be a promising solution for coupling the cross-technology coexistence. The difficulty of this category of methods lies in coordinating the devices working with heterogeneous physical-layer protocols to use the channel resources efficiently. For example, in the time domain, many redundant transmission gaps are generated in modern wireless protocols [25]. These blank time slots can be directly exploited for interference-free communication without the cooperation of interfering devices. WISE [25] proposes a statistic model to exploit the abundant white space during the WiFi traffic and enables ZigBee links under heavy WiFi interference. However, such solutions require one to suspend the transmission during the CTIs and are unsuitable for delay-sensitive tasks. A series of cooperative coexistence mechanisms are proposed to further balance the demands from different devices on the time domain [24, 26]. CBT [24] develops a cooperative scheme to enables a stable coexistence between ZigBee and WiFi. The core idea of CBT [24] is a cooperative busy tone for ZigBee node to enhance the mutual observability between ZigBee and WiFi for achieving coexistence on the time domain. Generally, solving the interference on time domain can not support full-duplex communication and requires nodes to keep sniffering the cooperative signaling, which is a huge power cost to IoT devices. The frequency-based isolation, such as fine-grained spectrum fragmentation techniques [30,31], adapting channel width [32,33], OFDM subcarrier suppression [29], can deal with this problem. The frequency multiplexing solution also increases the wireless spectrum efficiency.

The third approach uses mitigation strategy or interference nulling to over-

come the interference [35–37]. These methods are based on the characteristic that wireless signals are linearly superimposed in space. Thus, they can separate the interference to obtain the uncontaminated signal. On the basis of this theory, ANC [35] eliminates the interference created by known packets to increase the networking capacity. MIXIT [36] goes further, by directly leveraging network coding to transmit linear combinations of clean symbols for improving throughput. Interference is more common in backscatter systems without a carrier sensing scheme. Numerous backscatter interference mitigation systems are proposed. For example, Bigroup [37] and FlipTracer [38] propose backscatter collision recovery algorithms by clustering the samples on constellations.

2.2 Phase II: In-band Cross-Technology Communication

In the next phase, the cross-technology interference is considered the chance to communicate. A large body of related works [39–41] pays attention to the CTC, especially that among Bluetooth, ZigBee, and WiFi in the 2.4GHz ISM band. The basic idea of CTC is to make data packets recognizable to other protocols by changing the physical layer morphology of the packet. The challenge is to create a hidden pattern for modulating data without changing the original protocol. CTC technologies develop from the packet level to the symbolic level. In packet level, the data are modulated by changing packet length [42, 43], packet timing [40], packet sequence patterns [44, 45] and packet energy [41]. Basically, this category of methods regards each packet as a sample that can be used for modulation. However, the length of the symbol represented in the packet level is too long to achieve high-speed data transmission. The symbolic level modulation is explored for the CTC to further achieve high throughput communication [39, 46]. It attempts to emulate the desired signal samples by samples on the physical layer. For example, WEBee [39] first achieves high-throughput CTC between WiFi and Zigbee via physical-level emulation. The success of WEBee lies in the fact that

the high-order Quadrature Amplitude Modulation (QAM) used by WiFi is backward compatible with the low-order Phase Modulation (PSK) used by ZigBee. In other words, WEBee can only be used to communicate from high-speed devices (WiFi) to low-speed devices (ZigBee). BlueBee [46] also improves the throughput between Zigbee and Bluetooth by emulating the PSK modulation with GFSK modulation. These previous emulation tricks inspired us to enable CTC through fine-tuned physical layer tuning, which gave rise to Tagcaster (Chapter 5). Aside from the technologies on 2.4 GHz, a few works [47] explore the CTC on the sub-GHz ISM band. For example, LoRaBee [47] bridges the LoRa and ZigBee by payload encoding.

Another way to achieve CTC is to leverage the backscatter node to help build the connection between heterogeneous networks [20, 48, 49]. Similar to the RFID tags, backscatters are battery-free devices that modulate data by reflecting the source signals in different reflective states. The signal source can be any ambient RF source such as TV tower signal [50], Bluetooth [20], Wi-Fi [51], FM Radio [52] and so on. Complex modulation, including FSK [53], QAM [54], and OFDMA [55], can be achieved with dedicated current backscatter technology. Thus, backscatter node can directly communicate with many IoT devices including Wi-Fi [49, 55, 56], Bluetooth [53], ZigBee [20], FM Radio [52] and so on. Although many efforts have been made in backscatter communication over the past decade, key technical open problems remain under-explored [57]. The communication coverage of the backscatter network is especially very limited because the RF harvesting sensitivity of the backscatter node is low. Hence, there is still a long way to go before backscatter-based CTC making a real difference.

2.3 Phase III: Cross-Technology Mutualism Start-up

Our story starts from this stage, and the preliminary CTM is uncovered. The first part is to achieve cross-frequency communication. Traditional CTC can only

bridge the heterogeneous networks working on the overlapped frequency band. [1] However, IoT devices with different protocols work on different frequency bands in most cases. For example, some IoT devices (e.g., RFID) work on the 800/900 MHz ISM band, and the others work (e.g., WiFi) on the 2.4 GHz ISM band. The connection between devices working on different bands is an essential part of CTM. Only CFC can really enable an unblemished CTM. CFC, as a more advanced CTC, is difficult to be achieved and still an open problem. The reason is that the RF frontend of modern wireless systems is usually optimized for a specific band, and the signal from the other band will be filtered out. This task is generally considered impossible.

Recently, few attempts has used the nonlinearity effect in wireless systems to enable CFC [58, 59]. The nonlinearity effect is a common cross-frequency interference phenomenon in the wireless system. Wireless systems with nonlinear effects generate a series of undesired frequency components besides the fundamental signal. These nonlinear components decrease the communication efficiency and disturb the communication on the other band [60]. Thus, the nonlinear effect has been treated as a crucial threat to wireless communication for a long time. Most works focus on building analytical models and methods for eliminating the additional frequency components [61,62]. In the CTC area, some pioneers catch these new frequency components generated by nonlinear effects as a chance for CFC. For example, nonlinear phenomenon in RFIDs, which is called harmonic backscatter, was reported in [58, 63–68]. [58] and [69] explore the harmonics as a secondary communication channel. [58] use harmonics to enhance the communication between the reader and tags. [69] uses the harmonics to achieve the multi-frequency continuous wave ranging and further localize tags in 3D space. Besides the nonlinear phenomenon in RF devices, it has also been found in the acoustic system and been explored for bridging the ultrasound and ordinary acoustic devices [59, 70, 71]. Backdoor [59] leverages the nonlinearity ef-

[1] Backscatter-based CTC technology can shift carrier frequency only in a small range (hundreds of kHz) [20]. In other words, the backscatter node can only talk to devices working on the adjunct channels other than a different band.

fect in the microphone to construct a communication channel between ultrasound speakers and a microphone. DolphinAttack [70] and LipRead [71] further utilize such channel to attack the voice-enabled devices by sending inaudible commands. These early-stage attempts show that the nonlinearity effect can be explored as a bridge linking heterogeneous technologies. We also made our efforts in this area. Three works, TiFi (Chapter 4), Tagcaster (Chapter 5), and UPS+ (Chapter 6), in this dissertation all explore the nonlinearity effect to achieve CFC. Notably, nonlinearity is not perfect and is only one possible way to achieve CFC. We are still working on a better way to achieve CFC.

Meanwhile, we go beyond cross-technology communication to cross-technology cooperative localization in this stage. As we discussed before, the cross-technology localization system has many advantages, including better expansibility and lower cost. However, it has not received sufficient attention in the past and is only now taking off. We are fortunate to be involved in the early development of this field. We propose two cross-technology localization systems: UPS+ (Chapter 6) and iArk (Chapter 7). As far as we know, apart from ours, only one published work, Wibeacon [72], is available in this area. However, we believe there is huge potential in cross-technology cooperative localization to be explored.

2.4 Phase IV: Cross-Technology Mutualism

CTM is a goal for the future development of the IoT ecosystem. Cross-technology communication and localization are only the primary forms of CTM. The ultimate goal of CTM is to create a harmonic and intelligent IoT ecosystem similar to the natural ecosystem where IoT devices operate together as an intelligent organism. However, designing an intelligent harmonic ecosystem artificially is difficult. Thus, we need to empower the IoT networks with intelligence to enable them to learn the way to cooperate and develop by themselves. We should design the IoT ecosystem as we do in designing an artificial neural network. We only need to construct the basic network units and achieve the basic network func-

tions, and the networks learn all other specific details by themselves. The first three phases of the effort laid a good foundation for CTM. They give us the most basic but essential network infrastructures: cross-technology communication and cross-technology localization. We can finally realize the CTM along this path.

Chapter 3

Fundamentals on Cross-Frequency Communication: Nonlinearity

Cross-frequency communication (CFC) is essential for breaking down the barriers between wireless technologies to achieve flexible CTM. In this chapter, we introduce one rationale behind the CFC, namely, the *nonlinearity effect*, which will be further explored to build the following CTM systems. We first explore the principle of how the nonlinear effect arises and then .

3.1 Nonlinearity Effect

Consider a system with input signal x and the output of system is y. According to the Stone Weierstrass theorem [73], all continuous function defined on a closed range can be expressed by a polynomial expansion in any desired accuracy. Thus, any system function can be modelled as following:

$$y = \sum_{k=1}^{\infty} a_k x^k = \overbrace{a_1 x}^{\text{Linear}} + \overbrace{a_2 x^2 + a_3 x^3 + \cdots}^{\text{Nonlinear}} \tag{3.1}$$

where a_k is the strength of k-th component. The first component is linear to the input signal and is called *fundamental component*. The remainders are **nonlinear components**, which are called the *second-order* component, the *third-order* component, the *fourth-order* component, ..., successively. If the output only contains the first-order component, then the system is linear. Otherwise, the system is nonlinear. The linear system is preferred in communication due to its simplicity and ease of operation. However, the absolute linear system is nearly nonexistent, and everyone exhibits more or less nonlinearity. Nonlinearity is the essential nature of the world.

3.2 Nonlinear Phenomenon and Applications

The conventional wisdom considers that the nonlinearity effect might cause undesired new frequencies out of the desired band. Here, we observe the bright side of the nonlinearity effect: the newly created frequencies allow us to achieve the CFC between different wireless systems. In the following section, we show two typical nonlinear phenomenons and how to use them to achieve CFC.

3.2.1 Upconversion

The first phenomenon is conduct the upconversion of the input signal. Suppose the input signal of a nonlinear system is a single-tone sinusoidal signal, which is denoted by $x(t) = \cos(2\pi ft)$. Then, the output signal $y(t)$ is expressed as follows:

$$\begin{aligned} y(t) =& a_1 \cos(2\pi t) + a_2 \cos^2(2\pi ft) + a_3 \cos^3(2\pi ft) + \cdots \\ =& \frac{1}{2}a_2 + (a_1 + \frac{3}{4}a_3) \underbrace{\cos(2\pi ft)}_{\text{1st-order}} + \frac{1}{2}a_2 \underbrace{\cos(2\pi(2f)t)}_{\text{2nd-order}} + \frac{1}{4}a_3 \underbrace{\cos(2\pi(3f)t)}_{\text{3rd-order}} + \cdots \end{aligned} \tag{3.2}$$

The equation indicates that the spectrum of the output signal will have not only the fundamental frequency (1st-order) but also the harmonic frequency components (2nd-order, 3rd-order, $\cdots$). Those harmonic signals are out-of-band and

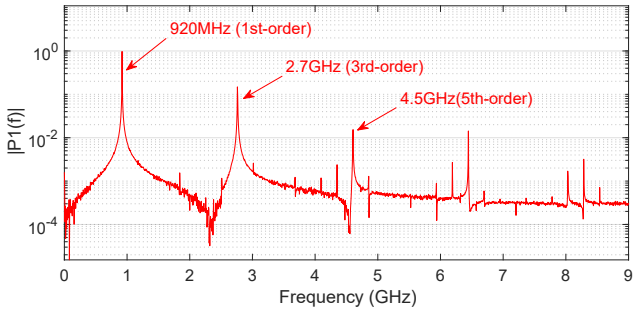


Figure 3-1: Spectrum of a tag's backscattering signal. The signal appears at 920 MHz, 2.76 GHz, 4.605 GHz, and 6.445 GHz, when the tag is queried at 920 MHz.

the interferences with the other wireless devices. Usually, the higher the order of the signal, the lower its intensity. Here we use the RFID tag as an example. [1] Fig. 3-1 shows a typical nonlinear upconversion in the RFID tags. Hence, we send a signal to a RFID tag at 920 MHz and the spectrum of RFID tag backscatter signal contains many harmonics including the 2nd-order (1840 MHz), 3nd-order (2760 MHz) and so on.

Such a scheme can also be used to connect the devices working on the harmonics. An obvious application for this phenomenon is to bridge the 800MHz UHF band to the 2.4 GHz ISM band. For example, when an RFID reader uses an 800MHz carrier, the backscattering signal of a tag will appear mainly at 800MHz, 1.6GHz, 2.4GHz, 3.2GHz, and so on. The third-order harmonic allows us to achieve CFC between RFID systems and 2.4GHz systems such as Wi-Fi or Bluetooth. This feature is fundamental for the TiFi. Additional details are shown in the Chapater 4.

3.2.2 Downconversion

Next, nonlinearity can also downconvert the signal. If the input signal is composed of two different sinusoidal signals, that is $x(t) = \cos(2\pi f_1 t) + \cos(2\pi f_2 t)$, then the output signal becomes more complicated. For clarity, we assume that the harmonics higher than the third-order are too weak to be detected, and only

[1] The rectifier of an RFID tag has a strong nonlinear effects and we deeply analyze the cause of its nonlinearity in Chapter 3.3.

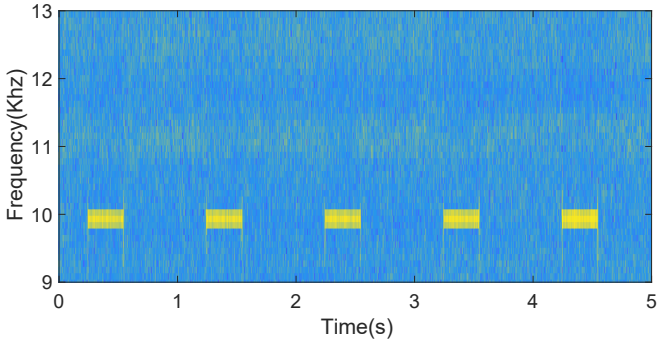


Figure 3-2: Time-frequency spectrum of microphone. The built-in recorder of an iPhone 8 is used to capture two ultrasonic signals that are transmitted simultaneously: pulses at 40 kHz and a single-tone continuous wave at 50 kHz.

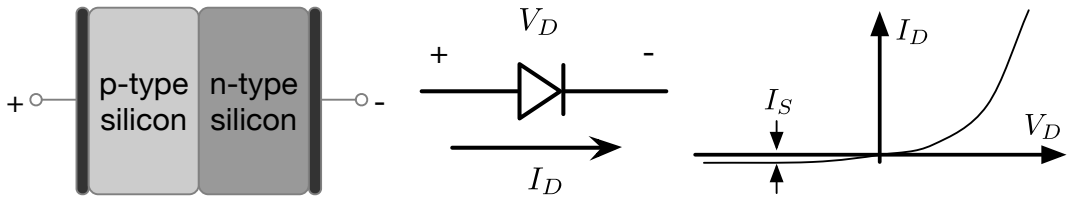
the first two orders play the key role. Then we have the following output signal:

$$\begin{aligned} y(t) =& a_1 x(t) + a_2 x^2(t) \\ =& a_1 \cos(2\pi f_1 t) + a_1 \cos(2\pi f_2 t) + a_2(\cos(2\pi f_1 t) + \cos(2\pi f_2 t))^2 \\ =& a_1 \cos(2\pi f_1 t) + a_1 \cos(2\pi f_2 t) + \frac{1}{2} a_2((2 + \cos(2\pi 2 f_1 t) + \cos(2\pi 2 f_2 t) \\ & + \cos(2\pi(f_1 + f_2)t) + \cos(2\pi(f_1 - f_2)t)) \end{aligned} \tag{3.3}$$

Interestingly, two new frequencies (i.e., $f_1 + f_2$ and $f_1 - f_2$) are created in the output. Here we use the microphone system as an example. The amplifier in microphone makes it nonlinear. [2] We send two acoustic signals at the same time: periodical pulses at 40 kHz and continuous wave at 50 kHz and use a microphone to record them. Fig. 3-2 shows the time-frequency spectrum of the captured signals. Clearly, the signal is downconverted to the difference frequency 10 kHz due to the nonlinearity.

This phenomenon is a kind of intermodulation [74] and shows that we can downconvert a signal at a high-frequency band (f_1) to the low-frequency band ($f_1 - f_2$) with the help of an nearby frequency component (f_2). Both Tagcaster and UPS+ use this idea to achieve CFC. It should be noted that such a down-conversion phenomenon also exists in the RF system. In Tagcaster, we broadcast the UHF RFID signal at 920 MHz with a shadow carrier at 920.5 MHz. The

[2] We will deeply discuss about this phenomenon in Chapter 3.3 and Chapter 6.



(a) P-N junction **(b) Circuit Model** **(c) I_D-V_D Curve**

Figure 3-3: Inner structure of diode and its current-voltage characteristics.

nonlinearity of the RF amplifier of an AM radio receiver can downconvert the 920 MHz to the 500 kHz AM radio band. On the basis of this idea, Tagcaster realizes the CFC between UHF RFID and AM radio. Chapter 5 and Chapter 6 provide more details on this aspects.

3.3 Cause of Nonlinearity

The nonlinearity in the communication system orginates from the nonlinearity effect of electronics. The most basic electronic elements, namely, diodes and transistors, have nonlinearity. They further construct the high-level electronic modules, such as amplifier and rectifier. These complex modules naturally behave nonlinearly. This section comprehensively analyzes the cause of nonlinearity from the basic electronic elements to the overall modules.

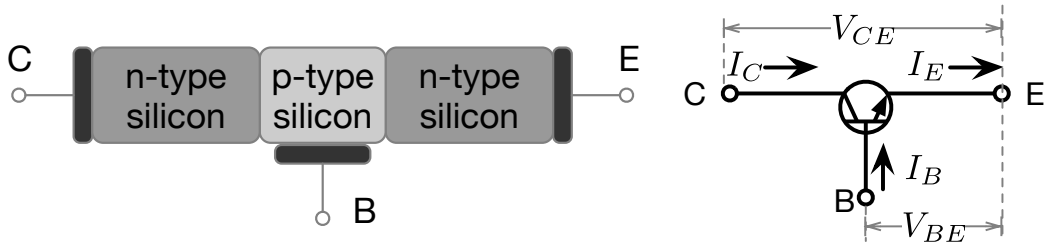
3.3.1 Nonlinear Elements

Diodes and transistors are the basic units of nearly all semiconductor systems.

■ **Diodes**: A diode is a crystalline piece of semiconductor material with a P-N junction connected to two electrical terminals. The inner structure of a diode is shown in Fig. 3-3. According to the Shockley diode model [75], the diode output current I_D and the across voltage of a diode V_D satisfy the following equation:

$$I_D = I_S \left(e^{\frac{V_D}{nV_T}} - 1 \right) \approx I_S \left(e^{\frac{V_D}{V_T}} - 1 \right) \tag{3.4}$$

where V_T is the thermal voltage, I_S is the reverse bias saturation current, and n is the quality factor ranging from 1 to 2. For clarity, n is usually approximated to



(a) Transistor **(b) Circuit Model**

Figure 3-4: The inner structure of a transistor.

1. Notably, the system function is an exponential function, which is a nonlinear function. The Eqn. 3.4 can be further expanded using a Taylor series as follows:

$$I_D = I_S\left(e^{\frac{V_D}{V_T}} - 1\right) = I_S\left(\frac{V_D}{V_T} + \frac{V_D^2}{2V_T^2} + \frac{V_D^3}{6V_T^3} + \cdots\right) \tag{3.5}$$

It can be seen that the output current is composed of a fundamental component of $\frac{V_D}{V_T}$ and many harmonic components (e.g., $\frac{V_D^2}{2V_T^2}$, $\frac{V_D^3}{6V_T^3}$, ...), which are nonlinearly proportional to the input V_D or V_T. Thus, a diode is a nonlinear component.

■ **Transistors**: Transistors are used to amplify or switch electronic signals. A transistor is a semiconductor device with three terminals. A voltage or current signal applied to one pair of the terminals of the transistors controls the current through another pair of terminals. According to their structure, transistors can be classified into two main types: field-effect transistor (FET) and bipolar junction transistor (BJT). Their difference is that FETs are controlled by voltage, but BJTs are controlled by current. As shown in Fig. 3-4, both types can be considered as two diodes (P-N junctions) sharing a common region that minority carriers can move through [76]. Such internal property of a transistor also triggers the nonlinearity effect. Here, we take the BJT as an example. A typical emitter configuration circuit of a BJT is shown in Fig. 3-4 (c). According to the Ebers-Moll model, the output current I_C is well modeled as:

$$I_{\mathrm{C}} = \beta I_{\mathrm{B}} = \beta I_0\left(e^{\frac{V_{\mathrm{BE}}}{V_{\mathrm{T}}}} - 1\right) \tag{3.6}$$

where β indicates the current gain, V_{BE} is the base-emitter voltage, and V_{T} is thermal voltage. Similar to the model of a diode, the abovementioned equation can be also expanded as follows:

$$I_{\mathrm{C}} = I_{\mathrm{B}}\left(e^{\frac{V_{\mathrm{BE}}}{V_{\mathrm{T}}}} - 1\right) = I_{\mathrm{S}}\left(\frac{V_{\mathrm{BE}}}{V_{\mathrm{T}}} + \frac{V_{\mathrm{BE}}^2}{2V_{\mathrm{T}}^2} + \frac{V_{\mathrm{BE}}^3}{6V_{\mathrm{T}}^3} + \cdots\right) \tag{3.7}$$

Thus, a transistor can also trigger the nonlinear effect as same as a diode does.

3.3.2 Nonlinear Modules

High-level electronic modules, such as rectifiers and amplifiers, also have a nonlinear effect because they are made of diodes and/or transistors.

■ **Nonlinearity in Rectifier**. The rectifier is a crucial circuit for low-power IoT devices, such as RFID tags or other backscatter nodes, to collect the RF energy in the air. It can convert the alternating current (AC) signal (e.g., RF signal in the air) to a direct-current (DC) signal, which can be used to power up the chip. A diode can form a simplest one-stage rectifier [77]. The output signal of such a rectifier is the same as that of a diode. Thus, the rectifier behaves nonlinearly in theory. In practice, we may use a multi-stage rectifier to achieve higher output voltage. Such rectification module consists of a series of nonlinear devices and will produce high-frequency spectral components at the output. Fig. 3-5 shows a typical rectifier circuit with two-stage design. In such a rectifier, the input signal passes through a rectification module and is then filtered by a low-pass filter. The low-pass filter is usually used to eliminate the additional high-frequency components. In the other view, such additional components can be used to increase the charging efficiency [67] or create the additional communication channel [78]. We will introduce additional details and potential applications of such nonlinear phenomena in Chapter 4.

■ **Nonlinearity in Amplifier**. The amplifier is a key module in the analog frontend of all wireless communication systems. It amplifies the signal in both receiver and transmitter. In the receiver, it is called a low noise amplifier (LNA),

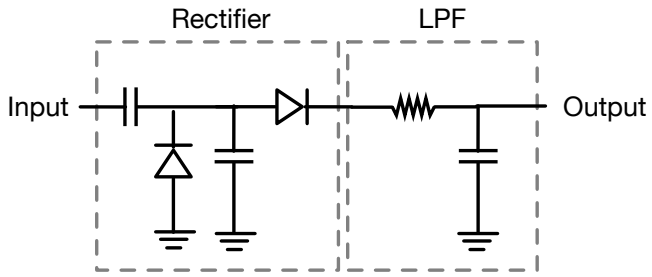


Figure 3-5: Schematic of a simple two-stage rectifier circuit

which is used to amplify the weak original RF signal for decoding. LNA determines the sensitivity and linearity of a receiver. In the transmitter, we use a power amplifier (PA) to increase the transmitter power for enlarging the communication range. PA dominates the overall system energy consumption, heat dissipation, and linearity.

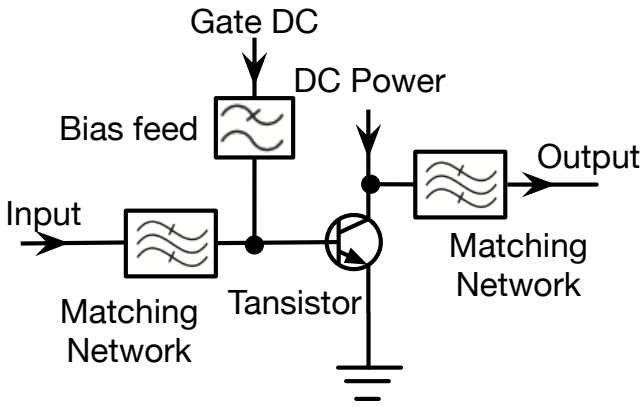


Figure 3-6: Schematic of a simple amplifier circuit

A general schematic of an amplifier is shown in Fig. 3-6. The figure shows that it can be seen that an amplifier is composed of a transistor and two matching networks for input and output, respectively. The matching networks are used to maximize the transmission power and minimize the signal reflection. A bias feed is used to set the DC bias of the transistor. The amplifier also exhibits nonlinearity due to the nonlinearity of the transistor. Suppose we input a single tone signal to an amplifier. The output of the amplifier will contain many harmonics, and the output power of desired single tone signal is reduced. Therefore, conven-

tional circuit designs generally require the amplifier to be as linear as possible [3], which makes the amplifier as efficient as possible. In the other point of view, such harmonics allow us to create a new communication channel between devices working on the different frequency bands. We will further introduce the potential applications of such phenomenon in Chapter 5 and Chapter 6.

In conclusion, the presence of harmonics caused by the nonlinearity effect provides us the fundamentals of CFC. In the following chapters, we will further present practical designs to show the realization of CFC among different wireless systems is achieved.

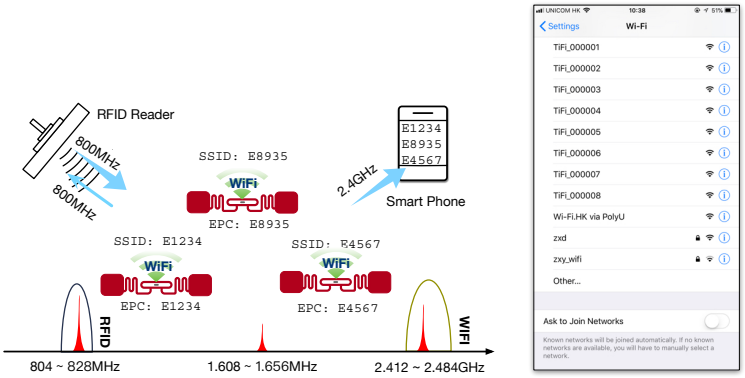
[3]We usually use 1 dB compression point (P1dB) to evaluate the linearity of an amplifier, which is the output power level at which the gain decreases 1 dB from its constant value. An amplifier with larger P1dB has the better linearity.

Chapter 4

Identifying UHF RFIDs with WiFi

As the major enabler of automatic ID technology, RFID systems are being increasingly used in everyday scenarios ranging from object tracking, indoor localization [79], and vibration sensing [80], to medical-patient management because of the extremely low cost of commercial RFID tags (e.g., as low as 5 cents per tag). Recent reports show that many industries, such as healthcare and retail, are moving towards deploying RFID systems for object tracking, asset monitoring, and the emerging Internet of Things [81]. A typical UHF RFID system consists of a reader and numerous tags and operates at a frequency band of $840 \sim 920$MHz. The tags are battery-free and harvest energy exclusively from the signals emitted by the reader.

After decades of development, RFID tag chips have made tremendous progress. RFID tags are becoming smaller and smaller and more powerful. However, the progress in RFID reader is tiny. Today, RFID tags can still only be collected by dedicated bulky readers rather than be supported by smartphone. Hence, the RFID is still unpopular in customer-orientated market. The industry has exerted considerable effort to bridge this gap. For example, Phychips Inc. [82] developed a small reader that can be plugged into a smartphone through its headphone jack. ImpinJ Inc. [15] released a special RFID holder, into which the user can



(a) Architecture **(b) Snapshot**

Figure 4-1: TiFi architecture. (a) At a high level, TiFi transforms each tag to a WiFi AP, which broadcasts legitimate beacons that regards the tag's EPC as its SSID. Any commercial smartphone could capture and recognize these beacons with a built-in WiFi AP scanner, thereby obtaining these tags' EPCs.; (b) The figure shows a snapshot of the built-in scanner of iPhone 8, which precisely explores our RFID tags. The SSIDs of these tags are in the form of TiFi_XXX.

insert his/her smartphone for RFID scanning. Alien Technology integrates WiFi and (or) Bluetooth modules into readers to provide temporary connection with smartphones. These trade-off solutions (additional details are provided in §4.1) aim to promote the integration of RFID technology into smart devices. However, they have achieved minimal progress a few years after their introduction because they either introduce extra hardware cost or increase the deployment complexity.

In this chapter, we propose TiFi, a CTM system that allows a commercial WiFi receiver (e.g., a mobile phone) to identify the commercial off-the-shelf (COTS) UHF RFID tags in a near field without changing hardware and firmware. Fig. 4-1 illustrates the usage scenario. At a high level, TiFi turns a tag into a virtual WiFi AP in accordance with the 802.11b protocol. Regarding the Electronic Product Code (EPC) as its service set identifier (SSID), the virtual AP periodically broadcasts *legitimate* WiFi beacons that can be recognized by unmodified WiFi receivers. Consumers can then identify RFIDs using their smartphones in the same way of discovering new APs. In addition, TiFi applies the fact that the near-field signal strength of a tag is hypersensitive to the distance in providing a proximity-based localization service, e.g., Near-Field Communication (NFC)

payment.

TiFi allows RFID technology to benefit from the WiFi economies of scale and significantly reduce the barrier of adoption. Thus, TiFi aims to enable new RFID applications. For example, massive assets in a warehouse can automatically broadcast their information to a staff's mobile phone. Consumers can order and pay for snacks in a vending matching by placing their smartphones close to the item they want. People can use smartphones to directly obtain advertisements from the tags embedded in bus stop posters and street signs.

Yet, transforming TiFi into a practical system may be unfeasible because of the following two issues.

• **Cross-Frequency Communication (CFC)**: *How can UHF tags be audible to a WiFi receiver?* The main challenge is the huge frequency gap between RFID and Wi-Fi. We observe that UHF tags resonate the reader's continuous wave (CW) not only at the fundamental frequency (e.g., first at 820MHz) but also at the harmonics (e.g., second at 1.64GHz, third at 2.46GHz) because of the nonlinearity effect of its rectenna. In particular, the absence of nonlinear treatment allows the antennas of RFID tags to radiate the harmonic signals, thereby leading to harmonic backscattering. Unlike the conventional wisdom that considers harmonics as a detrimental 'pollution', TiFi utilizes those around 2.4GHz as a second channel to communicate with WiFi receivers.

• **Cross-Protocol Communication (CPC)**: *How can unnoticed harmonic backscattering carry legitimate WiFi beacons?* Even if the harmonic backscattering is tuned at the WiFi band and is correctly sampled by hardware, a WiFi receiver does not recognize these packets for mismatching WiFi protocol. To transform a tag's packet to a WiFi AP, we craft the reader's continuous wave to simultaneously create RFID Gen2 packets as well as WiFi 802.11b packets, thereby achieving CPC.

Summary of the Results. We implement a prototype of TiFi using Universal Software Radio Peripheral (USRP) N210 software radios and test 7 types of COTS RFID tags. Our evaluation results demonstrate that TiFi allows a commer-

cial WiFi receiver (e.g., a mobile phone) to identify RFID tags within a range of 2m. TiFi performs comparably to the existing commercial mobile RFID reader, and has the additional serviceability to WiFi receivers. We demonstrate TiFi in two real-world scenarios, identifying RFID-tagged books on shelf and RFID-tagged objects in closed box. Across the two applications, TiFi could achieve the same level of effectiveness reported in the quantitative results above. In particular, the second scenario also shows the ability of TiFi to identify the RFID tags through their WiFi beacons in NLOS scenarios.

4.1 Motivation

Existing CTCs fail to work in our scenario for their incompatible CFC. This section thereby mainly examines other potential non-CTC solutions. Our objective is not to complete the list, but to motivate our design.

• **Limitations of Mobile Readers**: The first type of solutions extends the function of mobile phones to identify UHF RFID tags by using additional accessories. For example, the mobile readers released from ImpinJ [83] and Alien [84] can accommodate the user's smart phone through the holder on the top. The user can manipulate the reader through the Bluetooth connection and Wi-Fi connection [85]. However, they just provide data interfaces for the whole backend dataset. The user can not use this service to identify tags in the proximity directly because he can not know which one in the dataset is close. PHYCHIPS [82] invents a novel solution: the reader module can be plugged into a smartphone through the headphone jack and manipulated with the acoustic signals. However, these limited services and bulky accessories are not appreciated by the current market years after their introduction because of their inconvenience. TiFi allows a user to identify the tagged product instantly when putting its smart device close to the tag.

• **Limitations of HTTP Readers**: Many industrial grade readers (e.g., ImpinJ R410 [15], Alien ALR-9900+ [86], ThingMagic M6 [85], etc) can provide

Table 4.1: Comparison with other techniques

	Cost	CPC	CFC	NFI	Convenience
Http Readers	High	APL	No	No	Low
Mobile Readers	Median	APL	No	No	Low
Backscatters	High	PHY	No	No	Median
HF-NFC	Median	No	No	Yes	High
TiFi	Low	PHY	Yes	Yes	High

HTTP web service and work as stand-alone HTTP stations. Mobile phones can access these readers via LAN. In general, the HTTP reader is cumbersome for customers to pair its device with a reader in a supermarket especially when numerous readers are deployed.

• **Limitations of Backscatters**: Our work is inspired by a pioneering work (i.e., Passive WiFi [56]), which enables backscatters to emit WiFi signals. Similarly, there are many other backscatters, (e.g., Passive WiFi [56], FM backscatter [52], Ambient backscatter [50], Lora backscatter [87], HitchHike [88], BackFi [49] and Interscatter [20]). TiFi differs from them in three aspects. First, all these backscatters are required to modify the logics of "tags" (i.e., backscatter). TiFi does not have to change but work for commercial RFID tags, billions of which have been deployed around the world. Second, they must operate at the same frequency band with the receivers. Third, TiFi has the unique advantage of dual standard compliance (i.e., WiFi and RFID Gen2).

• **Limitations of HF-NFC**: The near-field identification (NFC) function of TiFi is similar to the HF-NFC that works at 13.56 MHz. Many mobile phones (e.g., iPhone 7/8/x and Samsung Galaxy series) have already integrated an NFC reader for mobile payment. Unfortunately, HF NFC tags use the magnetic induction for energy harvesting and communication. This underlying physical mechanism limits the operating range of NFC within a few centimeters. The short range limits the adoption of HF-NFC in the warehouse or supermarket management, which usually requires a long communication range to cover a large space for monitoring a batch of static or moving items in real-time.

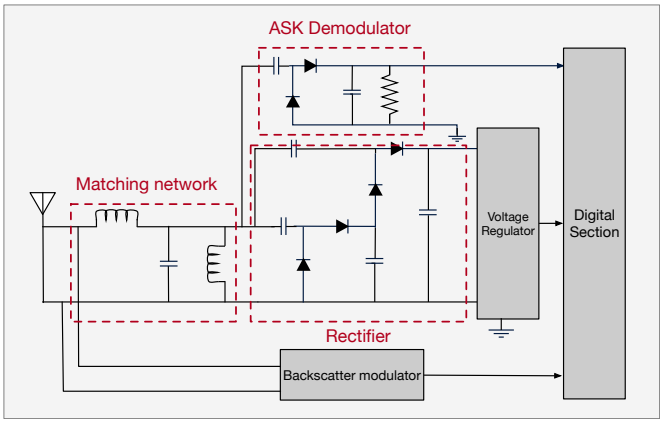


Figure 4-2: Simplified RFID tag architecture. The rectenna consists of two or more stages of voltage-doubling rectifier with nonlinearity effect that produces harmonics signals apart from the fundamental.

Advantage of TiFi: We summarize the advantages of TiFi over the potential solutions listed in Table 4.1. The table presents a generalized concept of CPC in which two protocols that can communicate with each other via gateway or directly are considered as CPC. NFI refers to near-field identification, supported by both HF-NFC and TiFi. In particular, the CFC functionality is a unique feature of TiFi. In short, TiFi is highly efficient in the physical-layer cross-protocol data exchange without pairing or connection procedure; TiFi provides fine-grained proximity localization, supplementing NFC-similar functions (e.g., mobile payment) to UHF RFIDs; TiFi can enable either non-HTTP or non-WiFi legacy readers to support WiFi communications; particularly, the CFC functionality is a unique feature of TiFi.

4.2 Towards Harmonic Backscattering as a Cross-Frequency Channel

In this section, we introduce the phenomenon of harmonic backscattering in detail as well as verify it through actual experiments across different models of tags.

4.2.1 Primer on Harmonic Backscattering

A UHF RFID system consists of a reader and multiple tags. The reader continuously generates a high-power continuous wave (CW), from which tags can harvest energy regardless of whether either the reader or the tag is transmitting. Tags will immediately lose the power if they are shielded from the CW even for a while. Reader and tags use ON-OFF keying (OOK) to modulate their data. Without conventional transceivers, tags transmit a "1" bit by changing the impedance on their antennas to reflect the reader's signal and a "0" bit by remaining in their initial silent state.

A passive UHF RFID tag consists of an antenna and an integrated circuit (IC). One of the tasks performed by the IC is the rectification, in which *rectifier* and the *antenna* (i.e., commonly called "rectenna" in literature) convert the alternating current (AC) induced by the CW sent by the reader into a direct current (DC), thereby providing the energy for the other part. Fig. 4-2 highlights the rectifier section as part of the passive RFID tag architecture. This section contains three common parts: (1) the antenna, (2) the N-stage rectifier circuit, and (3) the antenna-rectifier impedance matching network. In particular, the rectenna is based on a Cockcroft-Walton Circuit that consists of two or more stages of voltage-doubling rectifiers. The nonlinearity effect of these diodes produces *harmonics signals* in addition to the *fundamental signal* [89].

Conventional energy harvesting circuits typically use a harmonic confinement technique to suppress the harmonics and improve their RF-to-DC power conversion efficiency [90]. However, for the commercial interests and usefulness of RFID tags, tag chips and antennas are separately designed and optimized only at the frequency band of UHF RFID. Antenna design begins directly from the knowledge of one impedance value, which is the impedance of the IC at the fundamental frequency described on manufacturer data sheets. The design process then only ensures the matching at the fundamental frequency, and provides no treatment for the harmonic currents. Consequently, in accordance with the the-

ory [91] and the measurement [67, 92, 93], the *absence* of nonlinear treatment allows the tag antenna to radiate the harmonic signals generated by the rectifier, thereby resulting in *harmonic backscattering*.

4.2.2 Harmonics Formulation

As discussed in Chapter 3, tags can backscatter harmonics when queried with a modulated or unmodulated reader signal. Suppose the input reader's CW is denoted by S. Then, the backscattered signal denoted by S_{out} is given by:

$$S_{\text{out}} = \sum_{k=1}^{\infty} A_k S^k = \overbrace{A_1 S}^{\text{Linear}} + \overbrace{A_2 S^2 + A_3 S^3 + \cdots}^{\text{Nonlinear}} \tag{4.1}$$

where A_k are the gains of the various components introduced by the rectenna. If the incoming signal S is a sinusoidal signal with frequency f (i.e., fundamental frequency), then it outputs linear component $A_1 S$ with the same frequency f. However, the nonlinearity effect can produce many nonlinear components (i.e., harmonics). In particular, if $S = \cos(2\pi f t)$, then the output signal can be expanded using a trigonometry formula as follows:

$$\begin{aligned} S_{\text{out}} =& A_1 \cos(2\pi t) + A_2 \cos^2(2\pi f t) + A_3 \cos^3(2\pi f t) + \cdots \\ =& \frac{1}{2} A_2 + (A_1 + \frac{3}{4} A_3) \underbrace{\cos(2\pi f t)}_{\text{1st-order}} + \frac{1}{2} A_2 \underbrace{\cos(2\pi (2f) t)}_{\text{2nd-order}} + \\ & \frac{1}{4} A_3 \underbrace{\cos(2\pi (3f) t)}_{\text{3rd-order}} + \cdots \end{aligned} \tag{4.2}$$

The equation indicates that the frequencies of these harmonics (1st-order, 2nd-order, 3rd-order, $\cdots$) are exactly an integral multiple of the fundamental frequency. Imaging that the reader uses an 800MHz carrier, then a tag's backscattering signal will appear mainly at 800MHz, 1.6GHz, 2.4GHz, and so on. These backscattered harmonic will not interfere the reader's receiver because creating a filter to reject signals above fundamental frequency is easy. This condition is also another reason why harmonics do not attract considerable attention in RFID systems.

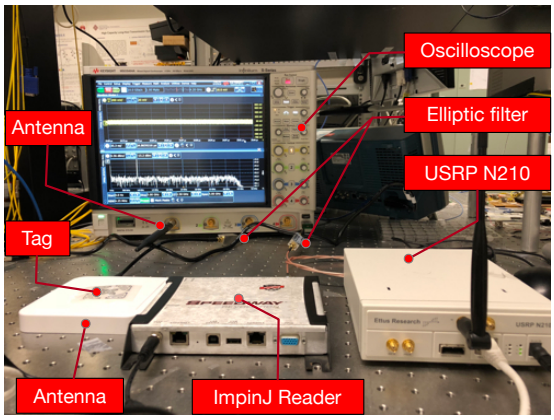


Figure 4-3: Experimental setup for the benchmark. A high-definition oscilloscope with 4G bandwidth is used to sniff the backscattered signals.

4.2.3 Experimental Verification

The fundamental idea of TiFi is to modulate the reader's continuous wave, such that the tag can harmonically backscatter at the WiFi band, thereby enabling the cross frequency communication. To investigate the feasibility of this idea, we run an experiment about the harmonic backscattering as below:

Experimental setup. The setup is shown in Fig. 4-3, where we use a commercial ImpinJ reader to perform the continuous reading at 920MHz. The test tag is a `Monza` tag from ImpinJ [15] and located at a distance of a few centimeters from the antenna and is oriented towards maximum reception and radiation. To observe the backscattered harmonics, we use a high-definition Keysight oscilloscope (i.e., MSOS404A) [94] to sniff the backscattered signals. The oscilloscope is equipped with 4GHz bandwidth and up to 20GSa/s sample rate and can produce a power spectral density (PSD) analysis over time-domain communication signals within GHz-level range.

Results. We acquire the spectrums of received signals when the RFID is backscattering (EPC) and when it is not backscattering (idle) by using the oscilloscope. Fig. 4-4 shows the whole spectrum when tag backscattering, which confirms that the backscattering signals contains numerous harmonics. The low-pass filter cannot attenuate the CW sent from the reader. Thus, the 1st-order signal is extremely stronger than the harmonics. As expected, the harmonics

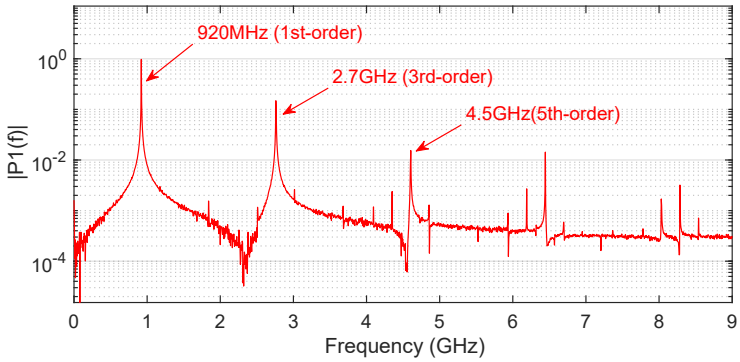


Figure 4-4: Spectrum of a tag's backscattering signal. The signal appears at 920 MHz, 2.76 GHz, 4.605 GHz, and 6.445 GHz, when the tag is queried at 920 MHz.

exactly backscatter at the frequencies of integral multiples of the fundamental frequency.

Second, whether the phenomenon is fairly ubiquitous in RFID tags is unclear. Thus, we repeat the aforementioned experiment across seven types of UHF tags, which are the best selling models in current market. We affirmatively observe the similar harmonic backscattering on these tags. Keeping the reader's transmitting power at 30 dBm, Table 4.2 lists their power of the first three order signals using the unit of dBm. We could see that the harmonic backscattering is indeed a ubiquitous phenomenon across various tags.

Third, as the model expected, only even harmonic components are clearly found. As reported in [91], the harmonic current distributions on a half-wave dipole working at fundamental frequency, which shows that the current at the antenna input is zero for even harmonics. Although if even harmonics are present from the IC, they would not be efficiently reradiated since the half-wave dipole

Table 4.2: Harmonic Power of Tag Response

#	Manuf.	Model	Polar.	Size (cm^2)	1st	2nd	3th
1	ImpinJ	H47	Circular	$4.8 \times 4.8 = 23$	-5	-85	-65
2	ImpinJ	B45	Circular	$2 \times 2 = 4$	-10	-81	-81
3	Alien	9640	Linear	$15.9 \times 1.5 = 23.8$	-6	-83	-68
4	Alien	9629	Linear	$2.55 \times 2.55 = 6.5$	-6	-84	-72
5	Alien	9627	Linear	$3.2 \times 5 = 16$	-6	-86	-73
6	Alien	9620	Linear	$3 \times 1.5 = 4.5$	-12	-85	-80
7	Alien	9610	Linear	$4.4 \times 1.03 = 4.5$	-12	-85	-80

would require zero admittance at even harmonics.

Fourth, the size of the tag is the key factor that seriously affects the backscatter signals at harmonics. For example, the sequence ordered by the size is 9640 ($23.8cm^2$), H47 ($23cm^2$), 9627 ($16cm^2$), 9629 ($6.5cm^2$), 9620 ($4.5cm^2$), 9610 ($4.5cm^2$) and B47 ($4cm^2$), which is consistent with the sequence ordered by the signal strength. In addition, the circular polarization behaves better than the linear polarization for the tags in a similar size. For example, the size of 9620 (-12 dBm), 9610 (-12 dBm) and B45 (-10 dBm) are around $4cm^2$ but the strength of B45 equipped with circular polarization is clearly 2 dBm higher than the other two linear-polarization enabled models. Normally, circular polarization allows tags to harvest much more power from the air than linear polarization. It is worth noting that the materials of tagged items are also a key factor. Many previous works released reports to study the implication of materials on RFID tags. Here, we encourage readers to refer to this detailed report [95], which results are confirmed by our preliminary experiments.

4.2.4 Summary

We explore the harmonic backscattering as the second communication channel to achieve the across-frequency (band) communication. It offers a clear advantage: the harmonic backscattering is induced by the internal hardware property of tags, and thus, it will not require extra circuits for the channel. In the subsequent sections, we will elaborate how to use the harmonic backscattering to communicate with WiFi receivers.

4.3 System Design

Our design engages with three actors: the reader, tags and mobile device. Only the reader is required to be upgraded. The TiFi reader, which operates in the same manner as RFID system, dominates the entire communication and transmits a persistent CW at approximately 840 MHz. It also performs carrier sense on behalf of the tags and helps coordinate medium access control (MAC) across multiple

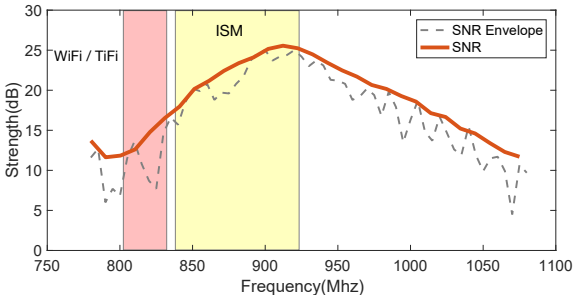


Figure 4-5: SNR of a tag's response when using frequency out of the usual band. The SNR (in blue) is consistently higher than 10dB over our required frequencies outside the ISM band.

tags. All communications completely comply with either RFID Gen2 protocol or the WiFi 802.11b protocol. This work considers the most common usage scenario in which tags are attached to the objects in a warehouse or supermarket and covered by readers. Our goal is to provide mobile device with the capability to identify tags by upgrading the readers. Thus, TiFi offers two core techniques:

- **CFC: Entrapping Tags into WiFi Band**. Tags perform backscattering not only at the fundamental frequency but also at harmonics caused by the nonlinearity effect of the tags' rectennas. The first technique (see Sec. 4.4) leverages such underlying physical property to entrap tags into harmonic backscattering within WiFi band to achieve the CFC.

- **CPC: Converting Tags to WiFi APs**. The second technique (see Sec. 4.5) upgrades a reader from physical layer (aka PHY) to link layer (aka MAC) by modulating its continuous wave. Consequently, the tags backscatter their WiFi beacons periodically and in a timely manner to achieve CPC.

The following two sections elaborate the two techniques.

4.4 Cross-Frequency Communication

In this section, we explain how TiFi makes tags backscatter at WiFi band.

4.4.1 Frequency Diversity

The unavoidable nonlinearity effect results in tags backscattering the third-order harmonic signals at a frequency that is extremely close to the WiFi band. This condition will offer us an opportunity of realizing the cross-frequency communication. WiFi contains in a total of 14 channels with frequencies of 2412MHz, 2417MHz, $\cdots$, and 2484MHz. Each channel has 22MHz band and is spaced 5MHz apart from one another. If we aim to entrap the third-order harmonics into one of these 14 channels, then the reader must operate at the subsequent 14 fundamental frequencies: 803MHz (i.e., 2412/3), 805.6MHz (2417/3), $\cdots$, and 828MHz (2484/3).

We investigate the regulated spectrums of UHF RFID systems across 21 countries or areas worldwide. We find that the current legitimate spectrum mainly distributes between $840 \sim 928$ MHz (highlighted with yellow background in Fig. 4-5). Even the lowest frequency (i.e., 840 MHz used in China) is 12 MHz higher than the highest frequency (i.e., 828 MHz) that we desire. Even so, as reported in the work [96], RFID tags are designed to be able to respond in a wide band (i.e., up to 300 MHz) in order to fit the various specifications although they can only support an extremely narrow communication bandwidth of hundreds of kHz. As shown in Fig. 4-5, our practical experiments confirm this phenomenon that the tag is supposed to operate during the spectrum of $840 \sim 928$ MHz, e.g., $902 \sim 928$ MHz in the US and $840 \sim 845$ MHz (or $920 \sim 925$ MHz) in China, but respond during $750 \sim 1050$ MHz. The design of wideband response is reasonable because products with RFID tags are typically transported all around the world and may be queried using a variety of readers, that conform to different regulations. The wanted fundamental frequencies all exactly follow in the response spectrum.

4.4.2 Dual-Frequency Solution

We adopt a similar dual-frequency solution as proposed in [96] to address the frequency gap. In particular, rather than transmitting a single frequency as

common in current RFID protocol, TiFi transmits two single-tone CWs at two frequencies: *primary* frequency f_r (e.g., 840 Mhz which falls within the legitimate UHF RFID band) and *secondary* frequency f_w (e.g., 828 MHz, which is set to one of the 14 frequencies we desire) to decouple RFID and WiFi communications. TiFi uses f_r to power up tags and drive the inventory procedure, and f_w to stimulate tags to reflect WiFi packets. Two frequencies are backscattered by tags respectively. Adopting two frequencies brings two main benefits: first, EPC Gen2 protocol requires the reader to hop every hundreds of milliseconds. The hopping of the primary frequency does not affect the WiFi transmission at the secondary frequency at all; second, as shown in Fig. 4-5, the tag has the maximum SNR at the legitimate RFID spectrum n(e.g., 840 $\sim$ 928 MHz). The dual-frequency solution does not worsen the performance of RFID systems.

Unlike the solution in [96], which requires two independent transmitters, we notice that our two frequencies get so close (i.e., < 25 MHz for some channels) that a single transmitter is sufficient. For example, the USRP N210 with SBX daughterboard has an instantaneous bandwidth of 40 MHz and the center frequency of current readers can be varied for more than 40 MHz in order to accommodate the differences in regulations on the UHF band across regions and countries [96]. Therefore, despite the use of dual frequencies, TiFi's design will unlikely require hardware upgrade. In addition, f_w is outside the legitimate RFID band, its power must be considerably lower than that of f_r to comply with FCC regulations. We refer to [96] for details. This setting will enable the emulated WiFi signals attenuate fast, and thus can be recognized by WiFi receiver in a relatively shorter range, e.g., dozens of centimeters. This setting is exactly what we want for the near-field identification, namely, to control the signals within a small region.

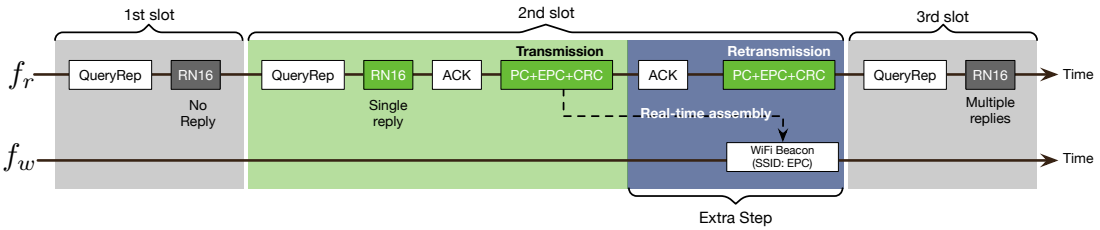


Figure 4-6: Illustration of TiFi procedure. This toy example shows three time slots in terms of two frequencies where no tag replies at the first slot, multiple tags reply in the third slot, and a single tag replies in the second slot successfully. The box highlighted in dark blue during the second slot is the extra step (i.e., retransmission) that TiFi inserts into the EPC Gen2 standard procedure.

4.5 Cross-Protocol Communication

We cannot control tags or WiFi receivers, so we have to craft the CW at f_w such that WiFi beacons transmitted at f_w are harmonically backscattered by tags to a WiFi channel at $3f_w$. In this process, the reader plays double roles: acting as an RFID reader to transmit RFID commands (e.g., `Query`) and acting as an assistant to generate WiFi packets (e.g., beacons). To better understand our design, we illustrate the high-level procedure of TiFi in Fig. 4-6. TiFi integrates the advertisement of WiFi beacons into the EPC Gen2 protocol seamlessly, by initiating a retransmission after a tag transmits its long reply successfully. The beacon is assembled with the `EPC` acquired in the first transmission. TiFi reader transmits the assembled WiFi beacon at the frequency f_w exactly during the retransmission. In this way, the WiFi beacon can be harmonically backscattered by the *right* tag to WiFi receivers. From the high-level, it seems that this tag broadcast its WiFi beacon at 2.4GHz like a normal AP. The following section introduces how the reader integrates two different protocols with this procedure in terms of the PHY and MAC layer.

4.5.1 Transmission Background of WiFi and RFID

Although many versions of WiFi protocols are available, our design only targets at an early version, 802.11b, which is far enough to meet our demand, that is, broadcasting a short 96-bit `EPC`. Current WiFi chipsets are all backward com-

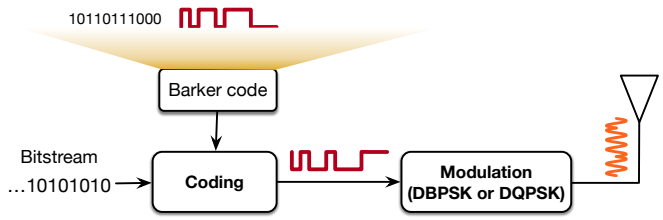


Figure 4-7: How a WiFi transmitter works

patible with this version. More importantly, the modulation of 802.11b is based on PSK, which doest not conflict with the OOK of RFID. In terms of RFID, our design targets at the EPCglobal Gen2 air protocol, which has been adopted world widely. We begin our design by briefly introducing these two protocols as follows.

WiFi Transmission. 802.11b is a set of WiFi physical layer specifications that use spread spectrum modulation. 802.11b PHY uses DBPSK/DQPSK at the physical layer and achieves 4 bit rates using different spreading codes, i.e., DSSS or CCK. We focus on the DSSS, which enables 1Mbps or 2Mbps transmission depending on the modulation. Fig. 4-7 shows how a WiFi transmitter operates. To improve the reliability, 802.11b uses pseudo-noise codes to spread the spectrum. It XORs each data bit with a Barker sequence (i.e., 10110111000), which is generated at a data rate of 11Mbps, achieving a spread spectrum of over 22MHz. In particular, data bit '0' and '1' are converted to '10110111000' and '01001000111', respectively. Each of these coded bits is then modulated onto the carrier using DBPSK or DQPSK, which offers 1Mbps or 2Mbps transmissions. DBPSK modulation carries '0' and '1' by changing the phase to either 0 or π, whereas DQPSK carries a pair of bits by changing the phase to one of $\{0, \pi/2, \pi, 3\pi/2\}$.

RFID Transmissions. The EPC Gen2 requires a reader to continuously generate a high-power CW. Two links are involved. The first link is data transmission from reader to tag, which is often called *downlink* transmission. The second link is the opposite, which is called as *uplink* transmission. Both links modulate data by OOK (i.e., changing the amplitudes of the CW), but involve

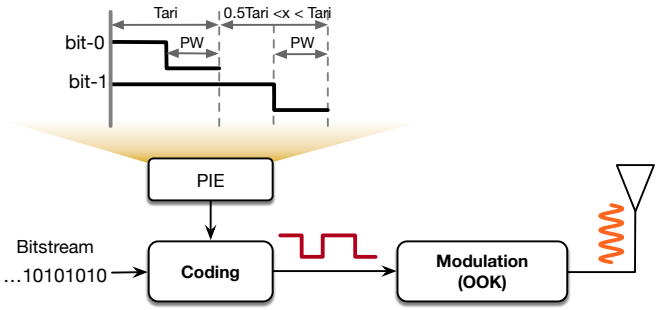


Figure 4-8: How a reader transmitter works

different channel coding methods.

- *Downlink*: A reader encodes the data bits using Pulse-Interval Encoding (PIE) for the downlink. As Fig. 4-8 shows, the PIE coding has three user-defined parameters: `Tari`, `PW` (Pulse Width) and `X`, where `Tari` is the reference interval for the downlink signaling. The duration of a bit '0' should be between $6.25\mu s \sim 25\mu s$. `PW` indicates the time duration of the lower edge, which can be set to a value between `0.265Tari` and `0.525Tari` but is capped at $2\mu s$. The duration of the bit '1' is `X`-μs longer than that of bit '0' and `X` must be between `0.5Tari` and `Tari`. The downlink rate varies between 27Kbps and 128Kbps with respect to the parameter choice.

- *Uplink*: A tag uses either FM0 or Miller coding [97], both of which are highly similar to PIE. We omit the coding details because tags cannot be controled in our design. Without traditional transceivers, tags adopt backscattering based modulation: transmitting a "1" bit by changing the impedance on their antennas to reflect the reader's signal; and a "0" bit by remaining in their initial silent state [98], as aforementioned.

Challenges: A TiFi reader is required to transmit RFID or WiFi packets with a *single* transmitter. Thus, *sharing the baseband processing is the heart of our design.* After comparing two types of transmissions, we find that achieving this goal is hindered by three main challenges. First, the data rate of an RFID system is limited to hundreds of Kpbs, whereas that of 802.11b is up to 2 Mbps. Second,

the reader uses amplitude-shift keying (i.e., OOK) , whereas WiFi uses phase-shift keying (either DBPSK or DQPSK). Third, the battery-free tags cannot hear from one another, hence, the reader is responsible for performing carrier sense on behalf of tags when conducting the cross-technology communication. In respond to these challenges, we elaborate the design of a TiFi reader from the physical layer to the mac layer subsequently.

4.5.2 PHY: Creating RFID and WiFi Packets

In TiFi, the baseband receives data from upper layer where each bit is labeled with "R" or "W", which represents the RFID or WiFi data respectively. The baseband encodes the input bits based on their labels. The coded bits are first modulated onto two intermediate frequencies respectively. The combination of the two bit streams are finally modulated onto the RF carrier.

4.5.2.1 Dual Coding Solution

As aforementioned, EPC Gen2 allows user to define three PIE interval parameters. For simplicity, we set `Tari`, `PW` and `X` to $25\mu s$, $2\mu s$ and $25\mu s$ as an example setting, leading to $25\mu s$ bit zeros and $50\mu s$ bit ones. In particular, the bit zero flips from high-voltage (HV) to low voltage (LV) at the $23\mu s$ and the bit one flips at the $48\mu s$ are expressed as follows:

$$\begin{cases} \text{PIE bit zero} = 23\mu s \text{ HV} + 2\mu s \text{ LV} \\ \text{PIE bit one} = 48\mu s \text{ HV} + 2\mu s \text{ LV} \end{cases} \tag{4.3}$$

Other settings are permissible provided that they are in compliant with EPC Gen2. Here, we use this setting as an example only.

To bridge the rate mismatch, existing CTCs typically use the high-rate transmission to emulate the low-rate transmission [39]. We adopt a similar idea, that is, using the WiFi transmission to emulate the RFID transmission. In particular, TiFi reader spreads the incoming bits based on their labels as follows:

- **[WiFi]**: If the label is "W", then the reader directly spreads it with the

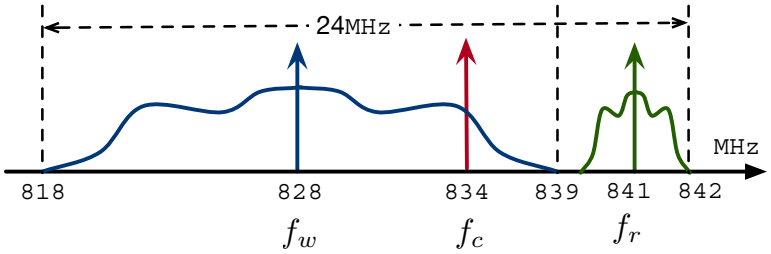


Figure 4-9: Dual modulation. Both WiFi and RFID data are carried onto a single RF carrier frequency.

Barker sequence directly. Each code bit has an interval of $1/11\mu s$.

- [**RFID**]: If the label is "R", then the reader uses a stream of constant "1"s or "0"s to emulate the PIE coding, which is equivalent to spreading the RFID bits with special spread codes. As shown in Eqn. 4.3, the PIE bit zero has 23μ HV and $2\mu s$ LV. Therefore, the reader spreads the incoming bit zero by $23 \times 11 = 253$ code "1"s plus $2 \times 11 = 22$ code "0"s. Therefore, the reader spreads the incoming bit "0' by $48 \times 11 = 528$ code "1"s plus $2 \times 11 = 22$ code "0"s.

Although no data originates from the upper layer, the reader must still maintain a high-level CW at f_r to keep the tags alive. In this case, the reader must mimic a series of meaningless bit "1"s. In addition, the WiFi beacon packets are always transmitted only during when tag is backscattering. At that moment, the reader maintains a single-tone at f_r.

4.5.2.2 Dual Keying Schemes

Generally, a carrier signal can be written as $(I(t) + jQ(t))e^{j2\pi f_c t}$ where f_c is the center frequency and $I(t)$ and $Q(t)$ correspond to the in-phase and quadrature-phase components of the coded WiFi or RFID data, respectively. *Keying* is the process of translating the input coded bits into a pair of (I, Q). The two types of packets use two different keying schemes:

- [**WiFi**]: 802.11b modulates the coded data using either DBPSK or DQPSK. Both schemes change the phase of the carrier signal to represent different

bits. To do so, TiFi shifts the carrier by one of the four distinct phases: 0, $\pi/2$, π and $3\pi/2$. In particular, (1) DBPSK: the coded "1" and "0" bits are translated to the IQ pairs, $(1, 0)$ and $(-1, 1)$, respectively. This results in two possible carrier signals: 1 and $e^{j\pi}$. (2) DQPSK: two consecutive bits are translated to one of the four IQ pairs: $\{(1, 0), (0, 1), (-1, 0), (0, -1)\}$. This process results in four possible carrier signals, 1, $e^{j\frac{\pi}{2}}$, $e^{j\pi}$ and $e^{j\frac{3\pi}{2}}$.

- [**RFID**]: The reader uses the OOK to modulate coded data. OOK is the simplest form of amplitude-shifting keying that represents digital data at the presence or absence of an RF carrier. In this case, the coded one and zero bits are translated to $(1, 0)$ and $(0, 0)$ respectively. This process results in a constant zero $Q(t)$, and $I(t)$ of 1 or 0, which correspond to a high or low voltage at the tags.

4.5.2.3 Dual Modulation Solutions

Modulation is the processing of moving the data onto the RF carrier and further propagating it in the air. The reader transmits data at two carrier frequencies (i.e., f_r and f_w) for the RFID and WiFi transmission, respectively. The naive approach is to adopt two transmitters (such as those used in [96]). An RF transceiver typically has over 40 MHz instantaneous bandwidth (e.g., USRP SBX Daughterboard), whereas the frequency difference between f_w and f_r is less than 40 MHz. Therefore, one RF transceiver is sufficient to simultaneously send data at two central frequencies. In this regard, we perform an *intermediate modulation* before modulating them onto the UHF carrier. In particular, the two intermediate frequencies (denoted by f_I^w and f_I^r) are set to $f_I^w = (f_w - f_r)/2$ and $f_I^r = (f_r - f_w)/2$ for WiFi and RFID data stream, respectively. Meanwhile, the final central carrier frequency is set to $f_c = (f_r + f_w)/2$. Consequently, the final carrier signal

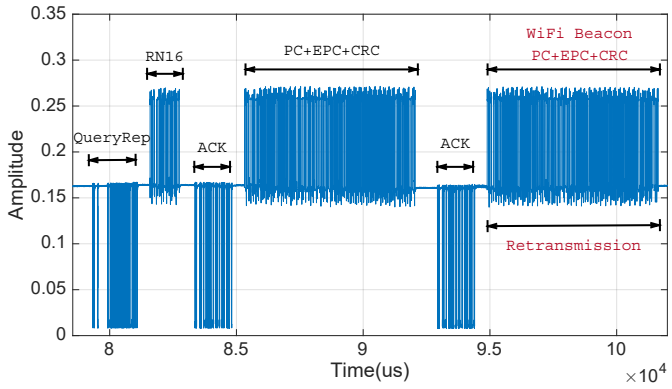


Figure 4-10: Baseband signals acquired by a TiFi reader. After decoding the long reply of the tag, TiFi reader riggers a retransmission, during which the WiFi beacon assembled with the EPC is transmitted such that it can be harmonically backscattered by the tag at a WiFi channel.

is expressed as:

$$\begin{aligned}
&\left\{[I^w(t)+jQ^w(t)]e^{j2\pi f_I^w t}+[I^r(t)+jQ^r(t)]e^{j2\pi f_I^r t}\right\}e^{j2\pi f_c t}\\
=&[I^w(t)+jQ^w(t)]e^{j2\pi(f_I^w+f_c)t}+[I^r(t)+jQ^r(t)]e^{j2\pi(f_I^r+f_c)t}\\
=&[I^w(t)+jQ^w(t)]e^{j2\pi f_w t}+[I^r(t)+jQ^r(t)]e^{j2\pi f_r t}
\end{aligned} \tag{4.4}$$

where $f_I^w+f_c=(f_w-f_r)/2+(f_r+f_w)/2=f_w$ and $f_I^r+f_c=(f_r-f_w)/2+(f_r+f_w)/2=f_r$ because of the careful design of the two intermediate frequencies. The modulated results are transformed into two central frequencies: f_w and f_r.

In addition, we also study if the two types of data interfere with each other in terms of the bandwidth. In particular, WiFi and RFID have approximately 22MHz and 2MHz respectively, as shown in Fig. 4-9. It is easy to find that a blank of 1MHz still remains even if the highest WiFi channel is targeted (i.e., 828MHz). Actually, when the reader transmits WiFi packets, only a single tone at the RFID band exists for keeping tags alive. Thus, both types of transmission do never interfere with each other in any manner.

4.5.3 MAC: Backscattering WiFi Packets

The design of MAC layer is to answer the question: *when does the reader transmit WiFi beacons*? The integration timing must meet three rigorous prerequisites.

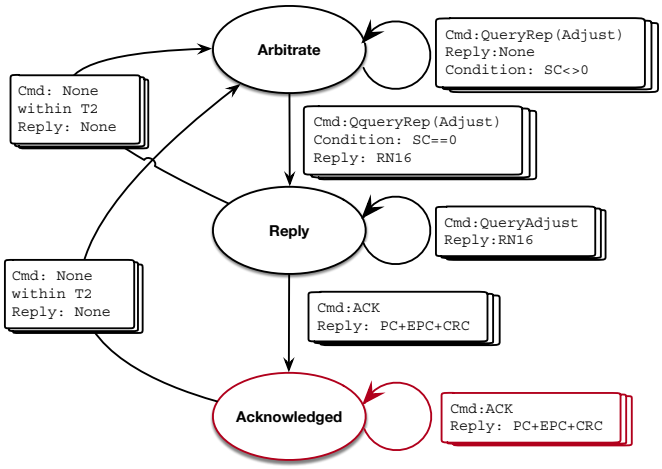


Figure 4-11: A simplified tag state diagram. Each tag should stay in one of three states: 'Arbitrate', 'Reply' and 'Acknowledged'.

First, the reader should acquire the tag's `EPC` already, because TiFi reader uses the `EPC` as the `SSID` and MAC address of the beacon. Second, the beacon packet must be transmitted exactly during the period when the tag is backscattering, because the WiFi receiver can sense the harmonic backscattering only at this moment. Third, only a single tag is allowed to backscatter WiFi beacons at any given moment; otherwise the mobile device is completely unaware of which tag is transmitting.

Background of Retransmission. Fig. 4-11 shows the state machine of a tag. It could be at three states: "Arbitrate", "Reply" and "Acknowledged". Suppose the tag is ready to transmit its `EPC` (i.e., at "Arbitrate" state). It firstly sends an `RN16` reply that contains a 16-bit random number for anti-collision and transmits to "Reply" state, after receiving the `Query` or `QueryAdjust` commands. As an acknowledgement, the reader sends back an `ACK` command. The acknowledged tag then transmits a *long reply* including its `EPC` and transits to the state of "Acknowledged". The Gen2 protocol specifies that the reader is allowed to retransmit the `ACK` command for a retransmission of the long reply, when the tag is at "Acknowledged" state. Note that the retransmission request must strictly follow after the last one in less than 20×`Tari`.

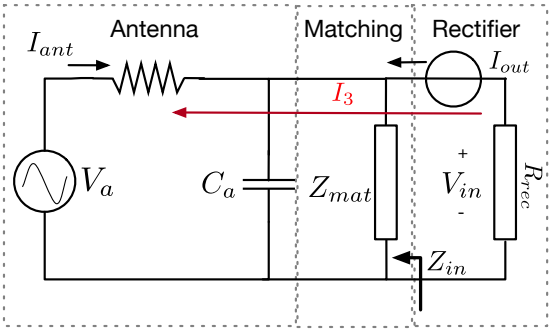


Figure 4-12: Simplified equivalent circuit of a tag's antenna [1].

4.5.3.1 Timing of Beacon Broadcast

Tags can actually reflect the incoming wireless signals even at the non-backscattering state, during which the rectenna is still working. We choose the backscatter duration as the timing of transmitting WiFi beacons for two MAC demands: (i) first, the transmission of WiFi beacon must be initiated *immediately after* the tag transmitted its `EPC` for two reasons. On one hand, the beacon cannot be broadcasted before the corresponding tag is identified (especially for the new arriving tags) because TiFi reader uses the `EPC` as the `SSID` and `MAC` address of the WiFi beacon; on the other hand, the *deferred broadcast* of WiFi beacons corresponding to the tags identified long time ago may produce the ghost tags, which have left the scanning field but still appear in the result UI. Thus, the best timing is exactly the duration immediately after the tag is identified; (ii) second, broadcasting WiFi beacons would introduce serious interference or jamming to other tags' transmissions. Choosing the timing of when a tag re-transmits its `EPC` can avoid any interference because the whole RFID system guarantees that only a single tag exclusively occupies the current slot.

In addition, we also find that the strength of third-order harmonics during the reflective sate is much higher than that during the non-reflective state. This feature could be validated from both the theoretical and experimental aspects:

- **Theoretical Analysis**. The equivalent circuit of a tag's antenna is shown in Fig. 4-12. When the tag receives the reader signal, the signal will first pass the matching module and be adjusted by the rectifier. The rectifier can

be modelled as a current source and generate the harmonics. Consequently, We can model the power of radiated third-order harmonic component P_3 as the function of the power of the fundamental component P_1 as follows.

$$P_3 = \frac{I_3^2 P_1 (Q^2 + 1)}{I_1^2 (64Q^4 - 7Q^2 + 1)} \tag{4.5}$$

where Q is the quality factor of the RF frontend and I_3 is the harmonic current. This model is proposed in [91]. From the Eqn. 4.5, we know that harmonics P_3 is proportional to fundamental component P_1; and P_3 approaches to the P_1 when the Q decreases [91]. As required by the Gen2 protocol [97], the value of Q is much smaller at reflective state than at non-reflective state. In other words, the power of third-order harmonics at reflective state far higher than that at non-reflective state.

- **Experimental verification.** To verify the above analysis, we use a commercial ImpinJ reader to perform the continuous reading at 920MHz. Meanwhile, we use an USRP to sniffer the harmonic band with 5 MHz bandwidth. According to our experiments, the strength of harmonic backscattering signal at reflective state is 6 dBm higher than that at non-reflective state.

In summary, the best timing to transmit the beacon packet is the duration of retransmission.

4.5.3.2 The Integration Timing

Our design integrates the WiFi broadcast into the *retransmission* as follows.

1. The reader initiates a reading session by broadcasting a `Query` or a `QueryAdjust` command; the tag transmits its `RN16` reply.

2. As a response, the reader sends back an `ACK` command to request the tag's long reply; the tag transmits the long reply including `EPC`.

3. After decoding the long reply, the reader immediately resends the previous `ACK` to the tag for triggering a retransmission of its long reply; the tag

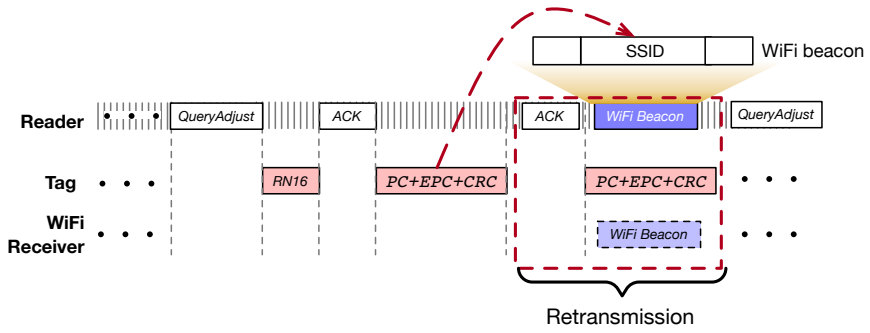


Figure 4-13: Reflection of WiFi beacons. TiFi broadcasts the WiFi beacons during the tag's retransmission.

transmits its long reply again.

4. During the retransmission, the reader transmits the WiFi beacon, whose `SSID` and MAC fields are assembled with the `EPC` decoded previously.

The above procedure is illustrated in Fig. 4-13. The only difference with the existing procedure is that TiFi adds a retransmission phase during which the WiFi packets are transmitted. Evidently, selecting a tag's retransmission as the timing to transmit its WiFi beacon efficiently meets the above timing constraints. Fig. 4-10 presents an example of the baseband signals that follow the aforementioned procedure, as acquired by the TiFi reader. The final effect appears similar to a tag backscatters its `EPC` and WiFi beacon simultaneously.

4.5.3.3 Demodulation

For clarity, we extract the signal modulated at the third-order harmonics as follows:

$$S_w(t) = e^{j2\pi 3 f_w t + \phi(t)} \tag{4.6}$$

Since WiFi beacons are modulated by changing the phases (i.e., BPSK), the $\phi(t)$ would change between two values (e.g., 0 and π), which represent the bit 0 and 1 respectively. Meanwhile, due to the OOK modulation of tags, the reflected signals from a tag becomes as below:

$$S'_w(t) = a(t)e^{j2\pi 3 f_w t + \phi(t)} \tag{4.7}$$

where $a(t)$ represents the ON or OFF states of the tag (i.e., OOK). In PHY layer, the tag changes $a(t)$ by by changing the reflection coefficient of its RF frontend no matter the packet is encoded with FM0 or Miller schemes in downlink. As a result, $S'(t)$ is a APSK (Amplitude and phase-shift keying) signal. At the receiver side, a WiFi receiver firstly conducts the downconversion via multiplying $S'_w(t)$ by the carrier of $e^{-j3f_w t}$ as follows:

$$S''_w(t) = S'(t)e^{-j2\pi 3f_w t} = a(t)e^{\theta(t)} \tag{4.8}$$

Since 802.11b uses the PSK, the receiver only detects the angles of $S''_w(t)$, namely, $\angle S''_w(t = \theta(t)$ for the decoding where the amplitude changes caused by $a(t)$ is ignored.

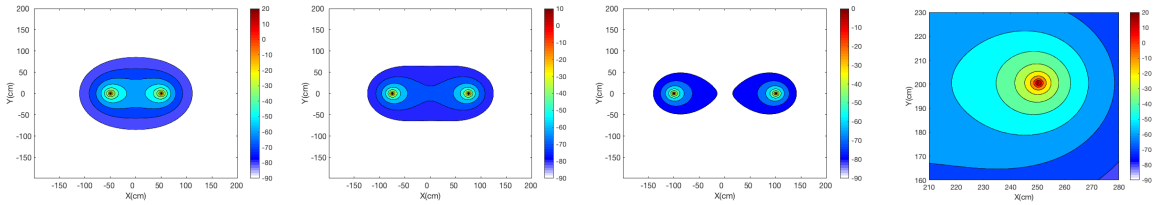
4.5.4 Practical Discussion

We elicit some practical concerns to discuss about the cross-technology communication procedure.

• *Is the time is sufficient to transmit an entire WiFi beacon during retransmission?* A long reply of a tag contains 128 bits, which are encoded via FM0 or Miller. When a general RFID setting is considered, a tag should take at least $128 \times 25\mu s = 3.2ms$ to backscatter the long reply. An 802.11b beacon contains 576 bits and takes $576 \times 1\mu s = 0.576ms$ on transmission. Therefore, a $3.2ms$ backscattering window is sufficiently large to transmit a 0.576ms WiFi beacon 5 times.

• *How frequently are WiFi beacons broadcasted?* The reader is typically configured to continuously and repeatedly scan tags round by round. Suppose that n tags exist. Each tag is queried and thereby generates a WiFi transmission every $(1/ne\ln(n))$ seconds [99]. For example, if 100 tags are available, then the WiFi beacons are broadcasted every $1.8ms$, which is considerably more frequent that the default $100ms$ setting of real WiFi APs.

• *How does TiFi deal with multiple tags?* TiFi does not need extra efforts to



(a) 100**cm** **(b)** 150**cm** **(c)** 200**cm** **(d) Zoomed into**

Figure 4-14: The theoretical coverage heat maps. (a) (b) shows the maps when the distance is set to 100cm, 150cm and 200cm respectively; (d) zoomed into the surrounding region of the WiFi receiver at the distance of 200cm.

deal with multiple tags but uses the existing EPC Gen2 Q-adaptive algorithm for anti-collision. As shown in Fig. 4-6, TiFi reader inserts the harmonic backscattering after when a tag is successfully identified, and does not act if the slot is empty or collided.

4.6 Near-Field Identification with TiFi

Suppose that the signal strength output from the reader is P_r. Then, the signal strength at the WiFi receiver, P_w, can be modeled using Friis path loss [56] as follows:

$$P_w = \left(\frac{P_r G_r}{4\pi d_1^2}\right)\left(\frac{\lambda_r^2 G_t^2}{4\pi}\frac{|\Delta\Gamma|^2}{4}\alpha_w\right)\left(\frac{1}{4\pi d_2^2}\frac{\lambda_w^2 G_w}{4\pi}\right) \tag{4.9}$$

The preceding equation describes two signal propagations: reader $\rightarrow$ tag $\rightarrow$ WiFi receiver. G_r, G_t and G_w are the antenna gains of the reader, the tag and the WiFi receiver, respectively. d_1 and d_2 are the distances among the three components. λ_r and λ_w are the wavelengths of the RFID and WiFi signals. $|\Delta\Gamma|^2$ is the backscatter coefficient. α_w indicates the loss in energy loss in the desired harmonics.

Suppose that P_r, G_r, G_t, and G_w are set to 30 dBm, 8 dBi, 2 dBi, and 0 dBi, respectively. Then, $|\Delta\Gamma|^2$ and α_w are around 1.1 dB and 3.3 dB, respectively. We simulate the power heat maps with respect to tags' locations in Fig 4-14, where each point indicate the power level that the WiFi receiver would receive if the tag is placed at the location. Suppose the WiFi receiver' sensitivity is -90dBm,

After simulation, two key points are obtained.

- The received power increases when the tag gets close to either the WiFi receiver or the reader, because maximizing the signal strength requires minimizing the product of $d_1 d_2$.

- The effective coverage range of TiFi is approximately $2m$, and power decreases by about 1dBm each time the tag is moved 10cm away from the WiFi receiver. This condition perfectly fits the demand of near-field identification, such as NFC, which requires distinguishing objects' locations via signal strength within a small area.

The above theoretical analysis shows that TiFi can work for NFI, which eliminates the pain point that the UHF reader cannot distinguish two close tags from locations.

4.7 Implementation

We implement a prototype of TiFi using a USRP N210 software radio and test it on a variety of commercial RFID tags and WiFi receivers.

• **TiFi Reader.** We implement the reader on a USRP N210, which is equipped with an SBX Daughterboard and an 8dBi directional antenna. Fig. 4-3 shows the implementation compared with that of a commercial reader (i.e., ImpinJ R420). In particular, UHD gain and baseband signal amplitude are set to 15dB and 0.5, respectively. The sampling rate is set to 25 MHz. The implemented prototype is built and executed on GNU Radio 3.7.10 under Linux Ubuntu 16.04 LTS operating system powered by an Intel Core i5-7200U processor, clocked at 2.50 GHz and a RAM memory of 12 GB.

• **Harmonic Suppression in Reader-Side**. The harmonics caused by the nonlinearity effect are not unique to RFID tags, but widely exist among electronic amplifiers. The commercial RFID readers should manage the harmonics to avoid RF interferences to other devices. We measure the signal spectrums of the TiFi

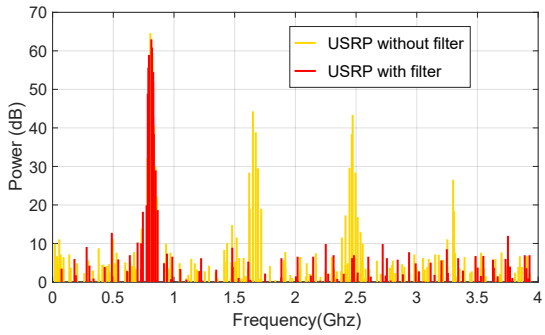


Figure 4-15: Harmonic signals from USRP reader. We add an elliptical lowpass filter to suppress the 1.6G and 2.4G harmonics leaked from the USRP devices.

reader with respect to its harmonic leakage. Fig. 4-15 illustrates the spectrum of the reader when it broadcasts query command without tag reply. We can observe that so many harmonics are leaked from the TiFi reader (without filter). This is because the reader is based on the USRP device, which is not optimized for commercial use. To block the harmonics from the reader, we customize an RF *elliptical lowpass filter*, i.e., a hardware component for harmonic suppression. The filter is designed to have a cutoff frequency of 1 GHz and 0.5 dB in-band ripple. Its restraint outside of the band at 2.4 GHz is over 50 dB. As Fig. 4-3 shows, the filter is used to bridge the reader and the antenna. Revisiting Fig. 4-15, we can see that the harmonics of the TiFi reader equipped with the filter are reduced exactly to the noise level (< 10 dB) by the filter. We use the filter in our experiments to ensure that tags are the sole devices that emit harmonic signals.

• **Commercial RFID tags**. Unless noted otherwise, our experiments are performed with the most widely deployed type: ImpinJ Monza 4 QT [15]. To demonstrate the generality of our technique, we also test 6 types of commercial tags, which are produced by two different manufactures, as listed in Table. 4.2. Each of these tags costs $5 - 10$ cents.

• **Commercial WiFi Receiver**. We choose a High-Definition Oscilloscope MSOS404A from KeySight. Inc to collect the raw signal and absolute signal strength. iOS does not provide the low-level APIs for us to measure the detailed PHY parameters such as RSSI, delay, phase, error rate, and so on. Thus, we employ the Android platform instead for the experiments to better evaluate the

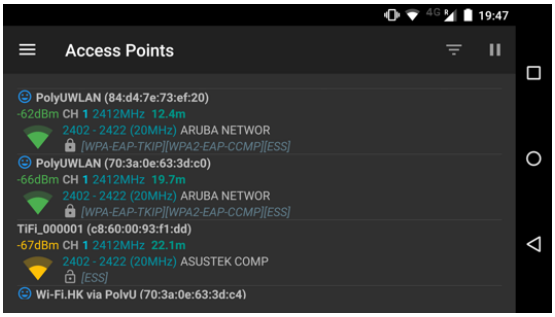


Figure 4-16: Snapshot of the WiFi analyzer app. `TiFi_XXX` corresponds to TiFi beacons.

performance of TiFi. However, TiFi can still work for iOS devices. We use the Huawei P10 (equipped with the WiFi chipset of Broadcom BCM43596) as our default WiFi receiver, and test the diversity across seven types of commercial tags. To measure the RSS of our Wi-Fi signal, we use an Android APP called Wi-Fi analyzer [100] as shown in Fig. 4-16. Unfortunately, Apple blocks the API of Wi-Fi low-level data in IOS system and we can only choose the smart phone with Android system. We implement a Wi-Fi baseband decoder by Matlab to measure the packet error rate.

4.8 Results

We deploy the TiFi reader and test tags with a distance of $2m$ by default in our office. The tag is placed at the position of $50cm$ away from the receiver. Unless otherwise noted, the WiFi and RFID carrier frequencies are configured to 828MHz and 841MHz by default. The UHD gain is set to 20dB. In our experiments, we number the `EPC`s of tags from one. The WiFi beacon packets have a payload of 68 bytes where the SSID is set to the form of `TiFi_XXX` and `XXX` indicates the last few bits of tag's `EPC`. We report the results of our experimental evaluation in this section.

4.8.1 Comparison to Sate-of-the-Art

For comparison, we profile existing related radio technologies in Table. 4.3 with respect to the dimensions of frequency and max range. The comparisons show

#	Technology	Max Range	Frequency
1	TiFi	2 m	2.4 GHz
2	USRP Reader [101]	3m	840 MHz
3	TSL-1128 Reader [83]	1.5m	840 MHz
4	ALR-S350 Reader [84]	1.5m	840 MHz
5	Phychips Reader [82]	0.2m	840 MHz
6	R420 Reader [102]	10 m	840 MHz
7	NFC (HF RFID) [103]	0.9m	13.56 MHz
8	Passive WiFi [56]	15.24m	2.4 GHz
9	FM Backscatter [52]	18.29m	92.1 MHz
10	Ambient Backscatter [50]	1.8m	539 MHz
11	Lora Backscatter [87]	2.8km	900 MHz
12	HitchHike [88]	54m	2.4 GHz
13	BackFi [49]	7m	2.4 GHz
14	Interscatter [20]	27.4m	2.4 GHz

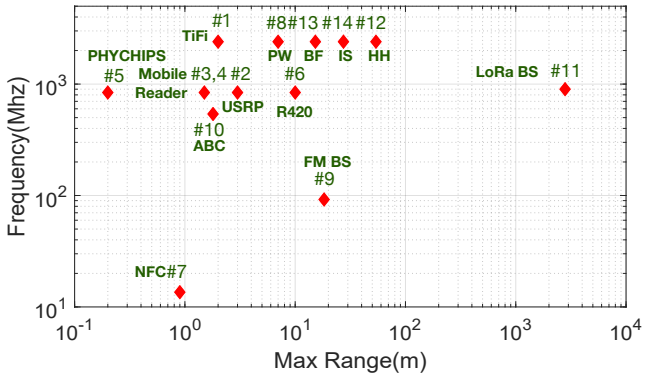


Table 4.3: Comparison to the Sate-of-the-Art

the following findings.

- **USRP Reader**: The USRP reader is implemented using the Open Project [101], which decodes EPCs of tags by using USRP at RFID frequency band. TiFi achieves nearly the same range as of the USRP reader (3m). This finding appears "surprising" because the USRP reader uses the 1st-order signal for the communication, which is 50 $\sim$ 60dB stronger than the 3rd-order harmonic backscattering used by TiFi. This can be explained by the sensitivity. The USRP has a 30dB weaker sensitivity than a mobile phone. It is a little hard for USRP reader to resolve the weak signals of below -60dBm. By contrary, current mobile phone has a quite sensitive transceiver, which can easily deal with signals of down to -90dBm. This ability extends the detection range of TiFi compared with USRP reader.

- **Mobile Reader**: The commercial mobile RFID readers (e.g., TSL-1128 [83] and ALR-S350 UHF readers) have mean ranges of 1.5m, which is even 0.5m shorter than TiFi's. This result show that TiFi can be exactly employed as good substitutes for specialized RFID mobile readers.

- **HTTP Reader**: TiFi achieves the one-fifth of the range of a commercial HTTP reader (e.g., R420 [102]). Unlike USRP reader, these commercial readers have good RF sensitivities as mobile phones. Their 1st-order communications are 60dB-higher than the 3rd-order communications, as listed in Table. 4.2.

- **HF-NFC**: The HF-NFC operates at the lowest frequency (13.56MHz), which determines its inductive coupling-based communication approach. The energy attenuation of this method is proportional to the cube of the distance. Therefore, the range of HF-NFC is upper bound at approximately 1m. We use the built-in NFC reader of the Moto X+1 phone for the test.

- **Backscatters**: Backscatters typically have a long communication range ($>$ 10m) because they are equipped with large capacitors, which absorb a considerable amount of energy from the air. In particular, Lora backscatter can achieve a maximum range of up to 2.8Km. In contrast to backscatters, TiFi is not required to modify tags (i.e., backscatters).

In summary, TiFi operates at the WiFi spectrum, similar to most of the backscatters, but performs comparably to the commercial mobile readers and better than HF-NFC in terms of the maximal range.

4.8.2 Characterizing TiFi's NFI

We evaluate the RSSI as a function of distance with different transmitting gains (i.e., a parameter of UHD). In the experiment, we hold the mobile phone in our hand and measure the reported RSSI values by moving the phone away from the test tag under a specific gain setting. Measurements are taken at increments of $10cm$. Fig. 4-17 presents the results of four gain settings, from which we can

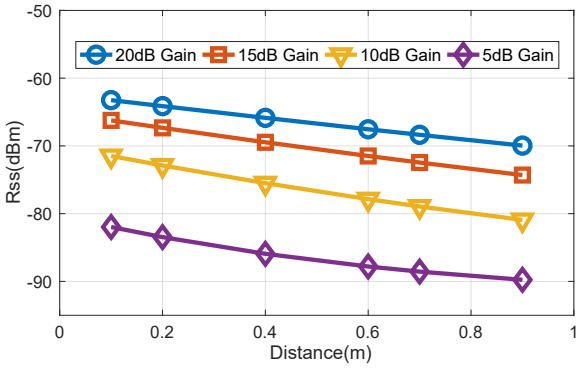


Figure 4-17: RSSI

observe that TiFi exhibits a good quality of linearity where the RSSI decreases as the distance increases. The finding indicates that RSSI has over 1 dBm changes when moving the mobile phone 10cm away from the tag. Note that RSSI is in the unit of dBm, which takes logarithm of the received power (defined in 4.9). This characteristic is derived from the backscattering communication, which is hypersensitive to the distance. It provides the experimental basis for NFI or proximity localization that the tag nearest to the WiFi receiver achieves the strongest RSSI.

A higher gain setting allows the transmitter to acquire stronger power, thereby extending the reading range. This finding is confirmed by our experimental results. The maximum gain of USRP N210 is up to 31dB. However, As reported in [104], when going beyond a 20dB gain for frequencies below 1.5GHz, USRP device exhibits a severe distortion on harmonics, even given a single tone wave waveform. This hardware defect constrains our experiment results. We believe the commercialized TiFi with customized hardware components would have longer range.

In summary, leveraging the signal strength of WiFi beacon backscattered from tags for near-field identification is completely feasible and effective.

4.8.3 Coexistence with WiFi and RFID Devices

4.8.3.1 Coexistence with other WiFi

One major concern might be about coexistence between virtual APs and nearby WiFi devices. We consider the coexistence from two perspectives:

• **Impact of TiFi on WiFi**. It can be seen from Fig. 4-17 that the power of our virtual AP is below −60dBm even when the gain of USRP is set to the maximum (i.e., 20dB). Actually, the power of harmonic backscattering is about 30dB lower than that of the fundamental signals employed in real APs. With respect to the 2 m effective range, virtual APs hardly exert any interference on nearby WiFi devices. Thus, TiFi is not a threat to normal WiFi devices.

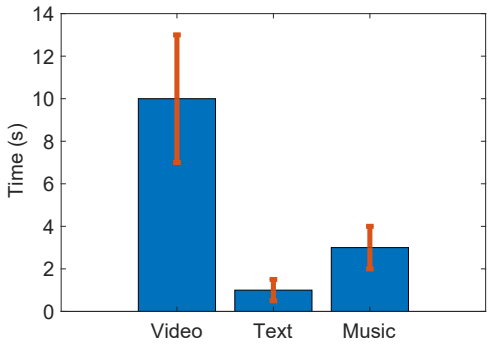


Figure 4-18: Impact from WiFi

• **Impact of WiFi on TiFi.** We firstly discuss the impact of WiFi on TiFi in different traffic scenarios. In the experiments, a real WiFi AP is placed half a meter away from a cluster consisting of 10 tags. Meanwhile, we connect a mobile phone to the real AP for acquiring the video or music streaming and text messages respectively. Both the real AP and virtual APs are set to the same WiFi channel. Without any traffic, each tag as a virtual AP can be identified every 1 s on average through the TiFi. As shown in Fig. 4-18, the average identification delay is increased to 10 s, 1.1 s and 3 s respectively in the three scenarios. This fully shows that nearby WiFi devices with heavy wireless traffic indeed affect the performance of TiFi, because the weak signals of virtual and passive APs might be drown in the strong signals from real active WiFi devices. However, even

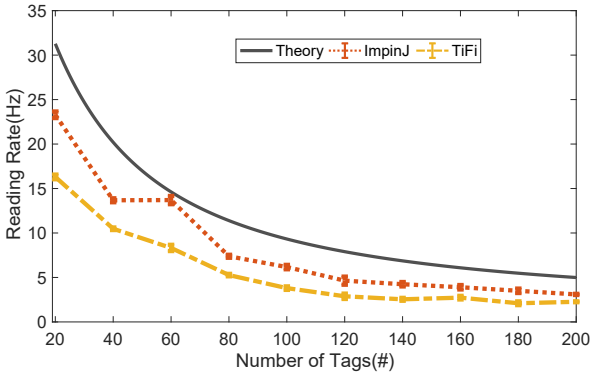


Figure 4-19: Reading Rate

though the strong interferences are present, tags can still be identified because both WiFi and TiFi employ the time-slotted random multiple access strategy. The similar case happens to the real APs, whose beacons are interfered by nearby WiFi devices sometimes. However, we can still find out them in a certain time.

In summary, WiFi devices with heavy traffic and close to tags might jam the wireless channel and further lead to significant identification delays in TiFi. However, WiFi devices at a distance (e.g., > 6) do not exert evident interference on TiFi. To avoid such potential interference , we advise TiFi to randomly hop among 14 WiFi channels every 500 ms. This duration is also the requirement of EPC Gen2 protocol in order to avoid interference with other devices. On the other hand, AP scanners are designed to scan all channels, and thus, the random hopping does not affect the discovery of virtual APs by using AP scanners.

4.8.3.2 Coexistence with other RFIDs

The mutual interference among RFID systems has been carefully considered by the current EPC Gen2 air protocol, which uses many schemes to avoid the mutual interference, such as channel hopping [97], CSMA [97], coloring-based multiple reader access [105], and so on. The design of TiFi exactly accords with the EPC Gen2 protocol. For example, TiFi reader broadcasts the WiFi packet only when the tag is acknowledged and re-transmits its `EPC`, during which only a single tag transmits the signals in the whole RFID systems. Thus, TiFi readers would not cause any interference to other tags or other readers thanks to the compatibility.

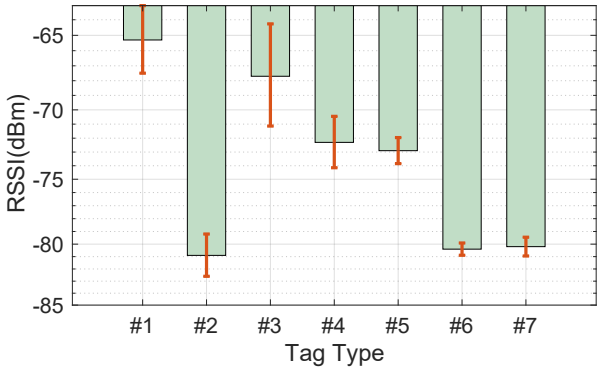


Figure 4-20: Tag diversity

4.8.4 Evaluation on Scalability

Next, we evaluate the scalability of TiFi in terms of the number of tags, i.e., how many tags does TiFi support? Since TiFi tightly integrates into Gen2 protocol, its scalability mainly depends on that of the Gen2 protocol. In theory, Gen2 protocol can read any number of tags as long as it is given enough time. In practice, we usually use a classic metric called *reading rate* (i.e., the number of reading times per second per tag) to imply the scalability.

To obtain the reading rate, TiFi reader and ImpinJ reader are configured to inventory all tags continuously and repetitively for 10 minutes in the experiments. Their average reading rates are shown in Fig. 4-19. It can be seen from the figure that the rate drops approximately linearly as the number increases. This is because more tags' participations incur more collisions and further lower the rate. The reading rate of TiFi is lower than the ImpinJ reader because each tag must transmit its *long reply* twice (see Fig. 4-10). We also plot the theoretical rate presented in [99] for reference. From the perspective of reading rate, we could present the scalability under a time constraint as follows: if each tag must advertise its WiFi beacons at least once *within one second* (i.e., 1Hz reading rate), TiFi can support about 200 tags; if advertising at least once *within* $100ms$ (i.e., 10Hz reading rate), about 42 tags are supported.

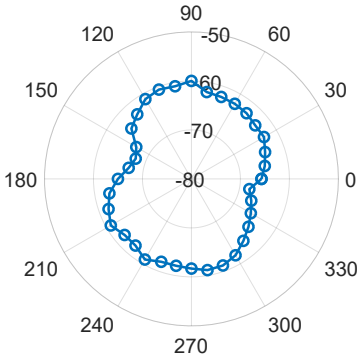


Figure 4-21: Tag diversity

4.8.5 Impacts of System Configurations

Next, we evaluate TiFi's RSSI as a function of different system configurations:

• **Type of Tag:** Fig. 4-20 shows the impact of tag type on RSSI and their hardware information are listed in Table 4.2. In the experiment, tags are placed in front of the reader antenna with a distance of $10cm$. We observe considerable differences among different types. In particular, tag#1 achieves the highest RSSI. This type tag adopts the antenna design of circular polarization, which is composed of a pair of perpendicular linear antennas. Thus, it can absorb energy or backscatter signals from various perspectives. However, although tag#2 also adopts circular polarization, its antenna area is only a half of that of tag#1 and thereby shows lower RSSI. This group of experiments suggests that one should employ a same type of tags for the proximity based localization or payment because the RSSI depends on tag types.

• **Orientation:** To evaluate the orientation's impact on harmonic backscattering, we use a smartphone to identify an Impinj H47 tag at different orientations and measure received signal strength. The result is shown in Fig. 4-21. It is shown that the strength of harmonic backscattering varies in different directions. This is because of the directivity of tag's antennas.

4.9 Related Work

We review the related works from the two fields.

• **Harmonic Backscattering.** Although the study and the exploitation of utilization of nonlinearity in diode based devices are not new, the harmonics in RFID system have only elicited attention in recent years. The harmonic phenomenon in RFIDs was reported in [58, 63–66, 68, 98, 106]. The work [67] characterized the harmonic signals in UHF RFID with extensive experiments. [58] and [69] are the most related works to ours. Similar to TiFi, [58] also explores the harmonics as a secondary communication channel. However, by virtue of harmonic backscattering, TiFi targets at talking with WiFi receiver while [58] is to enhance the communication between the reader and tags. This work [69] uses the harmonics to achieve the multi-frequency continuous wave ranging and further localize tags in 3D space. Unlike these prior work, our work utilizes the 3rd harmonics for cross-frequency communication.

• **Backscatters.** Similar to the RFID tags, backscatters are the battery-free devices that modulates data by reflecting the source signals. Dozens of backscatters have been proposed in the past five years [20, 49, 50, 52, 56, 87, 88]. The closest to our work is a recent work called Passive WiFi [56], which generates the 802.11b packets. However, Passive WiFi requires FPGA for signal processing which takes higher energy consumption. Unlike backscatters, our design is based on the commercial low-cost and lightweight RFID tags.

4.10 Discussion

This chapter presents TiFi, a novel CTM system that enables commercial WiFi receivers working at 2.4GHz to identify UHF RFID tags. Leveraging the harmonic backscattering for communication is a challenging technical problem. TiFi has taken an important step toward addressing this problem. However, the current version of TiFi still has two main limitations: first, the identification range is relatively short because of the severe harmonic attenuation and FCC spectrum constraint. Second, the identification still depends on readers, limiting to the usage scenario in practice. Third, WiFi started to employ QAM over OFDM

modulation technique as achieving higher data rate after the version of 802.11b. The QAM changes both the amplitude and phase to represent different bits. The OOK modulation at tags also affect the amplitudes of beacon packets, leading to the failure in beacon decoding at WiFi receivers. This is also a common limitation of the backscatter systems like [20, 49, 56, 87, 88], which aim to achieve the inter-communication with WiFi devices. All these work are designed to support 802.11b. The survey report from [107] suggests that almost all existing WiFi drivers are compatible with 802.11b. Thus, our solution can work for absolute majority of WiF receivers.

While there is scope for many improvements, we believe TiFi advances the state of the art in crosse-technology communication by using the harmonic backscattering of tags. The key innovation of this work involves two unique techniques, CFC and CPC - enabling tags to backscatter WiFi AP beacons without changing the hardware or firmware of RFID tags and mobile devices. This work will inspire plenty of new applications over UHF RFID systems.

Chapter 5

Activating Wireless Voice of ETC Systems with Zero Start-up Cost

Electronic toll collection (ETC) is a system enabling electronic collection of toll payments, thus allowing for near-nonstop toll collection and traffic monitoring. ETC (e.g., Z-Pass and I-PASS) has become the most successful and widespread application of Radio Frequency IDentification (RFID). Approximately, 70% to 89% of cars are equipped with ETC transponders (or tags) in the US. In particular, over 80% of Illinois' 1.4 million daily drivers currently use I-PASS [108]. ETC can eliminate delay on toll roads, HOV lanes, and toll bridge by collecting tolls without requiring cars to stop. Owing to this wide adoption, the industry and academia are looking into delivering new services via the currently deployed ETC infrastructure. Examples include paying for food at drive-through restaurants or parking lots with an e-toll transponder, tracking the number of cars at every road intersection for traffic control, and detecting and ticketing overspeeding cars without the need for car-mounted radars and hidden police officers. ETC offers an opportunity for using ETC systems to enable smart cities in the future [109].

As a typical technology of the Internet of Things, ETC systems are designed for machine-to-machine communication only (e.g., identifying cars or monitoring traffic). Drivers cannot *directly* obtain related information (e.g., charge amount,

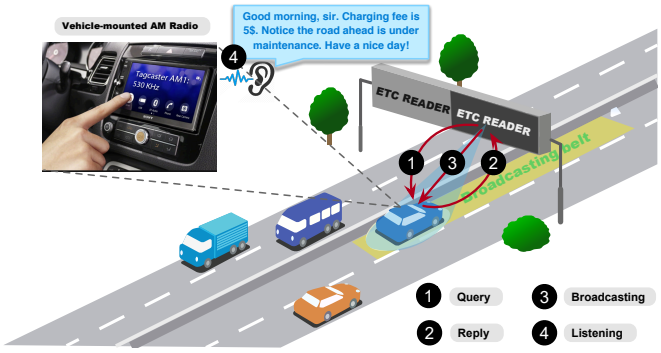


Figure 5-1: Tagcaster-enhanced wireless voice service for ETC system. ❶ The ETC reader queries the transponder attached on the vehicle. ❷ The transponder is identified through its reply. ❸ Tagcaster broadcasts the wireless voice to the identified vehicle. ❹ The driver can listen to the AM radio through the vehicle-mounted radio receiver.

credit balance, real-time traffic and road condition). Achieving direct machine-to-human (M2H) communication between drivers and ETC systems requires extra peripherals and efforts. For example, drivers must slow down their cars to view an LED screen installed at the tollbooth for charge details. The information acquired by drivers is usually quite limited due to the small screen size. Slowing down also worsens the congestion during rush hours. Facing this practical issue, we ask "Is there any convenient and user-friendly M2H communication way to provide informative interaction fast?"

In this chapter, we introduce a novel ETC service with CTM design, called Tagcaster, which supplements the function of an AM radio to the existing ETC infrastructure with zero start-up cost. Tagcaster can activate a deployed ETC tollbooth to provide the wireless voice service for direct M2H communication. Such cross-technology mutualism creates a user-friendly interface for a M2M communication system. Fig. 5-1 illustrates our service scenario. When a driver passes through a tollbooth, the ETC reader automatically identifies the vehicle's ID and retrieves its related data from backend servers. Then, the reader broadcasts the data in the form of an AM radio. The driver can listen to the broadcast using the vehicle-mounted radio receiver or smart-phone inside.

Applications: 'Wireless voice' is more user-friendly and spontaneous than

current interactive medias (e.g., LED screens or external speakers). It provides a new means to deal with the issue of poor visibility in bad weather conditions, such as stormy, foggy or smoggy, when viewing feedback from ETC screens is difficult for drivers. Such functionality is also useful in a wide range of application scenarios. Apart from charging information, Tagcaster can also broadcast greetings, real-time traffic conditions, account balance, advertisements, and so on. Moreover, drivers do not need to slow down to acquire information from an ETC system as they are pushed actively.

Challenges: The fundamental challenge in Tagcaster is in the seemingly impossible cross-technology communication between ETC RFID and AM radio due to the large frequency gap. ETC RFID systems operate at ultra-high frequency (UHF)(e.g., 800-900 MHz), whereas an AM radio works at radio frequency (e.g., 500-1700 kHz). A real AM station is usually equipped with a 60 m long antenna because the length of transmitting antenna must be close to half of the carrier wavelength. Clearly, a 16 cm long directional antenna for ETC reader fails to propagate AM radio signals into the air. Our insight is that non-ideal behavior (non-linearity effect) in the circuits of radio receivers can receive and pull the UHF signal down to the low-frequency band if the signals are transmitted via two UHF carriers. Specifically, on the transmitter side, the RFID reader broadcasts two signals at f_1 and f_2 (e.g., $f_1 = 820.5$ MHz and $f_2 = 820$ MHz) simultaneously. Given that both signals are at UHF, they can be propagated successfully by the existing UHF antenna. On the radio receiver side, a new signal is created at $|f_1 - f_2|$ (e.g., 500 kHz) due to the nonlinearity effect of the pre-amplifier at the radio receiver. The process is equivalent to performing an additional downconversion called the *zeroth downconversion* before the radio signal is further downconverted and decoded to an audio signal. Unlike traditional wisdom that regards nonlinearity as detrimental, we use it as a natural downconverter.

Solutions: Engineering a Tagcaster must address two practical issues that stem from the pursuit of zero start-up cost (i.e., without requiring modification in the hardware of the ETC system).

• *Generate two carriers.* The zeroth downconversion requires two signals from an ETC reader to operate at f_e and f_e+f_r, such that their difference of $|\cancel{f_e}+f_r-\cancel{f_e}|$ is exactly equal to the f_r that the radio receiver can process. Here, f_e and f_r are operating frequencies of the ETC reader and the AM radio, respectively. Therefore, we must enable the ETC reader to modulate broadcast signals at two carriers (f_e and $f_e + f_r$). Although total 52 frequency channels are available for an RFID reader, only a single channel can be used at any moment for the reading. In particular, our *transparent* design views the reader as a "black box", whose input is limited to predefined reader commands. To this end, Tagcaster whitens the baseband to a square signal by inputting a long sequence of `NAK` commands. Given that the reader uses pulse interval encoding (PIE) for the modulation, multiple harmonics can be observed on the receiving side. Inspired by this physical-layer characteristic, we set the parameter of `Tari` (i.e., length of bit zero) tactically to "frame" the first-order (i.e., fundamental) and fifth-order harmonics to appear at frequencies of f_e and $f_e + f_r$, respectively.

• *Modulate audio signals.* An ETC reader is a typical *digital* communication system whose baseband signal contains two different level voltages (i.e., high and low) only. The envelop of its modulated signals changes at two levels (i.e., OOK). On the contrary, AM stations and receivers are analog systems in which the quantized analog audio data are represented using multiple level voltages. Therefore, analog radio receivers cannot decode the digital binary signals transmitted from the reader. To deal with this issue, Tagcaster leverages the controllability of RF power to adjust the transmitting power dynamically for the required amplitude modulation. Specifically, Tagcaster initially quantizes the analog audio data into four-bit discrete values, then manipulates the transmitting power among 16 levels correspondingly. In such a way, the audio data can be carried onto the desired frequencies.

Contributions: We implement the prototypes of Tagcaster using an USRP N210. Nine off-the-shelf radio receivers including five vehicle-mounted and five general-purpose radio receivers are tested. Our results demonstrate that Tagcaster

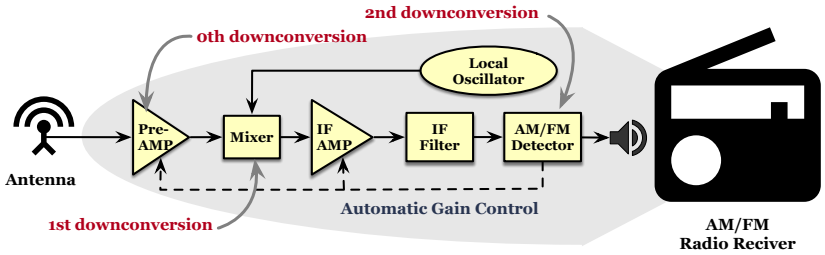


Figure 5-2: Internal structure of an AM Radio receiver. Input radio signals are processed through three downconversions where the zeroth downconversion is explicitly performed by the amplifier.

can fully provide AM radio service and enhance the ETC user experience on these devices. The perceptual evaluation of speech quality (PESQ) of the received voice is around 2, which is equal to that of the current telephone communication system. The coverage range is 30 m with two-way antennas. Demo audios are uploaded in [110].

In summary, this chapter presents Tagcaster, the first system utilizing the non-linearization phenomenon in radio receivers to provide high-quality radio service for ETC readers. The design of Tagcaster provides three key contributions. First, it proves the engineering possibility of down-converting communication with hardware non-linearity. Second, it introduces a new amplitude modulation by controlling RF power. Finally, the design of Tagcaster presents a practical prototype and a comprehensive evaluation.

5.1 AM Radio Primer

Most AM radio systems work at between 520 and 1700 kHz. Radio receivers usually adopt the superheterodyne design, as displayed in Fig. 5-2. Specifically, the RF signal (@ f_r) is amplified by the *pre-amplifier* to improve the signal-to-noise ratio (SNR). Then, the amplified RF signal enters a *superheterodyne mixer* along with the output of the local oscillator, which is tuned to a frequency (@ f_r') that is higher or lower than the intended reception frequency. As a result, the mixer output includes two signals, that operate at $f_r + f_r'$ and $f_r - f_r'$. The sum signal at $f_r + f_r'$ is immediately filtered out by the following IF filter. This

difference can always be at a fixed value of the frequency offset and is called the intermediate frequency (IF). This stage is called the *first downconversion* or *superheterodyning*. Superheterodyning exhibits good performance because radio components can be optimized to work at a single intermediate frequency. The desired baseband signal is then extracted by the *detector* (e.g., an envelope detector of Foster-Seeley discriminator), which performs the *second downconversion* by tuning the center frequency to the expected. The downconverted audio data are transmitted to the speaker.

5.2 System Design

In this section, we first analyze the nonlinear effect in AM radio and introduce how to leverage the nonlinearity to let AM radio communicate with UHF RFID reader. Finally, we discuss about the practical issues in building such cross-frequency communication channel.

5.2.1 Exploiting the Nonlinearity

As we discussed in Chapter 3, the RF amplifier has nonlinearity. If the input signal of an AM radio amplifier is two-tone signal $x(t) = \cos(2\pi f_1 t) + \cos(2\pi f_2 t)$, Eqn. 3.3 shows that there will be at least four new frequencies (i.e., $2f_1$, $2f_2$, $f_1 + f_2$ and $f_1 - f_2$) are created after magnification. Translating into actual numbers for AM Radio communication band, when $f_1 = 920$ MHz and $f_2 = 920.7$ MHz, the amplified signals appear at 920 MHz, 920.7 MHz, $2 \times 920 = 1840.7$ MHz, $2 \times 920.7 = 1841.4$ MHz, $920 + 920.7 = 1840.7$ MHz, and $(920.7 - 920)$ MHz $= 700$ kHz. The first five frequencies are immediately filtered out by the IF filter. However, 700 kHz remains. The net effect is that a completely UHF signal appears at a low frequency of 700 kHz, which radio receivers can process. The nonlinearity effect was considered a type of "pollution" in previous work. Nevertheless, we explore this underlying physical property as an opportunity to achieve cross-technology communication between ETC system and AM radio receivers. Since difference frequency is our interest, this item is extracted from

Eqn. 3.3 and expressed as follows:

$$S_{\downarrow}(t) = \frac{1}{2} a_2 \cos(2\pi(f_1 - f_2)t) \tag{5.1}$$

where $S_{\downarrow}(t)$ is the downconverted signal due to the nonlinearity effect. The previous discussion inspires us to leverage nonlinearity to pull down UHF signals to the radio band. Specifically, Tagcaster enables an ETC reader to transmit at two carriers, whose frequencies are denoted by f_1 and f_2. When the two signals pass through the pre-amplifier simultaneously, a low-frequency signal appears at $|f_1 - f_2|$ in the amplifier. This process is equivalent to conducting an additional downconversion before the first and the second downconversions at the mixer and detector. To distinguish them, we refer to the downconversion caused by the nonlinearity effect at the pre-amplifier as the *zeroth downconversion*. Fig. 5-2 shows their stages and process order. The next discussion is about the leveraging of the zeroth downconversion for Tagcaster radio service.

Upconversion at the reader side. Suppose that an ETC reader and a radio receiver work at frequencies of f_e and f_r, respectively. To activate the zeroth downconversion, our ETC reader modulates the audio data onto two carries at f_r and $f_e + f_r$ simultaneously. For clarity, we call the new carrier at $f_e + f_r$ the *shadow carrier*. The left part of Fig. 5-3 illustrates an example where $f_r = 550$ kHz and $f_e = 920$ MHz. The audio data are modulated onto the 920 MHz original carrier and 920.55 MHz shadow carrier. Formally, $v(t)$ denotes the audio signal, which is a low-frequency signal below 20 kHz. Then, the output upconverted RF signal from the reader is given as

$$S_{\uparrow}(t) = v(t)(\cos(2\pi f_e t) + \cos(2\pi(f_e + f_r)t)) \tag{5.2}$$

One might wonder why a single-tone signal is not generated at the shadow carrier for the downconversion only. The hardware limits the current commercial ETC reader to work at a single channel for each reading.

Downconversion at the receiver. The occurrence at the radio receiver side

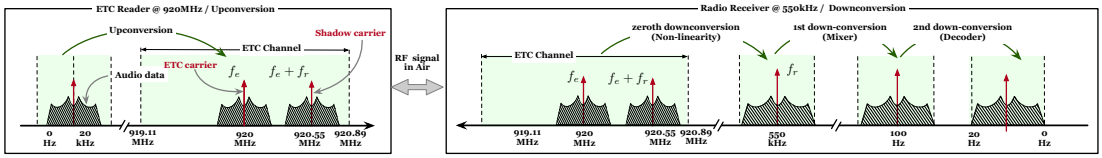


Figure 5-3: Illustration of Tagcaster design. At the reader side, the audio data are modulated onto two carriers: the real reader carrier at f_e and the shadow carrier at $f_e + f_r$. After receiving the RF signal, the receiver pulls down the signal by the zeroth, the first, and the second downconversion in turn. The zeroth downconversion is conducted due to the nonlinearity of the pre-amplifier.

is subsequently examined. After receiving the mixed signal transmitted from the ETC reader, the *pre-amplifier* in the radio receiver automatically performs the zeroth downconversion. Substituting the given signal into Eqn. 5.1, a new signal is produced as follows.

$$S_{\downarrow}(t) = \frac{1}{2}v^2(t)\cos(2\pi(\cancel{f_e} + f_r - \cancel{f_e})t) = \frac{1}{2}v^2(t)\cos(2\pi f_r t) \tag{5.3}$$

$S_{\downarrow}(t)$ is the result of the zeroth downconversion. It operates at the radio frequency f_r, which the receiver can process. That is, the zeroth downconversion can pull the mixed signal at f_e and $f_e + f_r$ down to radio frequency f_r. The right side of Fig. 5-3 presents the entire workflow at the receiver. $S_{\downarrow}(t)$ can be further downconverted twice by the mixer and the decoder to extract $v^2(t)$, which is finally played by the speaker. Notably, the side effect of our design is that the audio signal is distorted due to the squaring, i.e., $v^2(t)$. We can eliminate this distortion by taking a square root of the raw audio signal (i.e., $\sqrt{v(t)}$) before it is modulated onto the carriers, such that the downconverted signal $S_{\downarrow}(t) = \frac{1}{2}(\sqrt{v(t)})^2\cos(2\pi f_r t) = \frac{1}{2}v(t)\cos(2\pi f_r t)$.

5.2.2 Activating the Zeroth Downconversion

The previous discussion inspires us to leverage nonlinearity to pull down UHF signals to the radio band. Specifically, Tagcaster enables an ETC reader to transmit at two carriers, whose frequencies are denoted by f_1 and f_2. When the two signals pass through the pre-amplifier simultaneously, a low-frequency signal appears at

$|f_1 - f_2|$ in the amplifier. This process is equivalent to conducting an additional downconversion before the first and the second downconversions at the mixer and detector. To distinguish them, we refer to the downconversion caused by the nonlinearity effect at the pre-amplifier as the *zeroth downconversion*. Fig. 5-2 shows their stages and process order. The next discussion is about the leveraging of the zeroth downconversion for Tagcaster's radio service.

Upconversion at the reader side. Suppose that an ETC reader and a radio receiver work at frequencies of f_e and f_r, respectively. To activate the zeroth downconversion, our ETC reader modulates the audio data onto two carries at f_r and $f_e + f_r$ simultaneously. For clarity, we call the new carrier at $f_e + f_r$ the *shadow carrier*. The left part of Fig. 5-3 illustrates an example where $f_r = 550$ kHz and $f_e = 920$ MHz. The audio data are modulated onto the 920 MHz original carrier and 920.55 MHz shadow carrier. Formally, $v(t)$ denotes the audio signal, which is a low-frequency signal below 20 kHz. Then, the output upconverted RF signal from the reader is given as

$$S_{\uparrow}(t) = v(t)(\cos(2\pi f_e t) + \cos(2\pi(f_e + f_r)t)) \tag{5.4}$$

One might wonder why a single-tone signal is not generated at the shadow carrier for the downconversion only. The hardware limits the current commercial ETC reader to work at a single channel for each reading. Further details are discussed in §5.4.

Downconversion at the receiver. The occurrence at the radio receiver side is subsequently examined. After receiving the mixed signal transmitted from the ETC reader, the *pre-amplifier* in the radio receiver automatically performs the zeroth downconversion. Substituting the given signal into Eqn. 5.1, a new signal is produced as follows.

$$S_{\downarrow}(t) = \frac{1}{2}v^2(t)\cos(2\pi(\cancel{f_e} + f_r - \cancel{f_e})t) = \frac{1}{2}v^2(t)\cos(2\pi f_r t) \tag{5.5}$$

$S_{\downarrow}(t)$ is the result of the zeroth downconversion. It operates at the radio frequency f_r, which the receiver can process. That is, the zeroth downconversion can pull the mixed signal at f_e and $f_e + f_r$ down to radio frequency f_r. The right side of Fig. 5-3 presents the entire workflow at the receiver. $S_{\downarrow}(t)$ can be further downconverted twice by the mixer and the decoder to extract $v^2(t)$, which is finally played by the speaker. Notably, the side effect of our design is that the audio signal is distorted due to the squaring, i.e., $v^2(t)$. We can eliminate this distortion by taking a square root of the raw audio signal (i.e., $\sqrt{v(t)}$) before it is modulated onto the carriers, such that the downconverted signal $S_{\downarrow}(t) = \frac{1}{2}(\sqrt{v(t)})^2 \cos(2\pi f_r t) = \frac{1}{2} v(t) \cos(2\pi f_r t)$.

5.2.3 Practical Discussions

The following practical concerns must be noted.

5.2.3.1 How to deal with frequency hopping

Radio frequency (i.e., channel) f_r must be fixed and informed to drivers in advance so that drivers can adjust the receiver at f_r. However, FCC regulations specify that readers can have a maximum channel dwell time of 400 ms in any 10-second period to reduce interference in a channel. In other words, reader frequency f_e should hop among the 52 defined UHF channels every 400 ms. A logical question is whether the frequency hopping changes the radio channel or not? The answer is negative because the frequency after the zeroth downconversion is the difference of f_e and $f_e + f_r$. Regardless of how f_e is changed, the difference of two remains f_r. Thus, drivers do not need to change the receiving frequency at all.

5.2.3.2 How to deal with the Doppler effect?

The second concern is about the impact of the Doppler effect. At first glance, the Doppler effect does not appear to be a problem in practice because drivers can listen to AM radio inside their cars even while driving fast in an expressway. It does not matter if the audio data are carried at low-frequency radios (e.g.,

$<$ 1700 kHz). Frequency shift $\Delta f = \frac{\Delta v}{c} f$, where c is the speed of light, Δv is the relative velocity, and f is the carrier frequency. We are unaware of the frequency shift in an AM radio because carrier frequency f_r is small. Even when setting $f_r = 1700$ kHz (i.e., the upper frequency of AM radio) and $\Delta v = 100$ km/h (i.e., the upper limit of driving speed in the US), the resulting shift $\Delta f = 100\text{km/h} \times 1700\text{kHz}/(3 \times 10^8\text{m/s}) = 0.01$ Hz is too small to be perceived. In our scenario, carrier frequencies f_e and $f_e + f_r$ are up to 928 MHz. The potential frequency shift is raised to $100\text{km/h} \times 928\text{MHz}/(3 \times 10^8\text{m/s}) = 85$ Hz, which can no longer be ignored in practice.

Two carries are transmitted from the same ETC antenna and received by the same receiver in our scenario. Hence, they obtain an identical relative speed. As a result, the signal after the zeroth downconversion centers at the frequency below:

$$\left((f_e + f_r) + \overbrace{\frac{\Delta v}{c}(f_e + f_r)}^{①} \right) - \left(f_e + \overbrace{\frac{\Delta v}{c} f_e}^{②} \right) = f_r + \frac{\Delta v}{c} f_r \tag{5.6}$$

where the two items on the left indicate the shadow and real carriers, and ① and ② are the frequency shifts caused by the Doppler effect at two carriers respectively. Clearly, the final frequency shift after the zeroth downconversion is almost removed from the difference. The remainder is related to radio frequency f_r only. As previously discussed, such frequency shift is negligible (i.e., 0.01 Hz). Thus, the Doppler effect does not affect our design because of the differencing, although the two carriers are highly shifted.

5.2.3.3 How to protect privacy?

Audible messages may reveal the driver's private information (e.g., identifier, balance or destination). Different from traditional RFIDs, ETC systems usually use highly directional antennas to avoid reply collisions when multiple transponders are covered by a reader at the same time. This circumstance is reasonable because all vehicles must run within different lanes and are separated by a min-

imum safe distance. This condition also provides a simple but efficient privacy protection means that ensures that the broadcasted audio data are isolated from one another physically. For stronger protection, we can encrypt the audio data in advance. In this case, vehicle-mounted radio receivers may fail to decrypt the audio. Drivers can use their smartphones to receive the radio signal and decrypt the audio data in the application layer.

5.2.3.4 Why not to provide FM Radio?

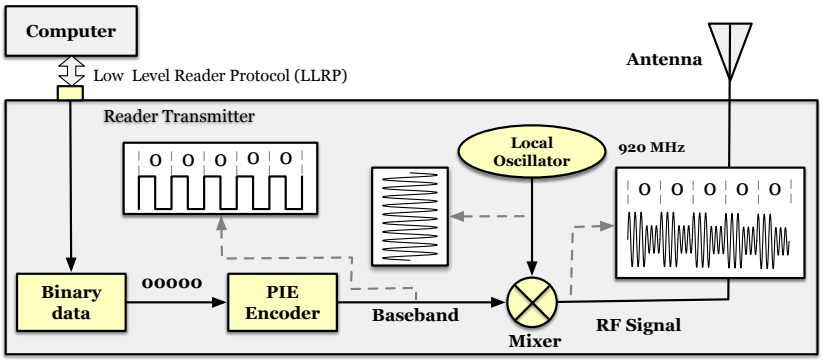
FM radio is another popular radio broadcasting using frequency modulation (FM) with an operating spectrum from 88 to 108 MHz. FM radio is not served at Tagcaster because it requires at least 88 MHz difference between the ETC carrier and the shadow carrier to activate the zeroth downconversion at receivers. This is far beyond the 20 MHz bandwidth of an RFID system.

5.3 Engineering Tagcaster

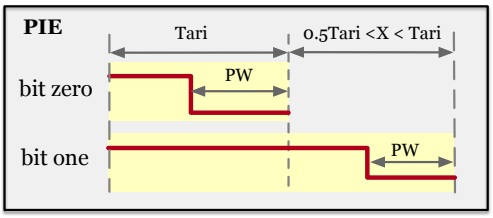
The core of Tagcaster is the engineering of dual-carrier upconversion at the ETC reader (i.e., modulating the audio data onto two carriers) because only such an upconversion can activate the zeroth downconversion at the radio receiver. However, achieving this task is challenging because Tagcaster is required to be a *transparent* service. The only way to change the behaviors of ETC readers is to feed data in the application layer in accordance with corresponding standards. Given such a strict constraint, two engineering challenges are discussed in this section.

- **Generating the Shadow Carrier.** The shadow carrier is used to pull RF signals from UHF down to the radio frequency at radio receivers. The first challenge is how to generate an undefined shadow carrier by complying with ETC regulations.

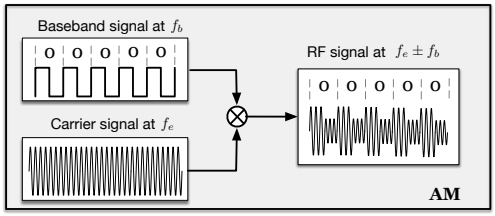
- **Modulating the Audio Signal.** An AM radio conveys the the *analog* audio data by changing the amplitude of the carrier, whereas the reader



(a) The schematic of reader transmitter



(b) Pulse-Interval Encoding (PIE)



(c) Amplitude Modulation (AM)

Figure 5-4: The RFID transmission. (a) The incoming bitstream from computer is firstly encoded through the PIE. Then The PIE-coded baseband signal is moved to the ultra-high frequency (e.g., 820 MHz) by multiplying the carrier generated from the local oscillator. Finally, RF signal is propagated into the air through the antenna; (b) shows the adjustable parameters of PIE used in RFID; (c) shows the amplitude modulation where the reader is transmitting a continuous stream of zero bits.

baseband only accepts the *digital* bitstream from the upper layer. The second challenge is how to carry the analog data through a digital wireless system.

5.3.1 Primer on RFID Transmission

To explain Tagcaster, how an ETC reader works must first be introduced. Readers are required to generate high-power CW, which persistently supplies energy to passive transponders [1] in the field. Two data links are involved. The first link is data transmission from readers to tags, called *downlink* transmission. The second link is the opposite, which is called *uplink transmission*. Given that Tagcaster broadcasts audio in a single way, we only introduce the downlink here. Fig. 5-4a

[1]Many ETC systems use active transponders for improved communication. Since our ultimate goal of Tagcaster is AM broadcasting from the reader to radio receivers, the types of transponders actually do not affect our design.

illustrates a schematic of a reader transmitter that contains four main components: *PIE encoder*, *local oscillator*, *mixer*, and *antenna*. The entire workflow is sketched.

Baseband Encoding. A reader encodes the data (e.g., commands) coming from the host using PIE in baseband. Fig. 5-4b illustrates the coding scheme. PIE coding uses different durations to represent bit zero and bit one. Bit zero has the duration of a single `Tari`, whereas that of bit one equals to `Tari`+`X`. `Tari` is the unit duration for the signaling reference. It can be set as from 6.25 to 25μs. The duration of bit one is always `X`-us longer than that of bit zero and must be between 1.5 and 2 `Tari`. Both bits start with a high voltage and end with a low voltage. The durations of low voltage for the two bits are the same and equal to the pulse width (`PW`). `Tari`, `PW` and `X` can be set by users to suit their scenarios.

Modulation. The PIE-coded baseband signal is then multiplied by the UHF carrier generated from a local oscillator to produce the output RF signal. Fig. 5-4c shows this procedure. Given that the multiplication only changes the amplitude of the carrier to carry the baseband signal, it is called as AM. In terms of carrier frequency, the ISO/IEC 18000-6 standard only specifies a broad spectrum (i.e., 820-920 MHz) and allows local agencies to regulate the channel division. For example, FCC 15.247 authorizes RFID readers to operate in the ISM band from 902 - 928 MHz with 52 channels, each of which has a maximum bandwidth of 500 kHz.

5.3.2 Generating the Shadow Carrier

The shadow carrier is the key for activating the receiver's zeroth downconversion. We must generate the shadow carrier in accordance with RFID regulations.

5.3.2.1 Rationale behind the Shadow Carrier

Modulation translates the entire spectrum of a baseband signal to a high band centering at the carrier frequency. Formally, $S_e(t) = \cos(2\pi f_e t)$ denotes the carrier. If the baseband signal is a simple sinusoidal signal denoted by $S_b(t) =$

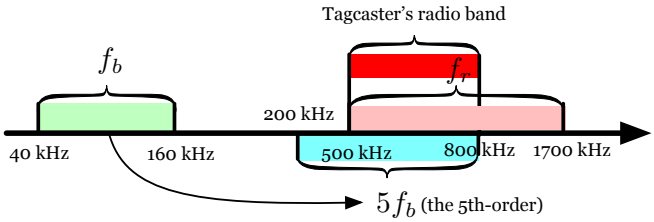


Figure 5-5: Spectrum comparison. The fundamental frequency of the baseband signal f_b in the ETC reader is from 40 to 160 kHz. Correspondingly, its fifth order harmonic $5f_b$ varies from 200 $\sim$ 800 kHz. A commercial AM radio works from 500 to 1700 kHz. The overlapping spectrum from 500 to 800 kHz is the band at which Tagcaster's radio operates.

$A_{\text{dc}} + \cos(2\pi f_b t)$, then the modulated signal propagated into the air is given by

$$\begin{aligned} S_{\text{air}} =& S_b(t)S_e(t) = (A_{\text{dc}} + \cos(2\pi f_b t))\cos(2\pi f_e t) \\ =& A_{\text{dc}}\cos(2\pi f_e t) + \frac{1}{2}\cos(2\pi(f_e + f_b)t) \\ &+ \frac{1}{2}\cos(2\pi(f_e - f_b)t) \end{aligned} \tag{5.7}$$

where A_{dc} is the direct constant. The equation implies that the output signal in the air appears at f_e and $f_e \pm f_b$. Inspired by this fundamental, we can generate a shadow carrier by inputting a well-designed baseband signal. Specifically, if we transmit a continuous stream of constant zeros, then the PIE-coded baseband signal of the reader becomes a *square signal* with the fundamental frequency of $f_b = 1/\texttt{Tari}$. The duration of the bit zero that is equal to `Tari` and `PW` is set to `Tari`/2 (see Fig. 5-4b). Therefore, readers are effectively "framed" to transmit continuous signals at f_e (i.e., direct constant) and $f_e + f_b$ (upconversion).

By changing the value of `Tari`, the fundamental frequency of the square signal at the baseband can be varied within 40 $\sim$ 160 kHz (i.e., $f_b \in [40, 160]$ kHz). However, we desire a frequency shift of $f_r \in 500 \sim 1700$ kHz (i.e., AM radio frequency). Even the upper limit of f_b (i.e., 160 kHz) cannot reach the lower limit of f_r (i.e., 500 kHz) because FCC regulation only allocates 500 kHz bandwidth to RFID reader for each channel (see Fig. 5-5).

The fundamental of signal processing indicates that the square wave is com-

posed of infinite sinusoidal harmonics. Thus, $S_b(t)$ can be expanded as follows by using the Fourier series.

$$\begin{aligned} S_b(t) &= A_{\text{dc}} + \frac{4}{\pi} \sum_{n=1,3,5,\ldots} \frac{1}{n} \sin(2\pi n f_b t) \\ &= A_{\text{dc}} + \frac{4}{\pi} \left(\underbrace{\sin(2\pi f_b t)}_{\text{1st-order}} + \underbrace{\frac{1}{3}\sin(2\pi 3 f_b t)}_{\text{3rd-order}} + \underbrace{\frac{1}{5}\sin(2\pi 5 f_b t)}_{\text{5th-order}} + \cdots \right) \end{aligned} \tag{5.8}$$

Instead of a single-tone signal, S_b is composed of infinite odd sinusoidal signals at frequencies of $f_b, 3f_b, 5f_b, \cdots$, which are called first-order, third-order, fifth-order harmonics, and so on, respectively. By substituting Enq. 5.8 into Eqn. 5.7, the modulated signal in the air is updated as follows:

$$\begin{aligned} S_{\text{air}} &= \left(A_{\text{dc}} + \frac{4}{\pi} \sum_{n=1,3,5,\ldots} \frac{1}{n} \sin(2\pi n f_b t) \right) \cos(2\pi f_e t) \\ &= A_{\text{dc}} \cos(2\pi f_e t) \\ &\quad + \frac{2}{\pi} \sum_{n=1,3,5,\ldots} \frac{1}{n} \left(\sin(2\pi (f_e + n f_b) t) + \sin(2\pi (f_e - n f_b) t) \right) \end{aligned} \tag{5.9}$$

Eqn. 5.9 indicates that the output RF signal actually appears at f_e, $f_e \pm f_b$, $f_e \pm 3f_b$, $f_e \pm 5f_b$, $\cdots$. Given that $f_b \in 40 \sim 160$ kHz, the fifth-order harmonic $5f_b$ falls into the range of $200 \sim 800$ kHz. It has 300 kHz overlapping (i.e., from $500 \sim 800$ kHz) with the allowable AM radio spectrum. To visually understand the spectrum, we illustrate various bands and their relation in Fig. 5-5. Therefore, the frequency of Tagcaster's radio service can be fixed at any frequency between 500 and 800 kHz (highlighted in red). Conversely, if radio frequency $f_r \in [500, 800]$ kHz, then we should set the duration of the bit zero at the baseband to.

$$\texttt{Tari} = 1/f_b = 5/f_r \tag{5.10}$$

where $f_r = 5f_b$. For example, if we choose $f_r = 500$ kHz, then $\texttt{Tari} = 1/f_b =$

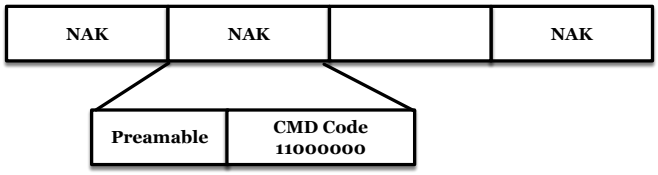


Figure 5-6: Whitening the baseband of the ETC reader. The reader is forced to keep transmitting NAK command, which contains an eight-bit constant code with six zeros.

$1/100$ kHz $= 10\ \mu s$.

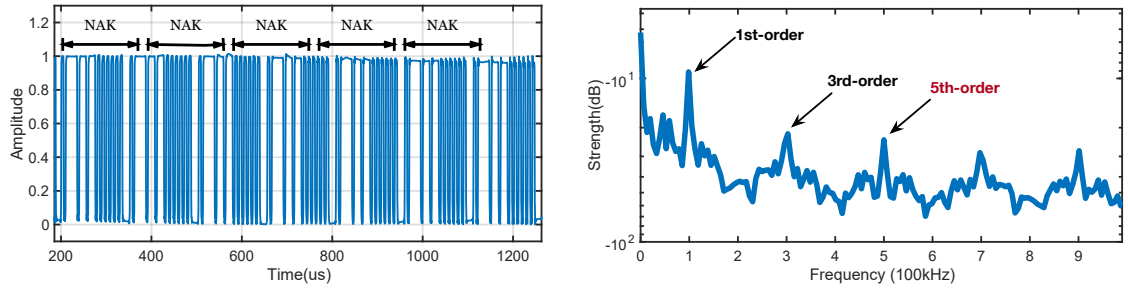
5.3.2.2 Whitening the Baseband with Zeros

Our basic idea of generating a shadow carrier is to force the reader to transmit a long sequence of bit zeros. This procedure is called as baseband whitening. However, commercial ETC readers only accept the predefined commands from hosts. A long sequence of zeros is unacceptable. About 30 commands defined in the ISO18000-6 or EPCglobal Gen2, among which we select the NAK command to whiten the baseband. Fig. 5-6 shows the structure of this command, which starts with a three-bit unmodified preamble and contains eight-bit command code. The time consumed for one NAK transmission is given as

$$(3 + 2 \times 1.5 + 6) \times \texttt{Tari} = 12 \times \texttt{Tari} \tag{5.11}$$

where 3 Taris are for the preamble, 1.5 Taris are for the two ones (i.e., X$= 0.5 \times$ Tari), and 6 Taris are for the six zeros. NAK command is selected for us in two reasons. First, the payload of the command is fixed to the bitstream of 1100000 where 75% of the bits are zeros. Second, NAK is a *mandatory* command that all commercial readers must support. In practice, we can command the reader to keep transmitting NAKs to achieve long-sequences of bit zeros approximately. The transmission of the remaining 25% bit ones almost does not affect the broadcast because their presence lasts for a short time (a few microseconds) in each cycle. Therefore, the instant pause in voice is hardly noticed by humans.

To verify this idea, we uses an ETC reader (refer to §7.6 for details) to transmit a sequence of NAKs by setting $f_r = 500$ kHz. Then, we set the parameter of Tari



(a) Time domain **(b) Frequency domain**

Figure 5-7: Illustration of the shadow carrier. The reader is forced to transmit a long sequence of `NAK`s. (a) shows the received signal in the time domain and (b) shows the spectrum of the signal.

to 10 μs (see Eqn. 5.10). We also employ USRP to receive the RF signal. Fig. 5-7 illustrates the baseband signal of the received signal in time and frequency domains. As desired, the signal spikes exactly at 100 kHz (i.e., first order), 300 kHz (third order), 500 kHz (fifth-order), and so on. This results verify that we can whiten the reader's baseband using the `NAK`s.

5.3.3 Modulating Audio Signal

Both AM radio and ETC reader adopt amplitude modulation to carry baseband signals. At first glance, Tagcaster can directly use the modulation component in an RFID reader for the AM modulation. Unfortunately, this naive approach fails to work in practice for two reasons. First, AM radios stations and receivers are designed for processing analog audio signals, but readers can only process digital signals. Specifically, the envelope of the reader's carrier only has two levels (Fig. 5-7a), whereas an AM radio uses different amplitude levels to represent quantized analog audio data (Fig. 5-8). Second, this approach requires the modification of reader hardware and firmware, which violats our design principle of transparency.

The essence of amplitude modulation is to carry data by changing the amplitude of the output RF signal. Commercial RFID readers can *dynamically* set the transmitting power in a *real-time*. For example, ImpinJ R2000 [111] has 31 power levels that the user can set. This functionality inspires us to modulate audio signal by adjusting the power of the output RF signal directly instead of

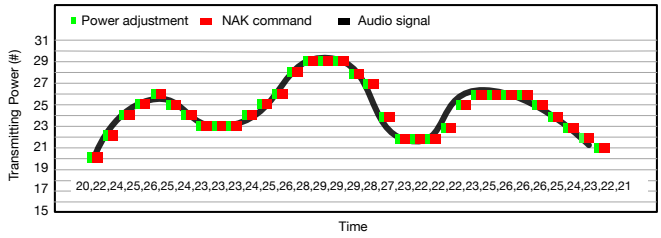


Figure 5-8: Modulation of audio data. Audio data are sampled every 12 `Taris` on the time line and quantized into 16 levels on the amplitude line. Each box corresponds to a sample.

modulating the multiplication. In doing so, we can skip the baseband processing to achieve amplitude modulation equivalently. The whole procedure is sketched as follows:

- **[Step 1] Sampling:** First, Tagcaster resamples audio data every 12 `Taris`. The sampling period is exactly equal to the duration of a `NAK` command (see Eqn. 5.11) because `NAK` is the minimum unit before which the transmitting power can be updated. Correspondingly, the sample rate is equal to $1/(12 \times \texttt{Tari}) = f_r/60$ Hz (see Eqn. 5.10). Given that $f_r = 500 \sim 800$ kHz, the sampling rate is equal to $8.33 \sim 13.33$ kHz. An 8 kHz sampling rate is regarded as adequate for human speech. For example, the telephone system usually uses 8 kHz ADC [112]. Thus, our sampling rate can fully address the common quality demand of radio broadcasting. Fig. 5-8 provides an example where each box represents one sampling.

- **[Step 2] Quantization:** Second, Tagcaster quantifies the amplitude of audio data into 16 levels, namely, four-bit quantization. Each quantified result corresponds to an output RF power level. A normal audio ADC adopts 8-bit or 16-bit quantization. However, we can only set the RF power to one of the 32 predefined levels in the reader. Moreover, we must ensure that the signal can propagate into air with sufficient energy. Only 16 levels (from 15^{th} to 31^{st}) are available for us (four-bit quantization). In Fig. 5-8, the audio signal is quantified to 16 levels indicated by horizontal gray lines. Our evaluation reveals that four-bit quantization is acceptable.

- **[Step 3] Broadcasting:** Finally, Tagcaster broadcasts the audio samples

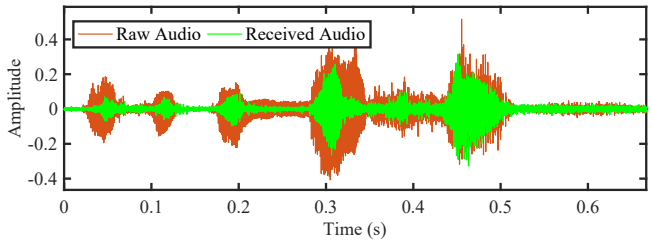


Figure 5-9: Raw audio vs. received audio

in such way: for each sample, it initiates a power adjustment and a subsequent `NAK` transmission. In Fig. 5-8, the green and red boxes indicate the RF power and `NAK` commands, respectively. The time cost for power adjustment is almost negligible because it does not require signal processing and is executed quickly (i.e., $< 1\ \mu s$).

In summary, `NAK` transmission holds the shadow carrier, whereas the power adjustment modulates audio data. The adjustment affects all RF signals coming out from the reader, so the audio data is actually modulated onto both carriers (f_e and f_e+f_r). This reason explains why we move $v(t)$ outside the sum of the two carriers in Eqn. 5.4. To validate the effectiveness of Tagcaster, Fig. 5-9 illustrates a comparison of the raw and received audio signals, both of which represent the sentence "Good morning, Mr. Bob!". The raw audio is generated by a text-to-speech software, while the received audio is recorded through a commercial radio receiver.

5.4 Implementation

Tagcaster Reader. We implement the prototype of the ETC reader for Tagcaster with an USRP-N210 SDR. It is equipped with an SBX daughterboard. An RF power amplifier [104] is used to magnify the max transmitting power to 31 dBm. The prototype fully supports Gen2 PHY [97]. Notably, the USRP emulated reader is used for evaluation purposes to measure low-level PHY information, such as harmonics and signal strength, which are inaccessible by commodity devices.

Radio Receiver. We test nine commercial radio receivers, including (1) five vehicle-mounted receivers (VMRs) at Toyota Sienna, Audi Q7, Audi Q5, Jetta Avant, and Jetta Sedan; and (2) four general-purpose receivers (GPRs), which are TECSUN ICR-110, Sony ICF-P36 [113], PANDA T-16, and AMHA 010. The main difference among them is sensitivity. VMRs are sensitive to work with low SNR.

AM Radio Channels. Seven radio channels (e.g., f_r) are listed in Table. 5.1. These channels are not uniformly distributed within 500 $\sim$ 800 kHz because the sampling rate is 2 MS/s in the reader so the adjustable step of `Tari` is 0.5 μs. Moreover, an AM radio receiver allows users to tune the frequency with a step of 5 kHz. However, only five channels are tested in our experiments because C5 and C7 are in conflict with commercial AM stations in our city.

5.5 Results

In this section, we evaluate the Tagcaster through outdoor experiments.

5.5.1 Communication Performance

We begin with a group of benchmark experiments to present the performance of the communication from the ETC reader to the AM radio receiver in terms of different parameter settings. The zeroth downconversion only occurs in radio receivers, which are composed of highly integrated circuits. We have no direct means to acquire downconverted low-level radio signals from these receivers. Recalling that our ultimate goal is to provide audio service, we use the audio data played from receivers to evaluate the link performance indirectly. The audio is recorded by a microphone in the format of WAV with a sampling rate of 48 kHz. A one-second audio is acquired for each setting. All experiments are conducted

Table 5.1: Radio Channel in Tagcaster

Channel(#)	C1	C2	C3	C4	~~C5~~	C6	~~C7~~
`Tari`(μs)	9.5	9	8.5	8	7.5	7	6.5
f_r(kHz)	530	555	590	625	665	715	770

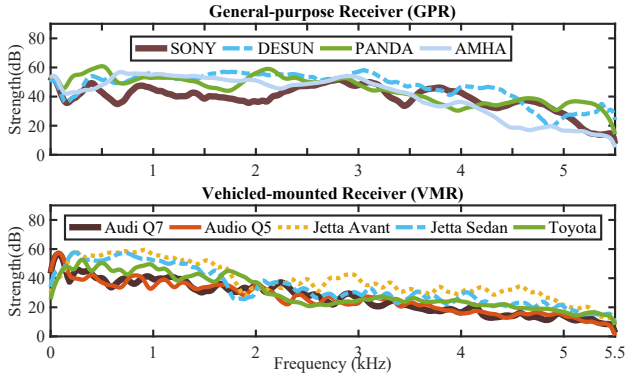


Figure 5-10: In-band frequency response

in a quiet environment.

5.5.1.1 Characterizing the In-band Response

We determine whether the nonlinearity effect works across diverse radio receivers or not. In our experiments, we use a reader to transmit a chirp-based audio signal at 530 kHz for every single receiver. The chip signal sweeps from 0 to 5.5 kHz. Fig. 5-10 shows the receiving power in the unit of dB at the receivers with respect to the two types of receivers. All receivers can output the chirp signals as desired in the entire spectrum, and the tendency is similar regardless of the types. This finding fully validates that the nonlinearity-enabled zeroth downconversion is a general physical characteristic of radio receivers. The average strength of GPRs is approximately 40 dB, whereas that of VMRs is around 30 dB. Therefore, GPRs performs better than VMRs. The additional 10 dB attenuation at VMRs is mainly caused by the car's metal body instead of the receiver's hardware. In addition, the strength gradually declines as the frequency increases. We could observe this similar tendency in other acoustic devices. This is mainly because the speaker and microphone (for recording) are designed for human voice and the transition bands of their low-pass filters are non-ideal, i.e., voice signal attenuates much more at higher frequencies [114].

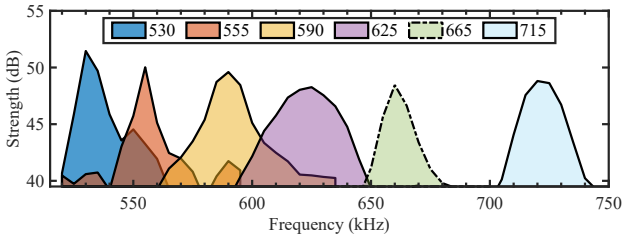


Figure 5-11: Out-of-band response

5.5.1.2 Characterizing the Out-of-band Response

Second, we measure the out-of-band frequency response in each radio channel. Unlike the previous evaluation, the broadcasting frequency f_r is fixed at the reader in this experiment. We tune the receiving frequency within the $f_r \pm 30$ kHz spectrum to measure the response. For example, if the reader broadcasts the radio at 715 kHz, then we measure the audio response within 715 ± 30 kHz. Fig. 5-11 shows the frequency response to the broadcasting in five channels. For comparison, we also measure the response of a commercial AM radio at 665 kHz in our city. Fig. 5-11 illustrates that Tagcaster's radio can be received between $f_r \pm 25$ kHz, whereas the commercial radio falls within ± 10 kHz. In practice, commercial radio channels are assigned at least 10 kHz intervals to avoid mutual interference. Tagcaster performs relatively worse in the spectrum control, because the ETC reader targeting RFID is not optimized for AM radio. These results guide us to select two channels with at least 30 kHz intervals for a single ETC station. ETC uses directional antennas, which can control RF signals in a narrow and specific direction. Thus, "leakage" has little impact in practice.

5.5.1.3 Characterizing the Broadcasting Range

Third, we evaluate audio strength as a function of the distance between the radio receiver and ETC reader. We notice that the reading range of an ETC reader for transponders is controlled under 10 m to ensure that a single vehicle closest to the station is identified. Tagcaster aims to broadcast related information about the vehicle. Thus, the broadcasting should be initiated exactly after when the vehicle is identified by the ETC system (i.e., when its distance to the reader is less

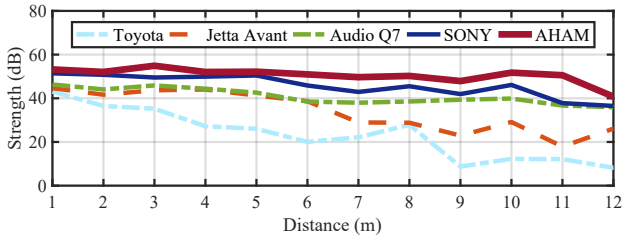


Figure 5-12: Impact of distance

than 10 m). Moreover, the broadcasting should only be received by the target for privacy protection. Thus, *a long-range broadcasting is unprofitable for* Tagcaster. Given these considerations, we only present the audio strength in the range of 12 m. The results are shown in Fig. 5-12. When the distance increases to 10 m, the strength is approximately 20 dB which is sufficient to provide a good quality for the audio decoding. In addition, current ETC stations are usually equipped with two independent antennas in the heading and leaving directions. Thus, the real coverage range for good-quality broadcasting is up to about 30 m in practice. Further, radio receivers are usually mounted in the head of the vehicle, so the length of the vehicle is irrelevant to the coverage.

5.5.1.4 Characterizing the Impact of Frequency Hopping

Finally, we show the impact of frequency hopping on the audio signal in Fig. 5-13. In the experiment, Tagcaster's reader randomly hops among the 52 channels between 902 and 920 MHz every 500 ms. The radio channel remains at 530 kHz. The mean strength with and without hopping is 41.1 dB and 41.9 dB, respectively. These results fully validate our theory in §5.2 that hopping does not affect the received audio due to the subtraction of two carriers.

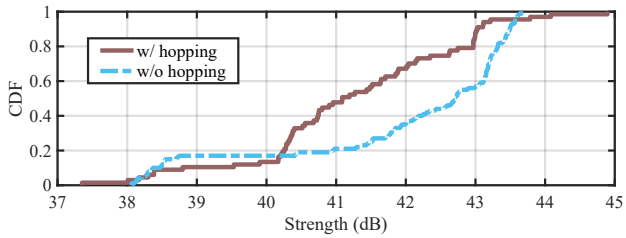


Figure 5-13: Impact of frequency hopping

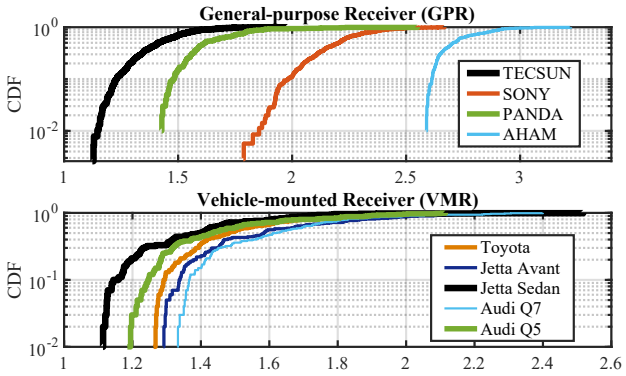


Figure 5-14: Audio quality in diverse receivers

5.5.2 Audio Performance

In this section, we evaluate the quality of the resulting audio signals in terms of PESQ. PESQ is a common metric used to measure the quality of telephony systems [115]. It outputs a perception score between 0 and 5, where a high score indicates good quality. Generally, the audio is good enough to be understood when the score is over 1.2. We manipulate the Tagcaster reader to broadcast the PESQ benchmark dataset [116] and use the official PESQ tool [116] to score the recorded audio data. Given that the PESQ tool only works for the audio data sampled with 16 or 8 kHz, we need to reduce the 48 kHz-recorded audio to 16 kHz. All experiments are conducted in our campus and in nearby noisy and busy streets (10 m away) where many vehicles run at every moment.

5.5.2.1 Audio Quality in Diverse Receivers

First, we evaluate audio quality with respect to different receivers. Fig. 5-14 displays quality comparisons for GPRs and VMRs. At a high level, the audio quality received by GPRs is better than that received by VMRs, because the radio signal is acquired by GPRs in a free space without the influence of the metal body. In addition, GPRs are analog systems that maintain higher fidelity qualification than digital VMRs. Particularly, the mean scores of Audi Q7 and Q5 are 1.6 and 1.5, respectively, whereas those of Jetta Avant and Sedan are 1.59 and 1.40, respectively. Usually, high-end vehicles are equipped with better audio

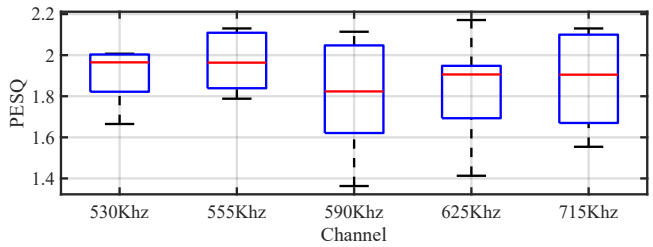


Figure 5-15: Audio quality in different channels

systems.

5.5.2.2 Audio Quality in Different Channels

Second, we evaluate the audio quality in different channels. The results are presented in Fig. 5-15. The average PESQ value of five channels is around 2, which is good enough for broadcasting service. The worst cases occur in the channels of 590 and 625 kHz, where the lowest scores are equal to 1.4. A commercial AM radio operates at 665 kHz in our city, and its leakage may interfere with these two channels. However, we can still understand the audio marked with a score of 1.4 even in a noisy situation.

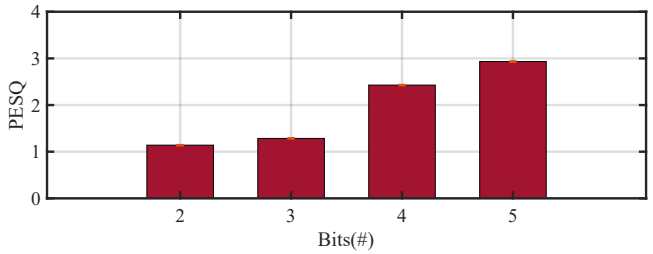


Figure 5-16: Audio quality with quantization

5.5.2.3 Audio Quality with Quantization Methods

Third, we evaluate audio quality as a function of quantization bits with a GPR. We display the quality of 2 to 5-bit quantization in Fig. 5-16. The R2000 reader has 31 power levels and can thus support four-bit quantization at the most. The first three quantization methods are tested in the R2000 reader, and the last one is transmitted from the USRP reader for comparison. The scores for the four cases are 1.2, 1.3, 2.5 and 2.8. In Tagcaster's modulation, audio data are

quantized into 16 levels, i.e., four-bit quantization, which have an average score of 2.5. Clearly, it increases as the quantization bits increases. However it can be found that the scores of bit 4 and bit 5 are very close. This shows that the 4 bit quantization is good enough in practice.

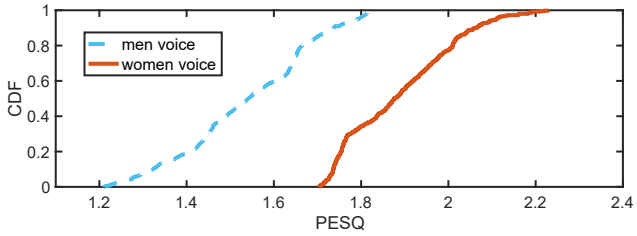


Figure 5-17: Quality vs. Contents

5.5.2.4 Performance on different contents.

Another important factor is the audio contents because different audio sources have different frequency distribution, which will impact the audio quality. We collect the sound from women's speech, men's speech and music and then broadcast them with Tagcaster. The other setting is the same as previous. As shown in Fig. 5-17, the women's speaking has better audio quality than the men's. This is because the men's voice has lower frequency than the women and most electronic noise of radio system is low frequency band from 0 to 100 Hz, which makes Tagcaster has worse performance for men's speech. However, even for men's speaking, the average PESQ score is still higher than 1.5 and this is enough for us to distinguish the speaking content.

5.5.2.5 Audio Quality at Different Driving Speeds

Fourth, we evaluate audio quality by considering the impact of driving speed. The Doppler effect can be an issue for UHF carriers, as indicated in §5.2. Fig. 5-18 illustates audio quality as a function of driving speed. In the figure, the positive and negative speeds indicate that the vehicle is heading to and leaving from the ETC station, respectively. The audio quality fluctuates only within the score of ± 0.2 compared with the stationary case where the speed is zero. When the vehicle is driving at 50 km/h, the 920 MHz carrier shifts to 42.6 Hz, but the radio

receiver only detects a 0.01 Hz shift.

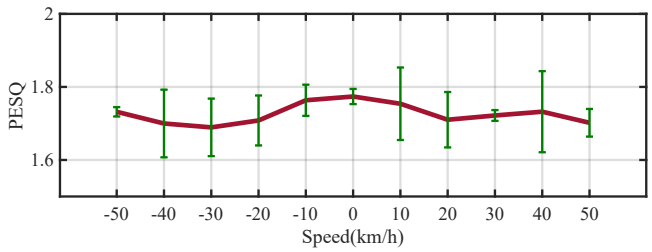


Figure 5-18: Impact of driving speed

5.5.3 Human Experience

We finally investigate the user experience of Tagcaster. We invite 20 drivers to experience Tagcaster service and ask them to rate the service. The rating score is from 0 to 5, where 5 is excellent. Fig. 5-19 illustrates a comparison of AM radio and ETC service. The subjective opinions of the 20 drivers are strongly positive. They appreciate the in-time wireless voice notification of charging fee using the AM radio, which is described as "extremely convenient and interesting". Specifically, approximately 60% of the volunteers have rated our service with a score of 4+ whereas only 30% have given scores to existing ETCs. However, 15% of the volunteers said that "the audio quality should be improved further" compared with commercial radios. This response is understandable because Tagcaster pursues zero start-up cost at the price of audio quality. Nevertheless, over 40% of the volunteers believe that the current audio quality is already sufficient for short-message notification.

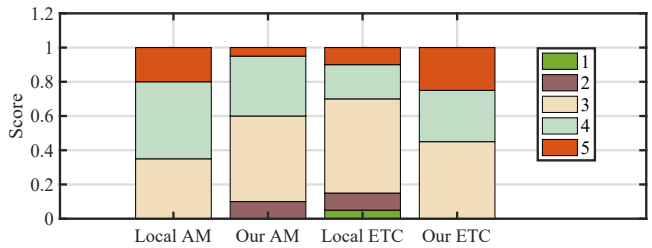


Figure 5-19: User feedback on Tagcaster service

5.6 Related Work

We review related work in three fields.

(1) Nonlinearity effect. Although the study and exploitation of the utilization of nonlinearity in diode based devices are not new, the harmonics in the RFID system have only elicited attention in recent years. The harmonic phenomenon in RFIDs was reported in [58, 64–66, 68, 98, 106], which focused on eliminating the negative impact of RF amplifier nonlinearity [117]. The work of [67] characterized the harmonic signals in UHF RFID via extensive experiments. The study of [69] used harmonics to achieve multi-frequency continuous wave ranging and further localize tags in 3D space. The work also explored harmonics as a secondary communication channel [58]. Deepak [118] introduced a new backscatter device for deep issue detection through the nonlinearity effect. However, unlike previous work that focused on tag's uplink communication, our work is the first to bridge the communication from ETC readers to AM radio receivers.

(2) Cross-technology communication (CTC). Many recent studies on CTC introduced deep cooperation between heterogeneous wireless devices. Most of them focused on the technologies in the same ISM band, such as Wi-Fi and Zigbee [26, 39, 40, 42, 43]. Specifically, WeBee [39] introduced a physical-level emulation technique to provide a high-throughput connection. The work of [119] utilized the harmonic backscatter technique to connect the UHF RFID and Wi-Fi. This work suggested a new type of CTC, that has never been used before.

(3) Backscatter and RFID. Similar to RFID tags, backscatters are battery-free devices that modulate data by reflecting the source signals. Dozens of backscatters have been proposed in the past years [20, 49, 50, 52, 56, 87, 88]. Our work is inspired by the FM backscatter [52], which reflects FM radio signals for broadcasting. Previous studies embedded the RFID reader into a bulb to make it easily deployable indoors [120]. ETC transponders have been used to localize and count vehicles for building smart cities [109]. By contrast, our work aims

to enhance RFID application in outdoor ETC service with powerful human to machine interaction.

5.7 Discussion

This chapter presents Tagcaster, a novel CTM system that enables commercial UHF ETC systems to provide additional broadcasting service with only a software update. Tagcaster is the first system to offer down-converting cross-technology communication. This system proves that the intermodulation can be used for creating the communication channel to the devices working on the lower frequency band. We believe there are huge potentials for building cross-technology communication with intermodulation. Another lesson we learned from this project is that we can emulate the analog communication signal with adjusting the high-speed digital communication data packets. Our extensive experiments indicate that Tagcaster can provide good-quality radio service with only a software updated ETC reader. However, the current prototype still has a limitation in the audio quality compared against the real AM radios due to the trade-off for cost. Our future work will generate a better audio by compensating for the hardware diversity identified via the transponder.

Chapter 6

Revitalizing Ultrasonic Positioning Systems for Ultrasound-Incapable Smart Devices

Tracking smart devices, such as phones or wearables, inside buildings where the GPS is not available has become a growing business interest. Indoor localization enables users to navigate indoor spaces similar to the functionality provided by GPS for outdoor environments. It sparks a series of key mobile applications, namely, indoor navigation (e.g., malls, factories, and airports), augmented reality, location-aware pervasive computing, advertising, and social networking.

Many efforts have been exerted to deliver an accuracy of decimeters for indoor localization. However, only a few of these solutions have actually reached today's mobile devices because previous works have failed in one of three main categories at a high level. The first group requires mobile smart devices to be equipped with specialized sensors, such as high-precision accelerators [121], magnetic sensors [122], LEDs [123, 124], or RFIDs [69, 96, 125–127]. These sensors do not become the compulsory components yet for the newly emerged smart devices

(such as wireless headphones, VR/AR glasses, smart watches, etc), despite being widely used in smartphones. The second group requires exhaustive fingerprinting of the environment to learn the spatial distribution of signal characteristics, such as FM [128], WiFi [129,130], Bluetooth [131], sound or light [132,133], and Zigbee [134], etc. The third group like ultra-wide band (UWB) systems (e.g., WiTrack [135]) can achieve cm- (even mm-) level accuracy but normally requires GHz bandwidth, which is usually not allowed in practice due to the spectrum regulations. On the contrary, as the rapid development of smart technology, the demand for highly accurate indoor localization services has become a key prerequisite in some markets, thereby resulting in a growing business interest.

In this chapter, we revisit a classical indoor localization solution, namely, the *ultrasonic positioning system* (UPS), which utilizes the ultrasonic sound as the ranging media. UPS outperforms RF-based systems in terms of accuracy. For example, MIT launched Project Oxygen [136] in 2004. In this project, a pioneering UPS named Cricket [16] was developed to provide centimeter-level accuracy location service. Subsequently, a large number of follow-up works [2,17–19,137–144] have further utilized sound or ultrasound for ranging or localization (see §7.8). However, despite the extensive efforts exerted on UPS research, only a few of the resulting solutions have been actually deployed in practice at present. For example, Cricket project ceased updating 10 years ago. One of the important reasons why UPSs do not receive enough attention today is that they suffer from a serious defect; that is, today's smart devices lack ultrasonic sensors, and therefore, cannot receive ultrasonic beacon signals from UPSs [17–19]. To bridge this gap, many follow-up works [17–19] have advised transmitting beacon signals at the spectrum of $20 \sim 22$ kHz, where 22 kHz is the upper bound of currently available microphones. The sound within this band is supposedly inaudible to the majority of humans. Nevertheless, it is still harmful to infants and pets who are hypersensitive to the high-frequency sounds. Moreover, the signals near the upper bound may be seriously distorted and attenuated due to the non-ideal transition band of the low-pass filter in the microphone.

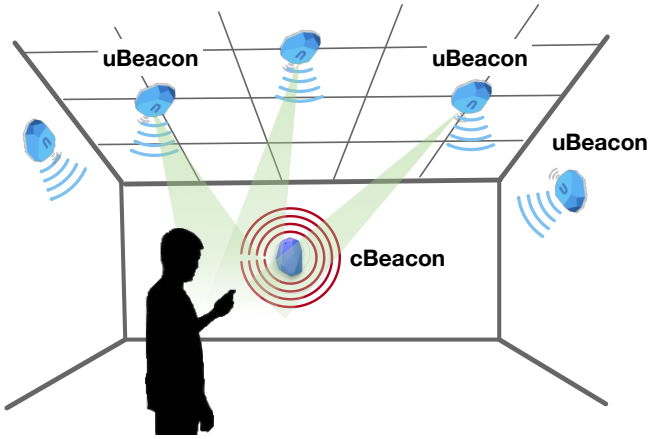


Figure 6-1: System architecture. Multiple uBeacons are deployed as location anchors for the trilateration, whereas a single cBeacon is installed for ultrasonic downconversion.

We present UPS+, an enhanced UPS that can provide sub-centimeter indoor localization for current smart devices and operate exclusively at ultrasonic spectrum, which is considerably beyond the hearing system of humans and pets. Our key innovation is the *renewed efforts* on promoting traditional UPSs to become audible to ultrasound-incapable receivers. A recent finding shows that a combination of two ultrasounds at two frequencies (e.g., f_1 and f_2) may get shifted to a lower differential frequency (e.g., $|f_1 - f_2|$) when they arrive simultaneously at a microphone [59, 70, 71]. Toward this hardware property as a natural downconversion approach, we downconvert ultrasonic beacon signals to the audible spectrum, which the receiver can process. To this end, UPS+ adopts a *heterogeneous architecture* that consists of two types of custom-made beacon devices: uBeacon and cBeacon, as shown in Fig. 6-1. They broadcast beacons at two *ultrasonic frequencies*, thereby ensuring that the beacon signals will not disturb humans or pets. In particular, a large number of battery-supplied uBeacons are deployed as location anchors, whereas a single cable-powered cBeacon is installed for the ultrasonic downconversion. Finally, periodic beacons from uBeacons are downconverted by the beacon signals from the cBeacon. After capturing the beacon signals from at least four uBeacons, the receiver computes the time of arrival and then locates itself via trilateration.

The nonlinearity effect has been verified and demonstrated in various acoustic attacks [59, 70, 71]. However, whether such property can be used for indoor localization remains unclear, unless the following three concerns are addressed:

• *How can we deal with the frequency-selectivity?* Several copied ultrasounds with the same frequency but propagated along different paths may induce constructive or destructive interference at the receiver, thereby leading to the frequency selectivity issue. The use of broadband acoustic signals (i.e., chirps) is considered as an effective means to address this challenge [2, 114, 142, 145]. However, a uBeacon is made of an ultrasonic transducer and only has a bandwidth of 2 kHz, which is barely adequate to deal with the selectivity. In this work, we propose the technique of *dynamical chirp spread spectrum* (DCCS) technique, which spreads the single-tone pulses of the uBeacons over the cBeacon's broadband chirps.

• *How can the beacons advertised from multiple beacon devices be distinguished?* Trilateration requires the receiver to acquire beacons from at least four uBeacons; thus, multiple beacon access is an unavoidable issue. Frequency division multiple address (FMDA) is typically adopted in previous UPSs. However, it fails in our scenario because each uBeacon sweeps an unpredictable dynamic band that depends on the receiver's location, which causes the downconverted beacons to overlap in the frequency-domain at the receivers. Instead, UPS+ is driven by a two-level multiple access mechanism. The receiver initially locates itself in a large space through the chirp slope of the cBeacon and then decodes the uBeacon's ID from its scheduled beacons.

• *How can the receiving energy of downconverted beacons be enhanced?* The nonlinearity effect is observed at the second-order harmonics, which has an amplitude that is less than the fundamental signals. Consequently, the signal-to-noise ratio (SNR) of the beacon at the receiver side in UPS+ is less than that in a traditional UPS at the same position. We alleviate this problem through the following efforts. On the transmitting side, a complicated custom-made driver circuit is designed to double the output power of the transducer. On the receiving

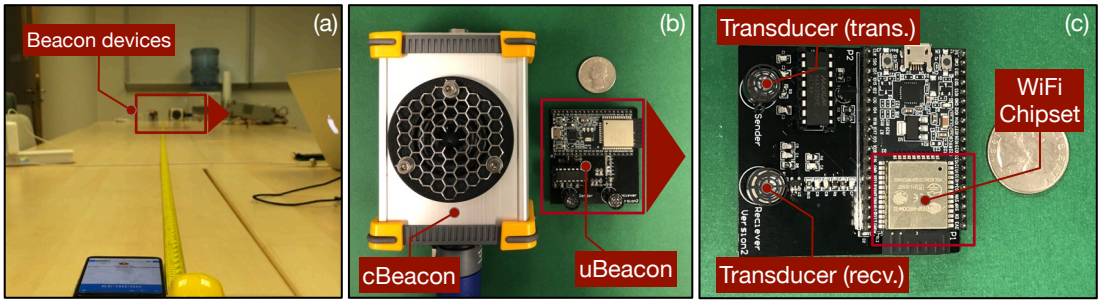


Figure 6-2: UPS+ experimental platform. (a) Experimental scenario; (b) two beacon devices, cBeacon and uBeacon; and (c) zoom-in image of the circuit board of the uBeacon.

side, a noise reduction algorithm is specially redesigned to enhance weak beacons by taking advantage of the dual-microphones in smart devices.

Summary of Results. Our evaluation is performed on 8 types of off-the-shelf smart devices (i.e., five smartphones, an iPad, an iWatch and a pair of AirPods). The results demonstrate that UPS+ can fully utilize the 22 kHz bandwidth for chirps spread on these devices. It performs localization with median and 90^{th} percentile errors of 4.59 cm and 14.57 cm, respectively. Such accuracy matches or even exceeds some previous UPSs. The effective range (median error $\leq$ 20 cm) of uBeacon equipped with three-transducers is around 6 m as equivalent to that of normal UPSs.

Contribution. This work presents UPS+, the first system that operates at an ultrasonic spectrum but remains serviceable in currently available ultrasound-incapable smart devices. The design of UPS+ introduces three key innovations. First, it presents a centralized device called cBeacon to boost previous UPSs, as shown in Fig. 6-2. Second, it uses the nonlinearity effect of microphone systems as an ultrasonic downconverter. Third, it spreads single-tone pulses with dynamic chirps in the air and enhances downconverted beacons at the receiving side. The study also presents a prototype implementation and evaluation of UPS+, thereby demonstrating its accuracy in localizing smart devices in our office.

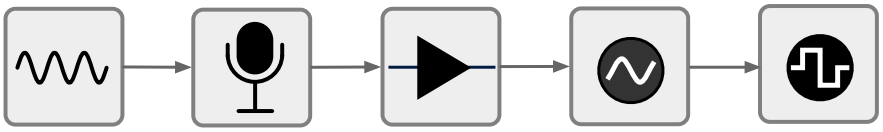


Figure 6-3: Sound processing flow

6.1 Primer on Microphone

Today's smart devices adopt a generic pipeline for sound processing as shown in Fig. 6-3. After being captured by the microphone, sound is first magnified by *amplifiers*. To capture audible sounds, the microphone system is designed to be sensitive to the spectrum of $0 \sim 22$ kHz by using a *low-pass filter* (LPF) to remove sounds that are higher than 22 kHz, even when they are recorded by the microphone. Finally, the sampling rate of the analog-to-digital converter is typically 44.1 kHz, and the digitized signal's frequency is limited to below 22 kHz according to the Nyquist sampling theorem. All modules within the microphone system are supposed to be linear; thus, the output signals are linear combinations of the input. However, as we discussed in Chapter. 3, a real amplifier exhibits *nonlinearity*. In particular, the previous workse [59, 146] show that the acoustic amplifier exhibits strong nonlinearity outside audible frequency range.

To operate the microphone system in an ultrasonic band, we use two off-the-shelf ultrasound speakers to simultaneously play two sounds $S_1(t) = A_1 \cos(2\pi f_1 t)$ and $S_2(t) = A_2 \cos(2\pi f_2 t)$ at frequencies of f_1 and f_2 respectively, where f_1 and $f_2 > 22$ kHz. After magnification, there will be a series of new frequency components (f_1, f_2, $2f_1$, $2f_2$, $f_1 + f_2$ and $f_1 - f_2$) output as shown in Eqn. 3.3. All the ultrasonic frequencies at f_1, f_2, $2f_1$, $2f_2$ and $f_1 + f_2$ are filtered out due to the LPF's cut off at 22 kHz. However, $(f_1 - f_2)$ remains. Translating to actual numbers, when $f_1 = 50$ kHz and $f_2 = 40$ kHz, the microphone will output acoustic signals at 10 kHz. The net effect is that a completely inaudible frequency is recorded by unmodified off-the-shelf microphones, thereby offering *natural down-conversion* for ultrasonic signals. For brevity, we extract the downconverted

sound signal denoted by $S_{\downarrow}(t)$ as our target signals:

$$S_{\downarrow}(t) = \frac{1}{2}a_2 \cos(2\pi(f_1 - f_2)t) \tag{6.1}$$

where a_2 is the attenuation coefficient of second order components. The feasibility and accuracy of the nonlinearity effect on existing microphone systems have been completely verified and demonstrated in recent previous works. Here, we omit to conduct our own verification and refer readers to the results in [59, 70, 71]. Especially, the work [70] tests dozens of devices to show that the nonlinearity is a ubiquitous effect across the currently available smart devices.

Nonlinearity effect has been considered a type of "pollution" or a security "back door" in previous works. Instead, we explore this underlying physical property as a novel and positive *downconverter* to engineer a practical UPS. Such downconverter does not require receivers to be equipped with any ultrasonic sensors. It offers two clear advantages. First, downconversion is generated by the hardware property of electronic amplifiers, and thus, it will not affect humans or animals. Second, the nonlinearity effect can downconvert ultrasonic beacon signals into any frequency below 22 kHz by manipulating two ultrasonic speakers, thereby implicitly providing up to 22 kHz broadband. Subsequently, we will explore this effect to design UPS+.

6.2 System Design

In this section, we present the system design at a high-level in this section.

6.2.1 Design Principles

For sake of brevity, we call the ultrasonic signal *beacon signal* and the devices that can transmit ultrasonic beacon signals, *beacon devices*. Here, we firstly introduce three technical principles that guid our design. First, the design should be backward-compatible with the existing UPS, that is, the past UPSs that are serving the ultrasonic receivers, can be non-intrusively upgraded to work for

Table 6.1: Comparisons of ultrasonic components

Type	Size (cm^2)	Price	BW	Rng.	Supply
Transducer	1.6×1.6	1 \$	2 kHz	8 m	Battery
Speaker	9×12	50 \$	200 kHz	50 m	Cable

the ultrasound-incapable smart devices as well with minimal efforts; second, the design should be absolutely harmless to the hearing systems of human being and pets; third, the design should be resistant with the severe distortions of acoustic signal caused by the multipath effect; finally, the design should be cost-effective and flexible.

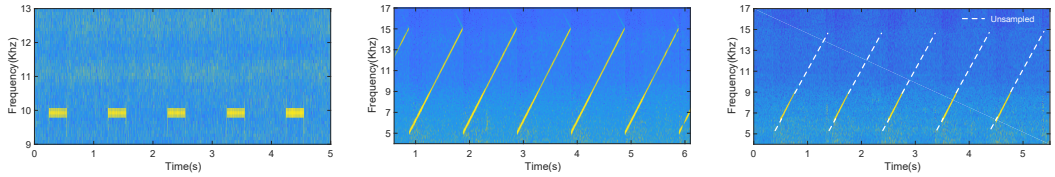
6.2.2 Design Challenge

Our fundamental concept is to deploy many beacon devices in the target space and utilize the nonlinearity effect to downconvert their beacon signals into an audible spectrum that can be recognized by smart devices. One of our engineering philosophies is to utilize commercial off-the-shelf components to build the entire system and to benefit from the economies of scale by being cost-effective. In the present market, two types of commercial ultrasonic components, namely, *transducers* and *speakers*, are available as listed in the Table 6.1.

- **Transducers**: Ultrasonic transducers (aka ultrasonic sensors, e.g., MA40S4S [147]) can convert AC into ultrasound, or vice versa. They are small, low cost (1 \$), and power-saving, but suffer from a limited range (i.e., max at 8 m) and a narrow band ($\sim$ 2 kHz). The majority of previous UPSs (e.g., Cricket) use transducers to build beacon devices.

- **Speaker**: Ultrasonic speakers (e.g., Vifa [148]) are bulky and expensive (50 \$), but have a broadband of 200 kHz (i.e., $0 \sim 200$ kHz) and a propagation range of up to 50 m. These speakers were used as anchors in PC [2].

To take advantage of the nonlinearity, the straightforward design should equip each beacon device with two ultrasonic transducers, which transmit ultrasound at two different frequencies. However, this design is not backward-compatible



(a) Pulse+CW **(b) CW+Chirps** **(c) Pulse+Chirps**

Figure 6-4: Downconverted beacon signals in the frequency-domain. The built-in recorder of an iPhone 8 is used to capture ultrasonic beacon signals with different settings: (a) the uBeacon advertises pulses at 40 kHz whereas the cBeacon transmits a single-tone continuous wave (CW) at 50 kHz; (b) the uBeacon advertises a CW at 40 kHz while the cBeacon transmits continuous chirps from 45 $\sim$ 55 kHz; and (c) the uBeacon advertises pulses at 40 kHz, whereas the cBeacon transmits chirps from 45 $\sim$ 55 kHz.

and cost-effective. A comparison of their characteristics presents an engineering dilemma in component selection; that is, deploying a large number of speakers is extremely expensive, whereas cheap transducers fail to satisfy our broadband demand.

6.2.3 System Architecture

To overcome the challenge, we design a *heterogeneous architecture* that jointly uses the two ultrasonic components by inventing two types of beacon devices.

Such heterogeneous architecture strictly accords with the aforementioned design principles in that: both devices work at absolutely ultrasonic spectrum and harmless to the hearing systems; uBeacons are exactly the same as the beacon devices used in past UPSs. Thus, our design can easily upgrade the past UPSs consisting of uBeacons to our architecture by simply adding a cBeacon; meanwhile, such heterogeneous architecture achieves a trade-off between the cost and effectiveness. Particularly, we can also immediately stop the potential interference to other recording activities (e.g., recording voice using smart phones.) by turning off the single cBeacon, rather than stopping distributed uBeacons one by one; or waking up cBeacon for localization when needed.

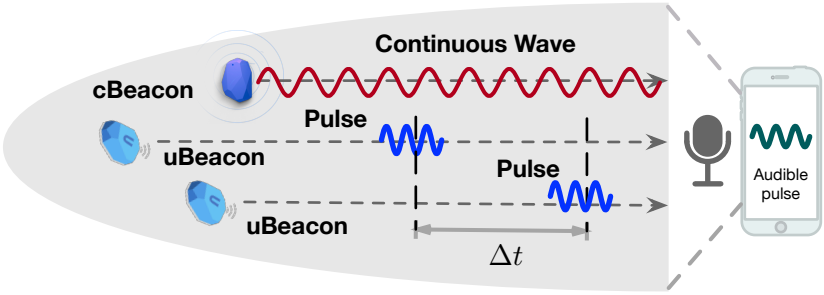


Figure 6-5: Downconversion timing

6.2.4 Timing for Downconversion

We assign two different ultrasonic frequencies to the uBeacon and the cBeacon, and they are deployed in different locations. The receiver can detect a downconverted beacon signal *only* when their ultrasonic beacon signals arrive simultaneously at the receiver. A natural question is how UPS+ synchronizes the arrivals of two types of beacon signals. To do so, the cBeacon persistently transmits a *continuous wave* (CW), whereas the uBeacons advertise short beacon signals every few milliseconds. Fig. 6-5 uses a toy example to explain the design. Given that signals from the cBeacon arrive continuously, downconversion occurs only when a uBeacon's signal arrives at the receiver. In this manner, timing is dependent only on the signal arrival of the uBeacon and irrelevant of the cBeacon. Furthermore, we demonstrate the timing through an actual experiment in Fig. 6-4a, where the uBeacon advertises pulses at 40 kHz, whereas the cBeacon transmits a single-tone CW at 50 kHz. We observe an apparent pattern of pulses at 10 Hz every second similar to the uBeacon's advertising pattern at 40 kHz, although the cBeacon never stops its signals.

6.2.5 Trilateration with ToA in UPS+

Suppose that the i^{th} uBeacon is located at position U_i and transmits a beacon signal at time t_s. The receiver detects the arrival of this beacon signal at time t_i. Thus, the beacon signal takes $t_i - t_s$ seconds to propagate in air. We can

compute the *pseudo-range* as follows:

$$c \times (t_i - t_s) = c \times t_b + |U_i - P| \tag{6.2}$$

where c is the speed of sound, $P(x, y, z)$ is the receiver's location, and t_b is the clock difference between the receiver and the uBeacons. All the uBeacons are assumed to be well synchronized in time and advertise beacon signals at time t_s. c, t_i and U_i are known to the receiver; hence, the preceding equation contains four unknowns, including three coordinate variables in $P(x, y, z)$ and $(t_b + t_s)$. The receiver's OS kernel may introduce additional internal delays due to multi-threading. These delays can be counted to t_b. To solve $P(x, y, z)$, the receiver must receive beacon signals from at least 4 different uBeacons. This approach is called *trilateration*. At a high level, UPS+'s trilateration has the following components.

6.3 Estimation of ToA

The key to trilateration is the estimation of ToA when the beacon signal arrives at the microphone. A naive solution is to allow uBeacons to transmit single-tone pulses. Then, the receiver can estimate ToA by detecting the existence of a signal at the downconverted frequency, as shown in Fig. 6-4a. However, such an approach is typically challenged by background noise, echoes, and the notorious frequency selectivity. The spread spectrum technique is widely recognized as a good solution to conquer these challenges. In this section, we present the unique spread spectrum technique adopted in UPS+ and then introduce the estimation approach.

6.3.1 Spreading Beacon Signals with Chirps

A chirp is a sinusoidal signal with a frequency that increases or decreases over time. The chirp spread spectrum (CSS) is a spread spectrum technique that uses wideband linear frequency chirps to modulate information. CSS has been proven

to be resistant against noise and multipath fading in acoustic channels [2, 114]. Suppose the sampling rate is f_s and t_0 is the time that the first sample is obtained. Then a periodic chirp is defined as

$$S[k] = \cos\left(2\pi(f_0 + \frac{1}{2}(k \bmod K)\Delta f)t_k\right) \tag{6.3}$$

where $t_k = t_0 + (k/f_s)$ and f_0 is the start frequency at time t_0. The transmitter periodically sweeps the spectrum of $[f_0, f_0 + K\Delta f]$. The swept spectrum is divided by K equal intervals in the unit of samples. On the receiver end, ToA can be computed by correlating the received signal with a predefined chirp template. If the correlation spikes in the middle of the reception, then it indicates a match. The position of the spike corresponds to the beginning of the chirp (i.e., ToA).

Spread Beacon Signal in the Air. The straightforward approach is to allow uBeacons to directly yield chirps. However, a chirp signal typically sweeps a wide band (e.g., 5 kHz), which considerably exceeds a uBeacon's bandwidth (i.e., 2 kHz). In UPS+, we allow the band-wider cBeacon to transmit chirps. Let S_u and S_c denote the beacon signals transmitted by a uBeacon and the cBeacon, respectively.

$$\begin{cases} S_u[k] = A_u \cos(2\pi f_u t_k) \\ S_c[k] = A_c \cos(2\pi(f_c + \frac{1}{2}(k \bmod K)\Delta f)t_k) \end{cases} \tag{6.4}$$

where the uBeacon advertises at f_u and the cBeacon periodically sweeps the spectrum within $[f_c, f_c + K\Delta f]$. When the above two equations are substituted into Eqn. 6.1, we obtain the downconverted beacon signal as follows:

$$S_{\downarrow}[t_k] = \frac{1}{2}a_2 \cos\left(2\pi(f_c - f_u + \frac{1}{2}(k \bmod K)\Delta f)t_k\right) \tag{6.5}$$

A comparison between Eqn. 6.5 and Eqn. 6.3 clearly indicates that the downconverted beacon signal is in the form of a chirp, but the sweeping shifts to the downconverted spectrum, i.e., $[(f_c - f_u), (f_c - f_u) + K\Delta f]$. To intuitively under-

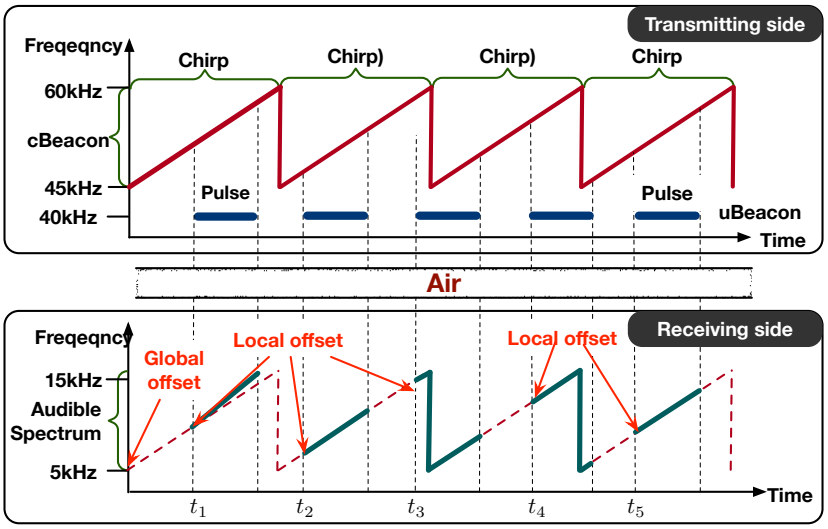


Figure 6-6: Detected dynamic chirp beacon signals. The cBeacon transmits periodic chirps during 45 $\sim$ 60 kHz, whereas the uBeacon transmits a 300 ms pulse beacon signal at 40 kHz every 500 ms. The receiver detects segments of the downconverted chirps.

stand this phenomenon, we present the experimental results in Fig. 6-4b, where the uBeacon advertises a continuous single-tone pulse at 40 kHz and the cBeacon transmits periodic chirps from 45 kHz to 60 kHz. Consequently, the receiver detects intact periodic chirps between 5 $\sim$ 15 kHz. Our technical trick is that *the spreading occurs in the air instead of at the transmitting side, and thereby no additional cost and bandwidth are required at uBeacons.*

Dynamic Chirp Spreading. Now, let us consider what happens when uBeacons periodically advertise *pulse beacon signals*. To illustrate our basic idea, we use the toy example in Fig. 6-6, where the cBeacon transmits chirps continuously but a uBeacon advertises short single-tone pulse beacon signals with spaces. Downconversion occurs during the window when a pulse appears at the microphone; thus, the receiver can no longer detect an intact chirp but only segments of it. In the figure, the green solid lines denote the captured chirp segments. These segments are identical in time and bandwidth, both of which are determined by the pulse interval. However, their starting and ending frequencies may differ, i.e., the spread chirps are *dynamic* and depend on the propagation distance. We call this spreading technique *Dynamic Chirp Spectrum Spread* (DCSS). Fig. 6-4c

shows the downconversion results under the same settings as shown in Fig. 6-4b, except that the signals of uBeacon are changed to periodic pulses. The figure indicates that we can only observe the chirp segments where the unsampled parts are filled with dashed lines.

6.3.2 Pinpointing Dynamic Chirp Beacon Signals

The conventional correlation method which uses a static template, fails in identifying dynamic chirp segments in the present case because the template is not static. In particular, two unknowns, namely, the *global offset* (denoted by Γ) and the *local offset* (denoted by τ), exist due to the dynamic condition. The global offset indicates the starting frequency of the cBeacon's chirp when it first arrives at the microphone. The local offset indicates the starting frequency of the uBeacon's pulse when it first arrives at the microphone. These two types of offset are annotated in Fig. 6-6. Thus, ToA correlation must be performed in two dimensions as follows:

$$\begin{cases} (\Gamma, \tau) = \underset{(\Gamma,\tau)=(0,0)}{\overset{(K,N-K)}{\arg\max}} \frac{1}{K} \sum_{k=\tau}^{\tau+K} \widetilde{S}[k] \cdot G(\Gamma, \tau, k) \\ G(\Gamma, \tau, k) = \cos(2\pi(f_c - f_u + (\Gamma + 0.5(\tau + k) \bmod K)\Delta f)t_k) \end{cases} \tag{6.6}$$

where $\widetilde{S}$ is the received signal and N is the total number of samples in the received signal. N should be the double width of the uBeacon's pulse to ensure that at least one chirp segment is captured. G is the *dynamic chirp template*. The output of the objective function suggests a tuple, which enables the correlation to spike at the appropriate offset. Fig. 6-7 illustrates the correlation results over the sample data shown in Fig. 6-4c.

Optimization. Solving the aforementioned objective function requires performing $N * K$ correlations, which will be a burdensome and energy-consuming task for a mobile device. We notice that once a smart device holds its position, all downconverted segments will share the global Γ, which is caused by the signal propagation delay from the cBeacon. Thus, Γ can be easily found by correlating

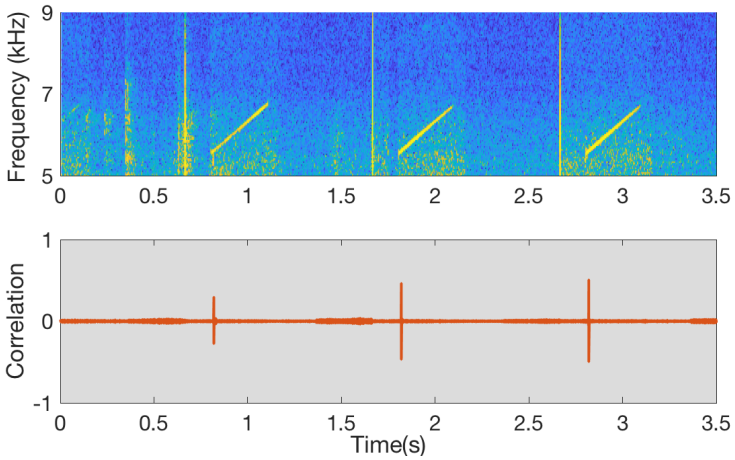


Figure 6-7: Correlation in the time domain. The above picture shows the spectrum via short-time Fourier Transform (STFT); the bottom picture shows the correlation result. The clear correlation peaks could be found exactly at the beginning of each chirp signal.

an intact chirp template with the audio. The correlation spikes at Γ, which allows the chirp template to exactly cover all segments. Then, we start to slide τ to determine the local offset. This optimization process can reduce the number of correlation to $N + K$. Note that although the global offset is involved in the above equation, it is actually not used in the trilateration.

6.4 Multiple Beacon Signal Access

So far, we have discussed how UPS+ can estimate the ToA of a beacon signal out from a single uBeacon. This section discusses how UPS+ distinguishes beacon signals from multiple uBeacons, i.e., *multiple beacon signal access*.

6.4.1 Time Synchronization

Trilateration requires all reference devices (i.e., uBeacons) to be appropriately synchronized in time; otherwise, each unsynchronized uBeacon will introduce an unknown variable to the clock difference, thereby rendering Eqn. 6.2 unsolvable. In this regard, we equip each device with a low-power WiFi chipset and adopt IEEE 1588 PTP for time synchronization. The classic Network Time Protocol (NTP) fails to satisfy our demand because it has an average delay of 10 ms, which

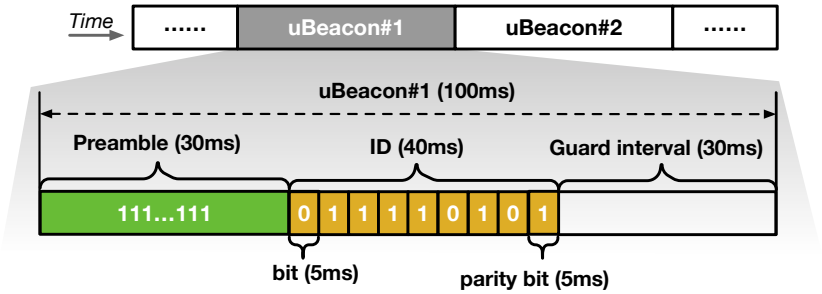


Figure 6-8: Multiple beacon signal access. Beacon signals are advertised in a predefined order to eliminate mutual interference. Each beacon signal contains an ID field to distinguish devices.

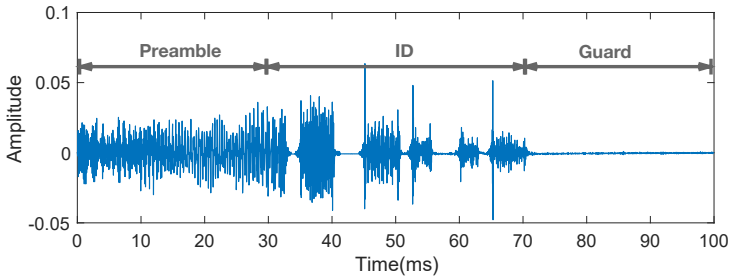
leads to a ranging error of $10\,ms \times 340\,m/s = 3.4\,m$ in UPS+. By contrast, PTP can achieve a high accuracy up to $100\mu s$ [149] or produce a ranging error of 3.4 mm. In UPS+, the receiver does not need to be synchronized with uBeacons because its clock offset is modeled in Eqn. 6.2 already.

6.4.2 Frequency/Time Multiple Access

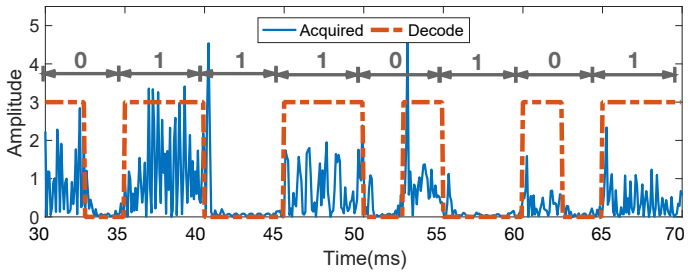
Previous works typically use frequency division multiple access (FDMA) to encode different beacon signals. However, this method does not work in our scenario because the downconverted beacon signal may sweep any segment of the chirp (Fig. 6-6), which is highly correlated with the location of the receiver. To address this issue, we adopt two-level encoding strategies in UPS+.

cBeacon Encoding. UPS+ assigns different *sweeping slopes* to different cBeacons, such that receivers can quickly locate itself at the room level. Since a cBeacon can cover about $50 \times 50\,m^2$ area, a single cBeacon is enough for each room. Multiple cBeacons are isolated through walls. We further apply on-off keying (OOK) to encode the IDs of uBeacons covered by the same cBeacon and schedule them in exclusive time slots.

uBeacon Encoding: Fig. 6-8 shows the encoding of pulse beacon signals. uBeacons advertise their beacon signals in a predefined order to avoid mutual interference. Each beacon signal contains two fields: *preamble* and *ID*. The preamble is composed of 30 ms pulses. The receiver uses the preamble to estimate ToA and to align at the ID field. The ID field contains 8 bits, each of which



(a) Received audio signals



(b) Decoded ID

Figure 6-9: Illustration of beacon signal decoding. (a) The whole beacon signals including the guard interval; and (b) zoom-in view of the ID field after the removal of the chirp carrier.

has an interval of 5 ms. These bits are encoded with FM0, which is also known as Biphase space encoding (For details refer to [150]). Since FM0 requires all bits to be flipped in the beginning, it could avoid the emergence of a continuous 8 bit pulses (i.e., preamble-like ID). The last bit is reserved for parity checking. Thus, UPS+ can completely support 128 uBeacons that are covered by a single cBeacon. An additional 30 ms guard blank is reserved at the end to avoid the interference from echoes. Different uBeacons are scheduled to advertise in various slots. Users are allowed to configure the scheduling based on their practices, e.g., allowing non-adjacent uBeacons to advertise simultaneously to decrease delays; or increasing guard intervals for low-duty cycle.

Beacon Signal Decoding: The receiver initially seeks the beacon signal preamble in the recorded audio data via correlation (see Eqn. 6.6). Once the preamble is obtained, the receivers can identify all the parameters of the chirp from the preamble. Then the audio data are multiplied with the chirp template to remove the chirp-based carrier and acquire baseband signals over the ID field.

Subsequently, '0' or '1' is decoded every 5 ms by determining if a transition occurs during each bit interval (i.e., Bi-Phase Space Coding, FM0). Finally, the receiver uses the decoded ID to determine the corresponding uBeacon's location. Fig. 6-9 illustrates an example of the received beacon signals, from which the ID is successfully decoded.

ToA Estimation. Suppose M preambles are found in the audio data. Let B_i denote the starting position of the i^{th} preamble in the samples where $i = 1, 2, \ldots, M$. B_1 is selected as the baseline and the i^{th} beacon signal's ToA (i.e., t_i) is calculated as follows:

$$t_i = \left((B_i - B_1) \bmod (100\,ms * f_s) \right) / f_s \tag{6.7}$$

The mod operation eliminates the scheduling delay, which is an integral multiple of 100 ms, i.e., uBeacons are scheduled every 100 ms. We also assume that beacon signal propagation does not traverse a scheduling period or beyond $100\,ms \times 340\,m/s = 34\,m$. This assumption is reasonable because a beacon signal propagating over 15 m becomes nearly undetectable. The receiver substitutes t_i into Eqn. 6.2 for the trilateration.

6.5 Enhancement of Beacon Signal

The SNR of the second-order harmonics is weaker than that of the fundamental, which affects the accuracy and the effective range. To mitigate this issue, we introduce multiple enhancement approaches at both transmitting and receiving sides.

6.5.1 Boosting Transmission

Firstly, we boost the transmission at beacon devices.

Transmitter Circuit. The conventional output circuit for the transducer is powered directly from the 5 V supply, i.e., standard input voltage. We firstly use a boost converter to convert the 5 V supply to 12 V DC voltage. Then,

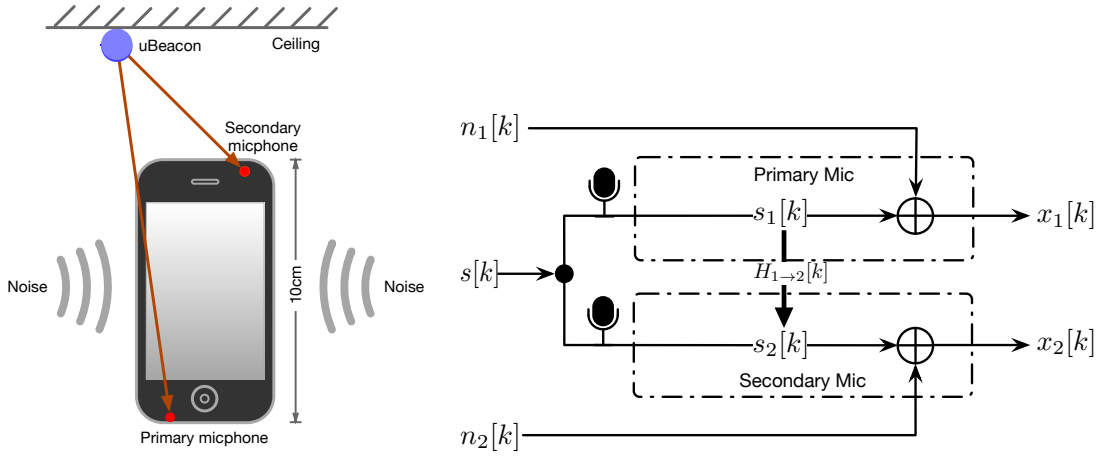
we design a driver circuit to raise and double the voltage. We use an inverter gate to provide a 180° phase-shifted signal to one arm of the driver. The other arm is driven by an in-phase signal. As a result, in spite of 12 V power input, the current from two arms are constructively superimposed at the output. This design doubles the voltage swing at the output and provides about 24 V peak voltage, which is almost twice than the conventional UPS.

Transducer Array. Using an array of transducers is a classical solution to improve the transmitting power of ultrasound. This solution has been adopted in many previous systems. For example, Cricket integrates 2 transducers on a single beacon device, achieving a range of up to 10m. Similar to ours, LipRead [71] utilizes the second-order harmonic for *long-range* voice attack. Although an array consisting of 61 transducers is used in LipRead, it only uses 5 of them for each frequency segment and totally supports 6 segments. Thus, it actually uses 5 transducers for a specific frequency to achieve a maximum range of 30 ft (or 9.1 m).

Boost from cBeacon. The effective range of UPS+ is longer than LipRead because the strength of a downconverted beacon signal (i.e., $A_{\downarrow}$) depends not only on the transmitting power of the uBeacon but also on that of the cBeacon (see Eqn. 6.1) . The transmitting power of the cBeacon is $10,000\times$ stronger than uBeacon, and thereby can enhance the strength of a beacon signal to a relatively higher level. With this in mind, we integrate one transducer in our uBeacon prototype for energy and cost saving. However, it is easy to extend the current prototype to accommodate an array of transducers.

6.5.2 Enhancement at the Reception

Secondly, we design an enhanced algorithm that runs on smart devices. It attempts to draw beacon signals out of background noise when the receiver is far away from the beacon devices. We notice that the majority of modern smart devices (e.g., smartphones) have two microphones, i.e., primary and secondary microphones, as shown in Fig. 6-10a. The primary microphone is typically mounted



(a) Dual-microphone **(b) Dual-channel**

Figure 6-10: Dual-microphone scenario. (a) illustrates a smart device equipped with two mics; (b) shows the common channel model of dual-microphones.

on the bottom to ensure a direct acoustic path from the mouth. The secondary microphone is mounted on top of the device to capture voice with a lower sound pressure level. They are separated by approximately 10 cm, and therefore, receive acoustic signals with different pressures. This condition offers an opportunity to enhance downconverted beacon signals.

6.5.2.1 Modeling Dual-Microphones

Let $s[k]$ denote the clean beacon signals. Then the received signals are given as follows:

$$x_{1(2)}[k] = h_{1(2)}[k] * s[k] + n_{1(2)}[k] \tag{6.8}$$

where $x_{1(2)}$ denote the signals obtained by the primary and secondary microphones, $h_{1(2)}$ are the channel parameters, and $n_{1(2)}[k]$ represent ambient noise. Fig. 6-10b shows the mathematical model. To calculate PSD, we divided the input signal into several *frames* with a constant length (e.g., 600ms). We calculate PSD frame by frame. Let $X_{1(2)}[i, j]$ denote the j^{th} frequency bin of the i^{th} frame for the two microphones. The following is obtained by performing fast Fourier

transform (FFT) on the i^{th} frame:

$$\begin{cases} S_{1(2)}[i,j] = H_{1(2)}[i,j]S[i,j] \\ X_{1(2)}[i,j] = S_{1(2)}[i,j] + N_{1(2)}[i,j] \end{cases} \tag{6.9}$$

where $X_{1(2)}$, $H_{1(2)}$, $S_{1(2)}$ and $N_{1(2)}$ denote the FFT results obtained from the corresponding signals defined in Eqn. 6.8. For clarity, we omit the index numbers unless otherwise noted. S_1 and S_2 are homogeneous because both microphones record the same audio source. Thus, we can introduce a transfer function $H_{1\to2}$ between impulse responses for the two microphones. We can rewrite the equation as follows:

$$\begin{cases} X_1 = S_1 + N_1 \\ X_2 = H_{1\to2}S_1 + N_2 \end{cases} \tag{6.10}$$

By considering both sides with square, the following PSDs are derived:

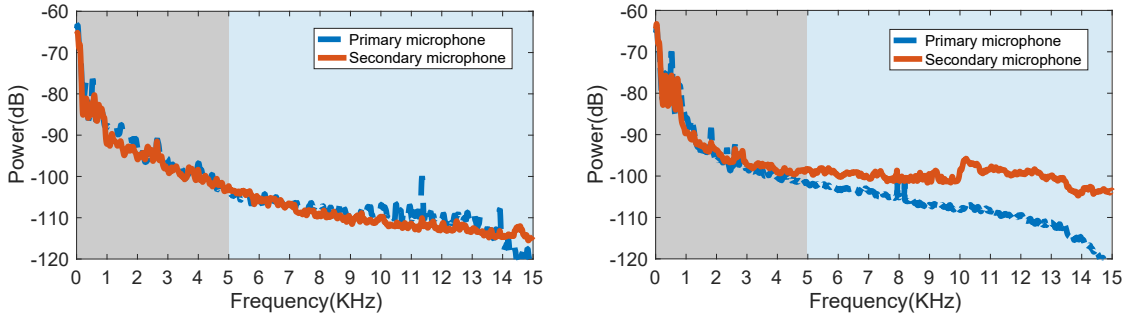
$$\begin{cases} \mathcal{P}_{X_1} = X_1^2 = \mathcal{P}_{S_1} + \mathcal{P}_{N_1} \\ \mathcal{P}_{X_2} = X_2^2 = H_{1\to2}^2\mathcal{P}_{S_1} + \mathcal{P}_{N_2} \end{cases} \tag{6.11}$$

where $\mathcal{P}_{S_1} = |S_1|^2$ and $\mathcal{P}_{N_1(N_2)} = |N_{1(2)}|^2$, $S_1N_1 \approx 0$ and $S_2N_2 \approx 0$. The last two equations are obtained from Wiener-Khinchin theorem, which states that the cross-PSD of the two signals is equivalent to considering FFT in their cross-correlation.

6.5.2.2 Understanding Dual-Microphones

To understand the dual-microphone mode, we conduct experiments to observe the responses of two microphones installed in an iPhone X.

Observation 1. First, we use the two microphones to record ambient noise (without beacon signals) for reference. The power spectral densities (PSDs) of the two audio data are shown in Fig. 6-11a. The signals are nearly identical or homogeneous in terms of PSDs regardless of the locations of the two microphones.



(a) Background noise **(b) Downconverted chirps**

Figure 6-11: Audio PSD at two microphones. The operating spectrum of beacon signals is between $5 \sim 15$ kHz. (a) Two microphones record ambient noise with identical power levels; and (b) two microphones record similar chirps but with different power levels.

This phenomenon is also confirmed in our other experiments, which are conducted under noise-only conditions using the same mock-up phones but in a crowded place or a busy lecture room. The noise in our effective spectrum ($5 \sim 15$ kHz) typically originates from thermal noise and household appliances. These noises are ubiquitously present in a room without definitive directions. Therefore, the signals collected by two microphone are homogeneous.

Observation 2. Second, we show the beacon-present PSDs picked up by the two microphones in Fig. 6-11b. The effective spectrum shows that the power level of the secondary microphone is considerably stronger than that of the primary microphone (20 dB higher) because our beacon devices are mounted on the ceiling. Their signals must travel additional centimeters to arrive at the primary microphone compared with the secondary microphone, thereby resulting in pressure difference. The directivity of ultrasound is stronger than that of audible sound due to its shorter wavelength. Such directivity further magnifies the difference.

In summary, the PSDs of the two microphones are quite similar in background noise, but become differentiable when beacons are present. These two key observations inspire our following design.

6.5.2.3 Parameter Estimation

Now, let us estimate the noise PSD and the transfer function.

Estimating Noise PSD. The first observation suggests that the PSDs of the two microphones over ambient noise exhibit similarity. Thus, we can determine whether any beacon signals are present by comparing the PSDs of the two microphones. If the two PSDs are similar, then the recorded signals should only contain noise. We obtain the PSD difference of the two microphones (denoted by $\Delta\mathcal{P}$) by subtracting two equations in Eqn. 6.11, as follows:

$$\begin{aligned} |\Delta\mathcal{P}_X| =& |\mathcal{P}_{X_1} - \mathcal{P}_{X_2}| = |1 - |H_{12}|^2|\mathcal{P}_{S_1} + \Delta\mathcal{P}_N \\ \approx& |(1 - |H_{1\to 2}|^2)|\mathcal{P}_{S_1} \end{aligned} \tag{6.12}$$

where $\Delta\mathcal{P}_N = \mathcal{P}_{N_1} - \mathcal{P}_{N_2} \approx 0$, i.e., the PSD difference of two noises. The last derivation is obtained from the first observation: $\mathcal{P}_{N_1} \approx \mathcal{P}_{N_2}$. Thus, in the case of noise-only periods, $\Delta\mathcal{P}_X$ approaches to zero because $\mathcal{P}_{S_1} = 0$. In practice, we set a threshold $\delta_{\min}$ for the decision. If $\Delta\mathcal{P}_X \leq \delta_{\min}$, then the estimated PSD of noise at the j^{th} frequency is given by the following:

$$\widetilde{\mathcal{P}}_N[i,j] = \alpha\widetilde{\mathcal{P}}_N[i-1,j] + (1-\alpha)\mathcal{P}_{X_1}[i,j] \tag{6.13}$$

where α is a user-defined learning factor, which indicates how much should be learned from the new round of estimation.

Estimating Transfer Function. We obtain the following by multiplying two equations in Eqn 6.10:

$$X_1X_2 = H_{1\to 2}S_1^2 + \cancel{S_1N_2} + \cancel{H_{1\to 2}S_1N_1} + \cancel{N_1N_2H_{1\to 2}S_1^2} + N_1N_2 \tag{6.14}$$

The beacon signals are assumed to be independent of noise, and thus their *cross-correlation* equals zero. The cancellation is due to such independence. After

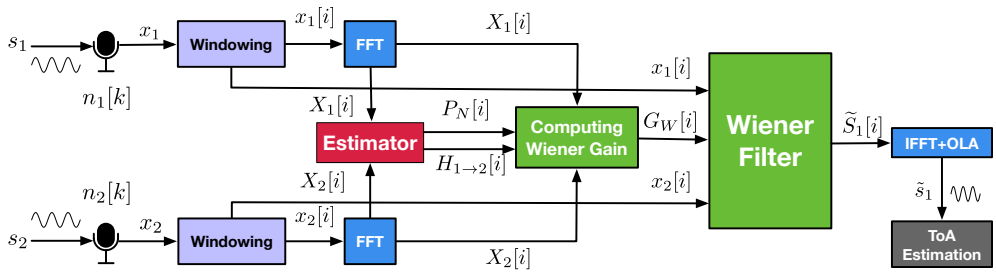


Figure 6-12: Block diagram of dual-microphone beacon signal enhancement algorithm.

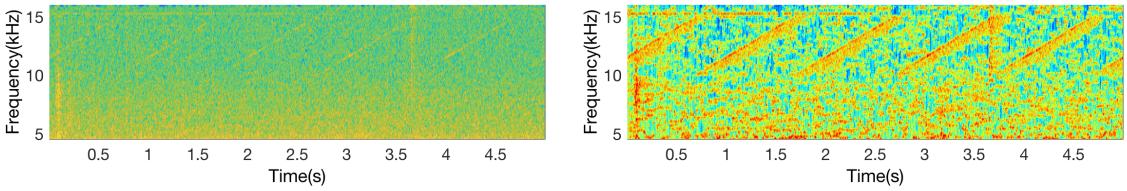
the manipulation, the transfer function can be estimated as follows:

$$H_{1\to2} = \frac{X_1X_2 - N_1N_2}{S_1^2} = \frac{X_1X_2 - \mathcal{P}_N}{\mathcal{P}_{S_1}} = \frac{X_1X_2 - \mathcal{P}_N}{\mathcal{P}_{X_1} - \mathcal{P}_N} \tag{6.15}$$

where X_1X_2 is the calculable cross-PSD. The last substitution is obtained from Eqn. 6.11. In the case of beacon periods (i.e., $\Delta\mathcal{P}_X > \delta_{\min}$), we can estimate the transfer function for subsequnt use.

6.5.3 Enhancement Algorithm

The above two observations suggest that the PSDs of the two microphones are quite similar regarding background noise, but become differentiable when beacon signals are present. These two key observations inspire us to design an enhancement algorithm at receivers. The block diagram of the beacon enhancement algorithm is shown in Fig. 6-12. Intuitively, the downconverted beacon signals are chirp segments whose spectrums are dynamic, that is, the segment moves among different frequency bins. Moreover, the beacon signals are advertised every hundred milliseconds. Both conditions provide idle windows for smart device to estimate the background noise. Specifically, when the PSD difference of two microphones is less than a threshold, the receiver starts to estimate the PSD of noise by assuming that no beacon signal is present currently; otherwise, the receiver estimates the PSD of signals and applies *Wiener filter* to enhance the beacon signals. The gain of the Wiener filter (denoted by G_w) is formally



(a) Before **(b) After**

Figure 6-13: Enhancement results. The two figures show the beacon signals before and after enhancement.

computed as follows:

$$G_W = \frac{\mathcal{P}_S}{\mathcal{P}_S + \mathcal{P}_N} = \frac{|\Delta\mathcal{P}_X|}{|\Delta\mathcal{P}_X| + |(1 - |H_{12}|^2)\mathcal{P}_N|} \tag{6.16}$$

Substituting Eqn. 6.12 into the preceding equation obtains the final derivation. In particular, $|\Delta\mathcal{P}_X| = |X_1^2 - X_2^2|$ (i.e., the difference between two microphones' PSDs), $\mathcal{P}_N$ is estimated using Eqn. 6.13, and $|H_{1\to2}|$ is estimated using Eqn. 6.15. Multiplying the input signals with G_W outputs the enhanced beacon signals. We omit the introduction of the filter due to space limitations and advise readers to refer to [151, 152] for details. Fig. 6-13 shows the effectiveness of the algorithm in an extreme case where the receiver is located 15 m away from the uBeacon. The figure shows that the enhanced chirp signals become distinguishable although they are nearly drowned in background noise before filtering.

It is worth noting that the above enhancement algorithm is an additional helpful functionality that can improve the detection range by about 1.5×. In practice, there must exist a special case that the two microphones receive beacon signals with same strengths, leading to a gain of 1. In this case, the signal cannot be enhanced. As a consequence, the system can work well only within the effective rang as the receiver does not use the algorithm. In other words, the algorithm would never worsen the localization performance even when it fails to work in the special case. It is very similar to the assistant GPS (A-GPS) which allows GPS receivers to obtain additional information from network like GSM to better assist in satellite location. However, GPS receiver still work in the most of cases

without A-GPS.

6.6 Implementation

This section presents the implementation of two types of beacon devices. Their prototypes are shown in Fig. 6-2.

• **uBeacons**. The uBeacons provide a functionality similar to those of devices designed in Cricket [16]. A simple solution is to directly apply their designs to our system. However, the Cricket design was released ten years ago and many of its components are already outdated. Thus, we have to design our own hardware. The core function is to drive an on-board ultrasonic transducer (MA40S4S [147]) to speak at 40 kHz at a specific time point. We also reserve an ultrasonic receiver for testing, which is not required for UPS+. The low cost WiFi chip ESP32 [153] from Espressif systems is integrated for the PTP protocol and online configuration. We synchronize uBeacons every $32s$ for energy saving where the beacon broadcast is advised between $0.25 \sim 64$ seconds. Each uBeacon is supplied by two 5 V-batteries.

• **cBeacon**. We use a general vector signal generator to produce periodic ultrasonic chirps, which are further transmitted to the air through an ultrasonic speaker (Vifa [148]). The output power is tuned to 10 W. It sweeps the 10 kHz band from 48 kHz to 58 kHz over 100 ms, i.e., 100 kHz per second. As a result, the downconverted spectrum is between $8 \sim 18$ kHz regarding to the 40 kHz of uBeacons. We intentionally reserve the top 4 kHz bandwidth (i.e., $18 \sim 22$ kHz) to avoid the signal distortion due to the non-ideal transition of the low-pass filters at microphones.

6.7 Results

A total of 15 uBeacons are deployed in a meeting room with an area of 9×3 m^2 in our department. The speed of sound is set to 344.38 m/s in our experiments. The speed of sound is dependent on the temperature, thus a calibration before

the experiments is performed. To test as many smart devices as possible and in a cross-platform manner, we directly use a built-in recorder of a mobile OS (Operating System, e.g., iOS or Android) to record audio clips for post-processing in MATLAB. We use uBeacons equipped with a single transducer by default and an iPhone X as the default receiver unless noted.

6.7.1 Evaluation in Zero-Dimension

We start by qualitatively comparing UPS+ with state-of-the-art UPSs and evaluate the accuracy of ToA estimation for static locations.

6.7.1.1 Comparison with State-of-the-Art

We compare UPS+ with past UPSs from nine different perspectives, as listed in Table 6.2. (1) Only a few UPSs use ultrasonic speakers as anchors due to their high prices. PC is an attempt in this direction. It achieves a good accuracy at an extremely higher cost. (2) Unlike normal UPSs, Dolphin deploys custom-made ultrasonic receivers to locate transmitters, but has a shorter range (i.e., 3 m); (3) ALPS designs an embedded ultrasonic speaker, which can achieve the operating range of 40 m. Due to the lack of hardware, we present the mean accuracy in the table only. (4) both BeepBeep and ApneaApp use the speakers in the smart devices for ranging. (5) TUPS is the UPS that we implemented across uBeacons but operates at the fundamental frequency. TUPS orientates to ultrasonic receivers, being similar to the ALPS, Dolphin and Cricket. Thus, we choose TUPS as a benchmark baseline in our evaluation. Particularly, we have the following observations: (a) Previous UPSs use either ultrasonic transducers or speakers, whereas UPS+ is a unique system that uses the two components jointly to build a hybrid UPS; (b) UPS+ achieves comparable accuracy and effective range as transducer-built UPSs, and meanwhile the unit cost is maintained at an acceptable level; (c) UPS+ is the unique system that operates at the ultrasonic spectrum, but remains compatible with current smart devices.

We also list the results of other four typical RF-based solutions for comparison. Usually, RF-based solutions have longer operating range (e.g., around

Table 6.2: Comparison to Past Ultrasonic UPSs

Scheme	Type	Accuracy	Range	Cost	Dim.	Spectrum	Modulation	Inaudi	Compt.[1]
Cricket [16]	Transducer	10 cm	10 m	10 $	2D	40 kHz	Pulse	Yes	No
Dolphin [142]	Transducer	2.34 cm	3 m	10 $	3D	$20 \sim 100$ kHz	DSSS	Yes	No
PC [2]	Speaker	4.3 cm	50 m	50 $	3D	$19 \sim 23$ kHz	Chirp	No	Yes
BeepBeep [19]	Speaker	0.8 cm	5 m	–	1D	$2 - 6$ kHz	Chirp	No	Yes
ApneaApp [154]	Speaker	2 cm	1 m	–	1D	$18 - 20$ kHz	Chirp	No	Yes
ALPS [155]	Transducer	16.1 cm	40 m	> 10 $	3D	$20 \sim 21.5$ kHz	Chirp	No	No
TUPS	Transducer	3.51 cm	8 m	10 $	3D	40 kHz	Pulse	No	No
UPS+	Hybrid	4.95 cm	6 m	10 $[2]	3D	$40 \sim 65$ kHz	DCSS	Yes	Yes
Tagoram [79]	RFID	8 mm	12 m	0.1 $	3D	820 MHz	ASK	No	No
ArrayTrack [156]	WiFi	57 cm	100 m	–	3D	2.4 GHz	FDMA	No	No
iBeacon [157]	Bluetooth	1 m	150 m	25 $	1D	2.4 GHz	DCSS	No	Yes
WiTrack [135]	–	13 cm	–	–	3D	$5.46 \sim 7.25$ GHz	FMCW	No	No

[1] The column of 'Compt.' indicates if the solution is compatible with today's smart devices.
[2] The cost in UPS+ does not contain the price of cBeacon, each of which cost about 60$ and is shared by multiple uBeacons.

100 m), but their accuracies are limited to meter or sub-meter level. RFID-enabled Tagoram behaves a good accuracy but requires the prior knowledge of track. Importantly, UHF RFIDs are unavailable by smart devices due to lack of UHF readers. Bluetooth-enabled iBeacon might be the most widely used commercial localization technology, which was initiated by Apple Corp. Our real experiments show that the average accuracy of an iBeacon is around one meter. WiTrack locates targets at cm-level using RF reflections at the cost of almost 2 GHz bandwidth. Compared against RF solutions, our work offers a trade-off solution between the practicality and the accuracy to the currently available smart devices and to meet the growing demand on indoor high precision localization.

6.7.1.2 Accuracy in ToA Estimation

As ToA estimation is the foundation of the trilateration, we evaluate the accuracy of ToA estimation. In the experiment, we place seven different uBeacons 2 m away from the receiver while holding the receiver at a static location (i.e., in zero-dimension). The uBeacons advertises beacon signals every 5 s. We record a 10 min audio across these uBeacons. Given that the distances between the receiver and the uBeacons remain unchanged at all times, the downconverted beacon signals should arrive exactly every 5 s. We compute the ToA of the remaining beacon signals by considering the ToA of the first beacon signal as the reference. These relative times should be an integral multiple of 5 s. Fig. 6-14 shows the errors of the estimated ToA across the seven uBeacons.

The result suggests that approximately 0.03 ms error occurs in the estimation

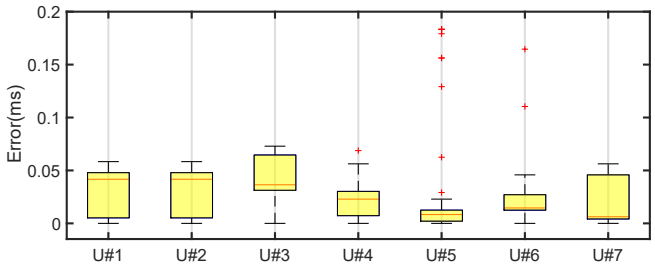


Figure 6-14: ToA estimation

even if the receiver remains at its location, mainly due to the synchronization time delay and the additional time consumption for processing audio data at the receiver, e.g., moving data from an audio capture system to a memory or disk. Such estimation delay will incur a potential 0.03 ms $\times$ 344.38 m/s = 1.03 cm error in the subsequent ranging or localization results. Nevertheless, 0.03 ms error is rather stable because the maximum standard deviation is around 0.02 ms. This experiment demonstrates the feasibility of using downconverted beacon signals for localization.

6.7.2 Evaluation in One-Dimension

Then, we evaluate the ability of UPS+ in one-dimension, that is, both the uBeacon and the cBeacon are fixed at their positions. The receiver is moved away from the uBeacon by following a straight line. Their distance is increased from 40 cm to 800 cm with a step interval of 10 cm. We are interested in determining UPS+ performance as a function of distance.

6.7.2.1 Accuracy in Ranging

In acoustics, the sound pressure decreases with the reciprocal of the distance from the sound source. In our case, the power of beacon signals attenuate linearly as the distance increases. As a result, the signal-noise-ratio (SNR)s of beacon signals decrease when the target device goes away from the beacon devices. Thus, the distance would affect the ranging accuracy directly. We initially investigate UPS+'s ranging accuracy as a function of the distance between the beacon device and the receiver.

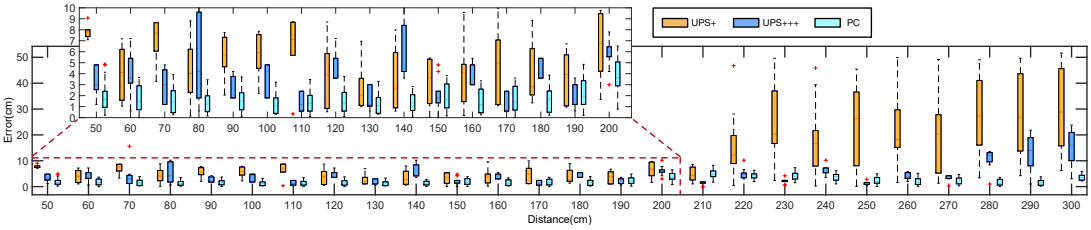


Figure 6-15: Ranging accuracy as a function of distance. UPS+ integrates a single transducer on a uBeacon; UPS+++ integrates three transducers on a uBeacon; PC refers to [2].

We compute the pseudo-ranges when the receiver is located at different positions. Adopting the pseudo-range at 40 cm as the reference, the displacements relative to the reference in other positions are calculated. The experimental trials are conducted 20 times in each position. The *effective range* is defined as the maximum distance when the median ranging error is less than 20 cm. The ranging accuracy is shown in Fig. 6-15, which only shows the results when the distance is less than 300 cm due to the space limit. We have the following findings:

• **Single-transducer made uBeacon**: Firstly, we equip a single transducer on the uBeacon. UPS+ achieves a mean accuracy of 4.85 cm and a standard deviation of 2.8 cm when the distance is less than 2.5 m. However, the mean error increases to 22.61 cm when the distance is beyond 2.5 m. The effective range is about 3 m. This is due to the facts: the transmitting power of a uBeacon is limited to 0.9 mW, and the adoption of second-order signals suffers from a larger attenuation compared with that of the first-order signals. Both factors decrease the SNR of downconverted beacon signals, thereby reducing the accuracy, when the receiver is away from the uBeacon.

• **Three-transducer made uBeacon**: Secondly, we equip three transducers (i.e., UPS+++) on the uBeacon to form a simple ultrasonic array. UPS+++ has a mean accuracy of 3.6 cm when the distance is less than 3 m. Correspondingly, the effective range is extended to 6 m. This result almost reaches the range of TUPS, which operates at the first-order signal but is equipped with a single transducer. The cost is only increased 3 $ in comparison to TUPS. This suggests that multiple transducers can indeed enhance the transmitting power of ultrasound.

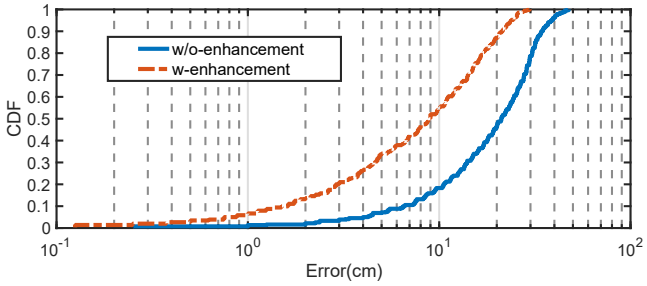


Figure 6-16: Enhancement

• **Speaker made beacon device:** Thirdly, another simple approach for lengthening the ranging distance is to increase the transmitting power similar to PC, which uses the ultrasonic speaker to transmit chirps at $19 - 23$ kHz. The transmitting power is tuned at 10 W, which is $10,000\times$ that of our uBeacon. Consequently, the ranging accuracy remains at 4.3 cm even if the receiver is located at a distance of 6 m.

In summary, UPS+ can achieve centimeter-level ranging accuracy and works comparably with ultrasonic ranging within an effective range of approximately 3 m, when uBeacons are equipped with single-transducers. We believe that 3 m range can fulfill major demands in practice, particularly when uBeacons are attached to ceilings or walls. Of course, we can increase the ranging distance by enhancing the transmitting power or integrating multiple transducers if necessary.

6.7.2.2 Accuracy after Enhancement

We evaluate the effect of the PSD-aware enhancement algorithm at a smart receiver equipped with two microphones. Android smart-phones are allowed to select audios through two microphones by turning on the stereo mode. We place a Huawei smart phone at a distance of 5 m. Fig. 6-16 shows the comparisons of ranging accuracy with and without enhancement. The median error *without* enhancement is approximately 21.13 cm due to the confusion from background noise. Intuitively, a 50-sample shift will incur a $0.1\,ms$ time shift or $27\,cm$ error in ranging. Dozens of samples were obscured by the noise in this case. To improve SNR, we apply the enhancement algorithm to ToA estimation. Consequently,

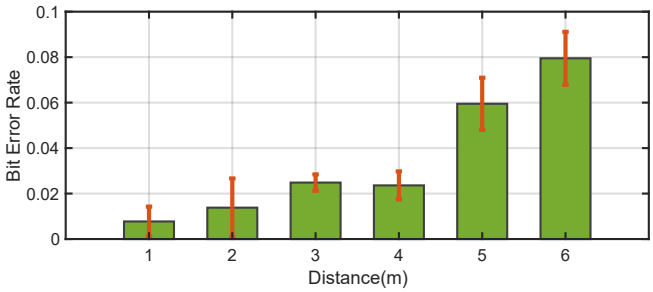


Figure 6-17: Bit error rate

the median error rapidly drops below 9 cm, which outperforms the result twice. This is because the enhancement can improve the energy of the correlation peak by a factor of three more than that without enhancement.

6.7.2.3 Accuracy in Decoding

Subsequently, we evaluate the UPS+'s decoding ability. We skip the identification test of the cBeacon because the results are always accurate regardless of how the cBeacon's sweeping slope is changed. The main reason for this condition is that transmitting signals of the cBeacon is relatively strong ($10,000\times$ that of the uBeacons), which results in easy decoding. Here, we only focus on the decoding of uBeacons' IDs. Fig. 6-17 shows bit error rate (BER) as a function of the distance. In contrast to ranging, UPS+ exhibits a strong decoding ability. BER remains below 8% even when the receiver is moved to a distance of 5 m.

A considerable difference exists between ranging and decoding because decoding has a loose requirement for preamble alignment. In UPS+, each bit has an interval of 5 ms. Our study implies that a bit can still be decoded even when a misalignment of 2 ms exists (i.e., involving 40% of the samples) because energy accumulation across 60% of the samples for a bit is sufficient for bit decision. By contrast, a misalignment of 2 ms causes a $344.35 \times 0.002 = 68.87$ cm error in ranging. Background noise can easily obscure a few samples in the preamble.

6.7.3 Evaluation in Two-Dimension

In the following section, we evaluate UPS+ in two dimensions with respect to its localization accuracy.

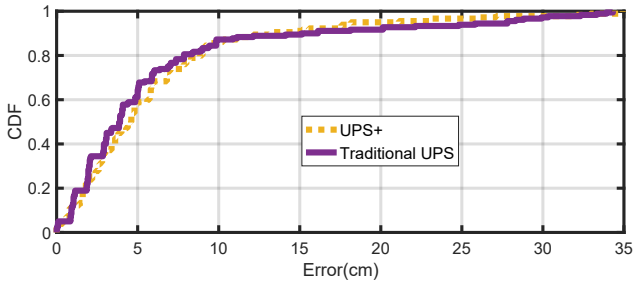


Figure 6-18: Localization accuracy

6.7.3.1 Accuracy in Localization

We compare UPS+'s localization accuracy with that of TUPS, which uses ultrasonic sensors operating at ultrasonic bands as its receiver. We perform 150 experimental trials. In each trial, the receiver is placed randomly in the evaluation environment. Fig. 6-18 plots the CDF of the 2D localization error. We observe the followings:

• UPS+ achieves a median accuracy of 4.59 cm and a 90^{th} percentile accuracy of 14.57 cm in 2D localization. By contrast, the median and 90^{th} percentile accuracy of the TUPS are 3.51 cm and 15.13 cm, respectively. These results are consistent with Dolphin's implementation [142, 158] which also reports a median localization accuracy of 3 cm. This result demonstrates that UPS+ can achieve nearly the same accuracy as the traditional UPS even if it works at the second-order harmonics. Interestingly, the 2D localization accuracy of UPS+ behaves slightly better than that under 1D condition, because trilateration in 2D uses the difference in ToA. Many uncertain common variables (e.g., sync delay at uBeacons or audio processing at the receiver) will be canceled out from the difference.

• UPS+ and Dolphin achieve more than 10× improvement for the mean accuracy over Cricket [16, 159], which reports a mean accuracy of 30 cm. This is due to the rapid development of hardware, which provides considerably higher resolution in sampling frequency compared with that in a decade ago.

In summary, adopting the second-order harmonic as localization media can achieve the same accuracy as that with the first-order. UPS+ even exceeds some past UPSs.

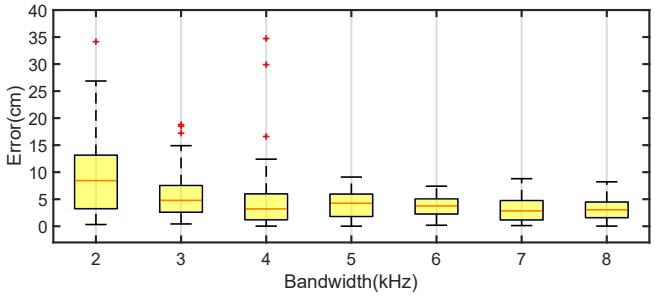


Figure 6-19: Error vs. bandwidth

6.7.3.2 Impacts of Parameters

Next, we evaluate the localization accuracy as a function of different system parameters:

Impact of Bandwidth. We evaluate UPS+'s localization accuracy as a function of bandwidth. The cBeacon sweeps the spectrum with a constant slope. Thus, increasing the length of preambles of the uBeacons' beacon signals is equivalent to increasing the bandwidth of the downconverted beacon signals. Fig. 6-19 shows the impact of bandwidth on localization accuracy. The plot demonstrates that the accuracy monotonically improves with increased bandwidth. In particular, if UPS+ uses a 2 kHz bandwidth, then the median error reaches 13.13 cm and rapidly drops below 10 cm for bandwidths larger than 4 kHz. This result is attributed to increased bandwidth, which provides finer granularity in separating the LOS path from echoes. In our default setting, 4 kHz is adopted to balance time delay and accuracy.

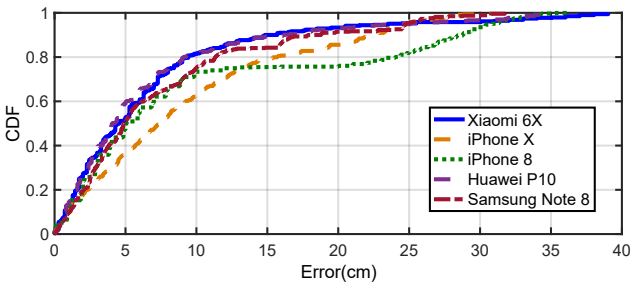


Figure 6-20: Accuracy vs. Manufactures

Impact of Receiver Manufactures. The nonlinearity effect of amplifiers in

a microphone system is a key in UPS+ for ultrasonic downconversion. A natural question is whether this effect can be applied to different models of receivers. To this end, we test 2D accuracy across four typical smart-phones from the top four manufactures, i.e., Apple, Xiaomi, Huawei and Samsung. They hold 10.2%, 9.7%, 17.7% and 22.9% market shares by 2018 respectively according to the latest report from IDC [160]. Totally, they occupy almost 60% market shares. The Fig. 6-20 shows the test results. From the figure, the median errors of the five devices are 4.42 cm, 7.39 cm, 5 cm, 4.28 cm and 4.9 cm. A slight difference (< 3 cm) exists among these devices, which demonstrates that the nonlinearity effect exists as a ubiquitous phenomenon in smart phones.

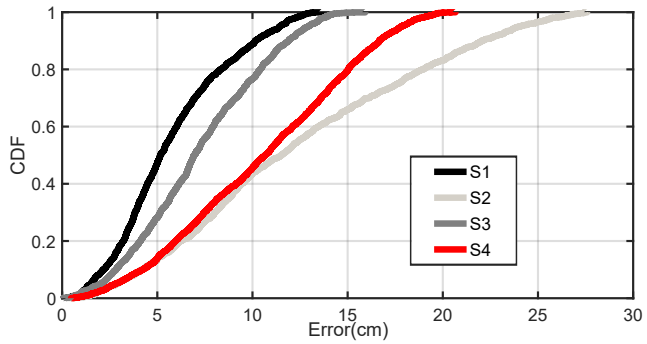


Figure 6-21: Accuracy vs. environment

Impact of Environment Settings. We repeat the experiments and evaluate the localization accuracy in four different environmental settings: (S1) in a quiet office room; (S2) in a noisy office full of loud music played using external speakers; (S3) in a noisy lecture room where over 50 students are participating in a discussion; (S4) in a 5×5 m^{2}small room full of echoes. We believe these settings can cover the majority of the cases in practice. The localization results are shown in Fig. 6-21. The mean accuracies of the four settings are 5.15 cm, 6.89 cm,10.63 cm and 11.33 cm respectively. We could see that there are about 5 cm changes in the median accuracy. The worst case happens when playing music which might contain some high-frequency items. Since the frequency band of human voice are below 5 K, the localization accuracy holds at a high level even 50 people move and talk in the room. In particular, UPS+ has a strong

ability to fight off the echoes in narrow space due to the DCSS technique. It can be seen from the experiments UPS+ is rather robust with respect to different environment settings.

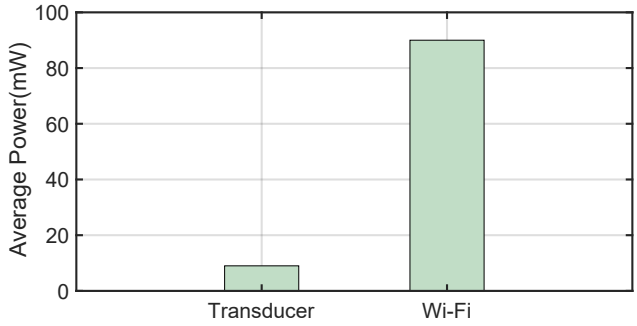


Figure 6-22: Energy vs. components

6.7.3.3 Energy Consumption

Finally, we present the energy consumption of a uBeacon with respect to components and cycles respectively. Fig. 6-22 shows the accumulated energy of the two main components, which are persistently consumed during one second. It can be seen that the WiFi communication for the PTP sync consumes 4 times energy than the transducer. The results show that wireless transceiver is indeed an energy hog in smart devices, which is consistent with many previous reports. Next, we estimate the actual energy consumption during one second of the broadcast cycle (for localization) and sync cycle (for clock synchronization). The results are shown in Fig. 6-23. It is clear that the low-duty cycle can save much more energy, especially when considering PTP sync. Specifically, a uBeacon can save 58% energy when adjusting its sync cycle from five seconds to one minute. With respect to our current settings (i.e., 1 s sync and 3 s broadcast), each uBeacon can work about 5 months. If the PTP sync cycle is increased to 64 s, the life can be prolonged to 8 months.

6.7.4 Evaluation in Three-Dimension

Finally, we envision three practical applications which can benefit from the high accuracy of UPS+. They qualitatively test 3D localization accuracy in several

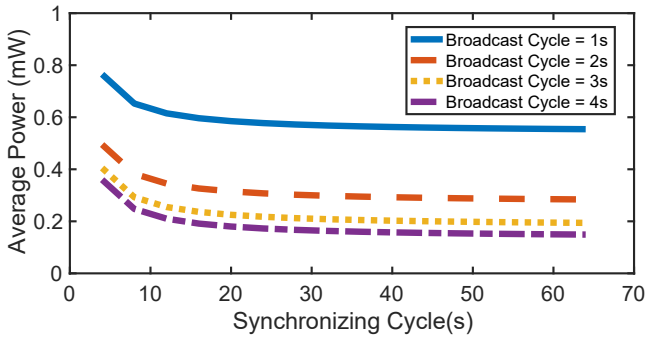


Figure 6-23: Energy vs. cycle

real-world applications, as shown in Fig. 6-24.

(1) **Finding Wireless Headphone**. This application shows to find a pair of AirPods (Ver. 1). Each AirPod integrates two microphones, but we are allowed to access the bottom one only. Its sampling frequency is limited to 16 kHz. For the better performance, we correspondingly adjust the chirp of cBeacon to ensure the downconverted spectrum falls into $2 \sim 7$ kHz.

(2) **Pairing Two Tablets**. UPS+ could provide fundamental location service to VR-system. This application enables to pair two iPads once their distance is less than 10 cm.

(3) **Tracking Hand-Arm**. We apply UPS+ into tracking a human's hand-arm through an Apple Watch (Ver. 3). The watch's highest frequency is up to 22 kHz. However, we find that the built-in voice recorder uses M4A format, which filters the frequencies above 16 kHz. This is also an important reason why we set the downconverted spectrum to $5 \sim 15$ kHz, which can adapt to many kinds of formats.

6.8 Related Work

This work touches upon many topics related to indoor localization, nonlinearity effect and noise reduction. These areas already have large bodies of research; thus, our review focuses primarily on closely related works.

(a) Sound-based Indoor Localization: Sound-based indoor localization

Figure 6-24: UPS+ in real-world applications. (a) shows how UPS+ enables finding the wireless headphone (e.g., AirPods); (b) shows how UPS+ enables pairing two tablets (e.g., iPads); (c) shows how UPS+ enables tracking hand arm through a wearable (e.g., Apple Watch).

can be classified under broad categories of range-based [2, 16, 18, 19, 142–144, 154, 155, 158, 161, 162] and range-free [132, 133, 141, 163, 164, 164–166] methods. Here, we focus only on range-based methods. We refer to [167] for a comprehensive survey on various indoor localization techniques. This category of solutions computes distances based on how long sound takes to propagate between a sender and a receiver [2, 16, 18, 19, 142–144, 154, 158, 161, 162]. For example, Active Bats [161] and Cricket [16] are pioneering range-based localization systems that use ultrasonic beacons. Dolphin [142] presents a new design for ultrasonic transmitters and receivers. BeepBeep [19] uses the same range approach to estimate the distance between two cellular phones. In [17], the authors attempt to identify the location of a mobile phone in a car using vehicle-mounted audio speakers. ApneaApp [154] uses an estimated range to track human breath. Pulse Compression(PC) [2] is the closest to our work but it uses 19 $\sim$ 20 kHz band, which may still be sensitive for infants or pets. By contrary, our beacons are advertised at 50 kHz or higher, which results in zero noise pollution to indoor lives. Moreover, our work uses dynamic chirps (§6.3) to spread the spectrum instead of the static chirps used in [2]. In summary, previous works either require ultrasonic receivers or advertise audible beacons. Additional comparisons are provided in §7.7.

(b) Nonlinearity Effect: Nonlinearity has been explored for many purposes [59, 70, 71] in recent years. Backdoor [59] constructs an acoustic (but in-

audible) communication channel between two speakers and a microphone over ultrasound bands with the nonlinearity effect. DolphinAttack [70] and LipRead [71] utilize the nonlinearity effect to send inaudible commands to voice-enabled devices such as Amazon Echo. Our system is inspired by these previous works. However, we use this effect for a different purpose and face challenges that vary from those in previous studies.

(c) Noise Reduction: Finally, several works [152, 168, 169] from the speech field have motivated us to design a dual-microphone enabled enhancement algorithm. In contrast to previous works, we use the features of beacon signals to determine their absence.

6.9 Discussion

In this chapter, we introduced UPS+, a CTM technique that demonstrated, for the first time, that ultrasound positioning system can benefit the ultrasound-incapble smart devices at actual high-frequency band without interfering the animals. Here, we further discuss the potential interference to other acoustic system and the practical deployment solution.

6.9.1 Interference to Voice Communication

One concern might be that the automatic downconversion at microphone may affect the voice communication when people make calls. Actually, majority of the voice communication systems (VCS) in phones are equipped with the noise suppression modules, which aim to enhance the quality of voice less than 8 KHz and filter the noises above 8 KHz [170]. To verify our theory, we conduct a group of comparative experiments in which Alice builds a voice communication with Bob through the GSM (Global System for Mobile Communications), WeChat and Skype respectively. The two instant messaging (IM) apps are connected through the Internet. At the Alice's side, we play a chirp sound sweeping from 10 kHz to 24 kHz (i.e., the horizontal line in Fig. 6-25). The signal is transmitted to Bob through the above three means respectively. We show the spectrum of

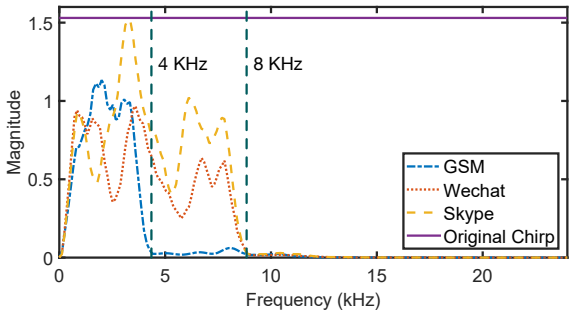


Figure 6-25: Spectrum of audio data transmitted through the VCS. A chirp signal sweeping from 10 Hz to 24 KHz played at the transmitting phone. This figure shows the spectrum of the corresponding chirp signal received through GSM, WeChat and Skype.

the audio data received at the Bob's side in Fig. 6-25. We can observe the following findings: the chirp signal is clearly truncated at 4 kHz when using the GSM network. This is because the traditional cellular network specifies to use the audio sampling rate of 8 KS/s at cell phones for saving bandwidth, i.e., all frequency components above 4 KS/s should be filtered by the VCS before the transmission. This rule is still reserved nowadays in the 4G and 5G cellular networks for the backward compatibility although their capacities actually reach a higher level. The two IM apps adopt the sampling rate of 16 KS/s for the better quality of voice. In fact, many previous studies [171] on the perceptual evaluation of speech quality (PESQ) suggest that the 16 KS/s sampling rate (i.e., $16/2 = 8$ KS/s upper limit on the voice frequency due to the Nyquist's sampling theory) can fully meet the high-quality requirements on the VCSs. On the contrary, in our system, the beacon signals are downconverted to $8 \sim 18$ KHz. Even if the beacon signals are captured by the microphone during the calls, they will be totally filtered by the VCS modules. Therefore, UPS+ does not affect the voice communications at all regardless of through GMS or through Internet based IM apps.

6.9.2 Interference to Voice Recording

UPS+ would affect the voice recoding using smartphone. Conversely, this potential influence would become a good means to silently jam spy microphones from

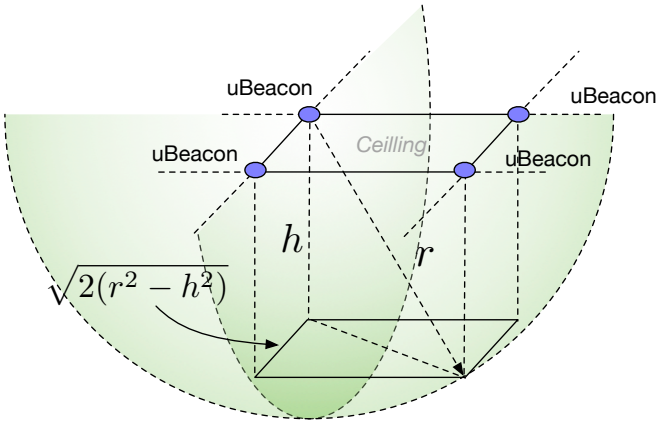


Figure 6-26: Illustration of the uBeacon deployment. All uBeacons are deployed in a grid with a space of $\sqrt{(r^2-h^2)/2}$ on the celling.

recording [59]. For example, military and government officials can secure private and confidential meetings from electronic eavesdropping; cinemas and concerts can prevent unauthorized recording of movies and live performances. The down-converted audio signals at the audible spectrum would appear only when two ultrasonic beacon signals arrive simultaneously. For the legal recording activities, we can immediately stop the interference by sending a wireless request to turn off the cBeacon.

6.9.3 Deployment Density

It is a long-standing topic about how to deploy a minimal number of beacon devices for covering the whole surveillance region, which has been studied well by a large number of previous work [172–178]. Here, we present a simplified coverage model (due to the space limit) for guiding the deployment. We assume that all uBeacon devices are deployed in a grid on the ceiling. Let r and h be the effective range of a uBeacon and the room heigh respectively. As illustrated in Fig. 6-26, the distance of two adjacent uBeacon devices in the grid should be equal to $\sqrt{(r^2-h^2)/2}$ such that any one of the four uBeacon devices can cover the whole cube and any device in the cube can receive beacon signal from the four uBeacon devices at least. In this way, the deployment density of the uBeacon devices are equal to $1/\left(\sqrt{(r^2-h^2)/2}\right)^2 = 2/(r^2-h^2)$ $\#/\mathrm{m}^2$. Suppose $h = 2$

m and $r = 3$ m, then the density should be 0.89 $\#/m^2$, namely, deploying one single-transducer made beacon device per square meter. Actually, our deployment solution provides coverage redundancy because it is sufficient for the receiver to acquire beacon signals from three uBeacons as aforementioned. This topic is out of our discussions in this work. We strongly recommend to read the previous work [172].

6.9.4 Limitation

The current prototype still has a couple of limitations. We outline them for future work.

• **Relatively shorter range.** Although we have demonstrated that three-transducer made uBeacon can achieve an effective range of 6 m, the range is still far shorter than those of RF-based solutions. Actually, the short operating range is a common drawback of sound based localization system compared with the traditional RF-based solutions. because acoustic wave attenuates faster than electromagnetic wave. However, the advantage of UPS+ is in its extremely high accuracy. Our solution is more competitive for accuracy-sensitive applications, such as AR/VR systems, localizing tiny IoT devices, indoor navigation, and so on.

• **Post-maintenance.** Since all uBeacon devices are battery-driven in UPS+, users are required to update batteries every a few months. Two methods can be utilized to mitigate this issue. First, low-duty cycle algorithm is able to significantly reduce the energy consumption as shown in the evaluation. For example, all uBeacons transit to low-powered silent states and are waken up by the smart devices when used; second, the recent advances in the wireless charging technology also provide a direction to design battery-free uBeacons in our future work.

Chapter 7

General-Purpose Deep Tracking Platform across Protocols for the Internet of Things

Today, numerous IoT protocols have been rapidly developed in recent years to meet a variety of demands such as RFID, ZigBee, LoRa, Sigfox, and NB-IoT. Fig. 7-1 shows how various IoT devices create smart warehouses[1]. Their target application scenarios are clearly diverse. NB-IoT and LoRa nodes are installed to track transporters outdoors and forklifts indoors, respectively; RFID tags are attached to goods for auto-identification and management; Sigfox tags are used to monitor raw materials and dangerous chemicals; temperature or light sensors over Zigbee are deployed to measure the warehouse environment. Such diversity in wireless protocols, however, has become a serious challenge in device management and tracking.

At present, the academia and the industry usually use the divide-and-conquer paradigm to handle diversity. Existing solutions have been tailored for specific types of devices. For example, several works [69, 96, 125–127] targeted the localization of RFID tags; one work [134] was designed for Zigbee mote positioning,

[1]A practical smart warehouse may only adopt some of IoT protocols. This toy example shows an extreme case.

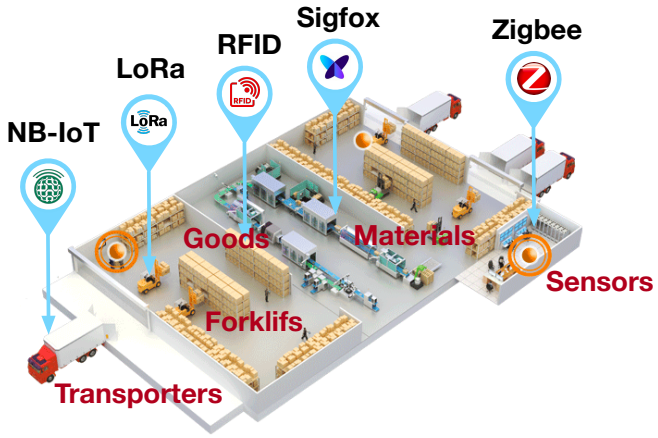


Figure 7-1: Smart Warehouse Equipped with Different Types of IoT Device.

and another [179] concentrated on LoRa nodes. An all-in-one platform for tracking all types of IoT devices does not exist. Thus, users must deploy multiple tracking systems in a warehouse or a smart home where various IoT devices are running or equip a single device with multi-radios. We predict that the coexistence of multiple protocols in IoT is more likely to occur than the outright success of any single one in the near future for two main reasons. First, no single technology can fully meet the complex application requirements. For example, RFIDs are low cost but have a limited range (a few meters). Second, replacing legacy devices completely is time consuming even though technology evolves rapidly. Therefore, a single technology is unlikely to dominate the current IoT market. Consequently, the difficulties in management and tracking due to the coexistence of multiple protocols will persist in the industry for a long time. In particular, the long-term maintenance cost for the warehouse management system will persistently introduce much overhead and cause a huge rise in the manufacturing cost, make the manufacturer lost their competitiveness.

In this chapter, we present a CTM sensing solution, named iArk, which is the first general-purpose and all-in-one tracking platform for the *non-intrusive*, *high-precise* and *flexible* localization of different types of IoT devices in the presence of severe multipath effects. This triangulation-based tracking platform is powered by a large-sized antenna array with two merits. First, this solution is non-intrusive. That is, the platform acts as a protocol-free sniffer simply report-

Table 7.1: Summary of Mainstream IoT Technologies

Type	**Frequency** (MHz)	**BW**	**Date rate**	**Modulation**
RFID	902 - 928	500 kHz	40 kbps	ASK/PSK
NB-IoT	880 - 915	180 kHz	15 kbps	BPSK/QPSK
Zigbee	868/915	1 MHz	250 kbps	BPSK/OQPSK
LoRa	868/915	500 kHz	50 kbps	CSS
Sigfox	868/915	200 Hz	600 bps	DBPSK/GFSK

*The above IoT technologies can work at multiple bands (e.g., 2.4 GHz). This table only shows their technical parameters at UHF band (i.e., $860 \sim 960$ MHz) in accord with the North American standards.

ing location results with a timestamp over the phase measurement. It does not need to communicate with IoT devices. Thus, the platform itself is essentially protocol-free without knowing the specific models. Second, the antenna array can separate the line-of-sight (LOS) direction in a complex setting full of multipath propagations to perform accurately.

Table 7.1 summarizes the technical parameters of the mainstream IoT technologies, which prefer the ultra-high frequency (UHF) ISM band as one of operating bands for improved penetrability indoors (especially in an industrial setting) and a license-free spectrum. This condition provides an opportunity to build a platform that is compatible with multiple types of IoT devices at the UHF band.

However, constructing such a tracking platform requires addressing the following challenges of engineering hardware, middleware, and learnware:

Hardware: Achieving a *high-precision* tracking platform requires considerable effort in hardware design. Specifically, building an RF frontend composed of large-sized antenna array is nontrivial. Scaling the baseband processing, transmission synchronization, reducing onboard losses and cost control raise serious system challenges. Thus, only testbeds with few antennas for indoor localization have been reported, such as 16 antennas in ArrayTrack [180] and 12 antennas in SWAN [181]. Argos [182] is a pioneer work in building large-scale antenna array but does not work for indoor localization. In this work, we build an iArk prototype with an 8×8-element antenna array plus a side antenna, by using commercial off-the-shelf radio modules, as shown in Fig. 7-2. Notice that the RF

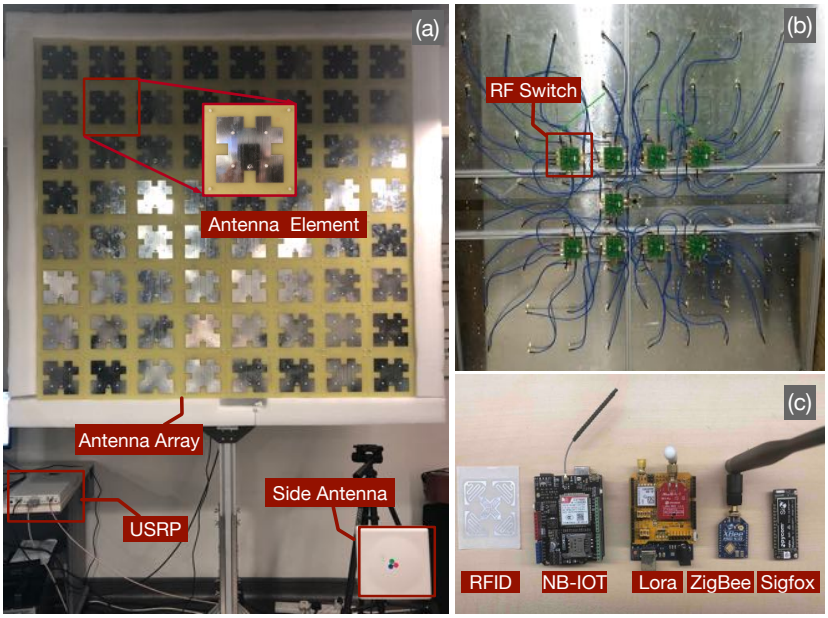


Figure 7-2: Deep Tracking Platform. (a) and (b) show the frontside and backside of the antenna array, respectively; (c) shows the tested IoT devices.

frontend is the most expensive component in the communication system. Given the cost-effectiveness of the prototype, we leverage 9 RF switches to string the 64 array elements, that share a single *main-channel* (i.e., connected to a single RF frontend) in a time-slicing manner. One of our innovations is the design of the *side-channel*, which shows considerable merit in dealing with the protocol diversity. Additional details are presented in §7.3.

Middleware: Previous indoor localization systems are usually bound to a specific type of device because the existing signal-processing algorithms for estimating physical-layer parameters (e.g., RSS or phase) are protocol-specific. For example, the computation of the angle of arrival (AoA) only considers the carrier phase rotation over the distance. However, the phase is also affected by other factors, such as preamble structure, modulation and encoding schemes, and operating channel. As a result, the cross-protocol phase estimation is a daunting undertaking. To address this challenge, we present a novel *protocol-free estimation algorithm* that leverages the signals from the side channel to counteract the negatives from the main channel. In §7.4, we explore the preceding idea further and describe the manner by which the algorithm can work with the dual signals

from the frontend.

Learnware: For a general-purpose platform across protocols, past tracking approaches can be easily migrated to iArk with minimal efforts. To demonstrate this capability, we develop an AI-boosted localization system. Specifically, we logically divide the 8×8 array into 25 overlapping 4×4 subarrays as inspired by the binocular stereo in computer vision; each subarray can estimate a direct AoA of the device. The targeted device will be exactly located at the intersection of the lines along the estimated angles. The system works well in a relatively stationary environment in the absence of multipath. However, a warehouse is full of indoor reflectors (e.g., ceiling, walls, and furniture) in reality and these reflectors produce many multipath propagations. Consequently, the estimated AoA is severely skewed. To address this challenge, we use two deep neural networks, the *AoA neural network* (`ANN`) and the *Triangulation neural network* (`TNN`), which are used to compute the AoA and perform the triangulation respectively in §7.5.

Summary of Results. We evaluate five types of off-the-shelf IoT devices. The results demonstrate that iArk can fully identify the direct AoA with median errors of $(0.46°, 1.5°)$ at azimuthal and elevation angles. It performs localization with median errors of 8.7, 5.8 and 14.2 cm in X, Y and Z dimensions over a medium-sized office (10×20 m^2). The accuracy of iArk matches or even exceeds that of past AoA systems.

7.1 Antenna Array Primer

Before introducing our system, a brief review is given in terms of the antenna array and spatial spectrum.

Antenna Array. The direction of the RF source can be uniquely represented with two angles, azimuthal angle and elevation angle. The Angle of Arrival (AoA) of the RF source is computed by comparing the phases of the received signals at multiple antennas. Suppose an antenna array with $\sqrt{K} \times \sqrt{K}$ antennas (aka elements). The spacing of two adjacent antennas is L, which is less than the

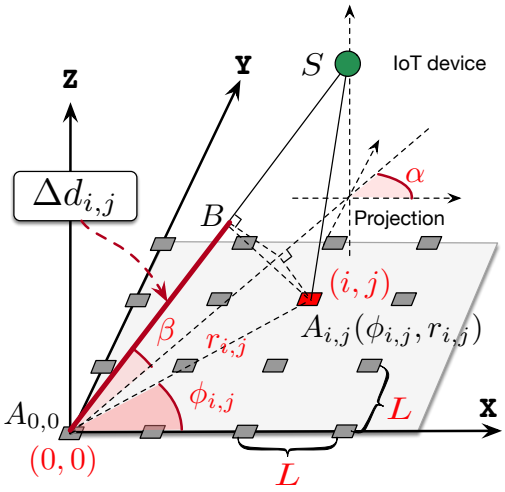


Figure 7-3: AoA Computation with an Antenna Array. K antennas (aka elements) are deployed in a grid. The distance difference $\Delta d_{i,j} = |SA_{i,j}| - |SA_{0,0}| = -r_{i,j}\cos(\alpha - \phi_{i,j})\cos(\beta)$.

wavelength λ, i.e., $L < \lambda$. Fig. 7-3 shows an example of 4×4 antenna array. For the convenience of calculation, the positions of antennas are transformed to the polar coordinate system where the antenna $A_{0,0}$ on the bottom-left corner is selected as the origin. The polar coordinate of the antenna $A_{i,j}$ is given by:

$$\begin{cases} r_{i,j} = L\sqrt{i^2 + j^2} \\ \phi_{i,j} = \arctan 2(j, i) \end{cases} \tag{7.1}$$

where $i, j = 0, \cdots, \sqrt{K} - 1$. Let $d_{i,j}$ be the distance from the RF source S to the antenna $A_{i,j}$. When S is relatively far from the array, $|SA_{i,j}| \approx |SB|$ where $A_{i,j}B \perp SA_{0,0}$. The distance difference between the source S and the antennas of $A_{i,j}$ and $A_{0,0}$, denoted by $\Delta d_{i,j}$, is given by:

$$\Delta d_{i,j} = d_{i,j} - d_{0,0} = -r_{i,j}\cos(\alpha - \phi_{i,j})\cos(\beta) \tag{7.2}$$

where α and β denote the azimuthal and elevation angles, respectively, and $\alpha \in [0, 360°)$ and $\beta \in [0, 90°]$. Let $\theta_{i,j}$ denote the phase of the RF signal measured at the antenna $A_{i,j}$, $\theta_{i,j} \in [0, 2\pi)$. The phase and the distance have the following

known relation at antenna $A_{i,j}$ due to the phase rotation:

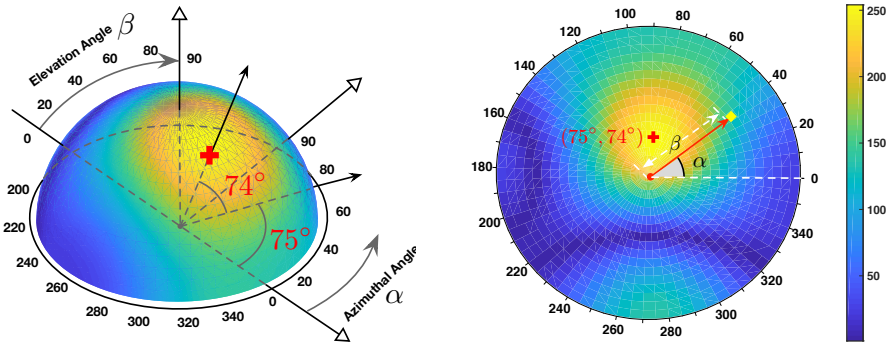
$$\theta_{i,j} = 2\pi d_{i,j}/\lambda \bmod 2\pi \tag{7.3}$$

Hence, the phase difference between the received signals at the two antennas, denoted by $\Delta\theta_{i,j}$, relates to the difference in their distances from the source as follows:

$$\begin{aligned} \Delta\theta_{i,j} &= \theta_{i,j} - \theta_{0,0} = 2\pi\Delta d_{i,j}/\lambda \bmod 2\pi \\ &= -2\pi r_{i,j}\cos(\alpha - \phi_{i,j})\cos(\beta)/\lambda \bmod 2\pi \end{aligned} \tag{7.4}$$

Spatial Spectrum. Copies of a signal traveling along different paths may overlap at the receiving antenna. To separate the copies of the device's signal arriving from different directions, we must measure the power coming from each direction. A large phased antenna array can achieve this goal by forming a very narrow beam and steering it around. When steering its beam to the line-of-sight (LOS) direction, the array essentially filters out the power coming from all other directions. Equivalently, we can compute a *spatial spectrum* (SS) [79,96,126] that indicates all possible AoAs. Formally, SS is defined as $\mathbf{P}(\alpha, \beta)$, where the power of the signal received in the beam from azimuthal angle α and elevation angle β. Let $\theta_{i,j}$ and $\widehat{\theta}_{i,j}$ denote the theoretical and measured phases of the RF signal received at the antenna $A_{i,j}$. We can derive the relative power of the source along the angle of (α, β) as follows:

$$\begin{aligned} \mathbf{P}(\alpha,\beta) &= \left| \frac{1}{K} \sum_{i,j}^{\sqrt{K},\sqrt{K}} e^{\mathbf{J}\left(\Delta\widehat{\theta}_{i,j} - \Delta\theta_{i,j}\right)} \right|^2 \\ &= \left| \frac{1}{K} \sum_{i,j}^{\sqrt{K},\sqrt{K}} e^{\mathbf{J}\left((\widehat{\theta}_{i,j} - \widehat{\theta}_{0,0}) - \frac{-2\pi r_{i,j}}{\lambda}\cos(\alpha-\phi_{i,j})\cos(\beta)\right)} \right|^2 \\ &= \left| \frac{1}{K} \sum_{i,j}^{\sqrt{K},\sqrt{K}} e^{\mathbf{J}\left(\widehat{\theta}_{i,j} + \frac{2\pi r_{i,j}}{\lambda}\cos(\alpha-\phi_{i,j})\cos(\beta)\right)} \right|^2 \end{aligned} \tag{7.5}$$



(a) 3D Spatial Spectrum **(b) 2D Spatial Spectrum**

Figure 7-4: Illustration of a Spatial Spectrum. The spatial spectrum is generated by a 2×2 antenna array. (a) and (b) show the same spatial spectrums but in 3D and 2D forms respectively. The 2D spatial spectrum is the projection of the 3D spectrum.

where $\Delta\widehat{\theta}_{i,j} = \widehat{\theta}_{i,j} - \widehat{\theta}_{0,0}$ and $\Delta\theta_{i,j}$ is defined in Eqn 7.4. $\widehat{\theta}_{0,0}$ is a constant term that can be extracted to outside of the sum, and $|e^{-\mathbf{J}\widehat{\theta}_{0,0}}| = 1$. The aforementioned equation actually correlates the measured phase difference with a "template" across the K antennas given an AoA of (α, β). The SS would spike at the LOS direction of the RF source. Thus, the direction leading to the maximum power in the SS is the direct AoA (i.e., LOS direction). To visually understand the SS, we show an example in Fig. 7-4. The SS is generated by a 2×2 antenna array. The maximum relative power of SS is achieved at the angles of $(75°, 74°)$, which is considered as the direct AoA of the RF source. The polar 2D SS is a projection of the 3D SS, where the polar angle and distance in the 2D SS indicate the azimuthal and elevation angles respectively. For the sake of clarity, we will use 2D SS in the subsequent sections by default.

Triangulation. After two direct AoAs at two different positions are obtained, an RF source can be located at the intersection of the two directions. This approach is called *triangulation*.

7.2 System Architecture

We build a novel platform called iArk, which consists of a large-sized two-dimensional antenna array, to measure the AoA for the triangulation. The platform is for all

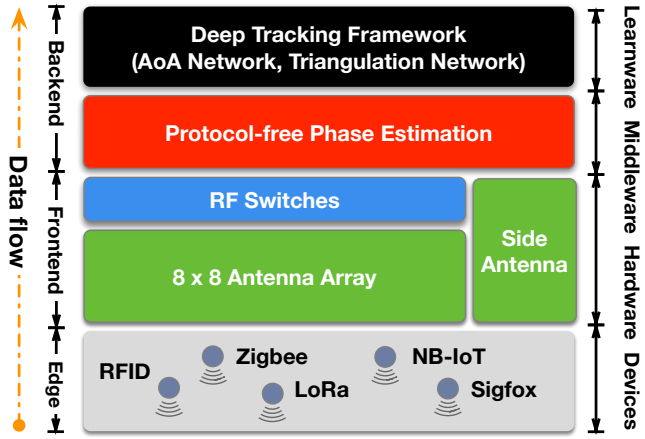


Figure 7-5: System Architecture

types of IoT devices operating at the UHF ISM band (i.e., approximately 915 MHz). It can be also easily extended to other frequency bands (e.g., 2.4 GHz) by simply replacing the RF frontend. iArk adopts a layered and loose-coupled architecture, as shown in Fig. 7-5. From a high-level perspective, the system contains three layers, namely, the hardware, the middleware, and the learnware.

- **Hardware**: We design an "$8 \times 8 + 1$" frontend to acquire RF signals from the IoT devices. The "8×8" - antenna array can significantly increase the spatial diversity in the AoA estimation (§7.3).
- **Middleware**: We design a middleware layer to hide the tracking algorithm from the heterogeneity of wireless protocols. In particular, the protocol-free phase estimation (PPE) is constructed for this purpose (§7.4).
- **Learnware**: We develop AI-boosted localization framework called *learnware* over the middleware for case study. The learnware is composed of two deep neural networks for AoA estimation and triangulation respectively (§7.5).

The platform acts as a sniffer, so it does not need to know what kinds of devices in the warehouse and not to communicate with them. The platform also does not need to know which device transmits the signal or whether a signal collision occurs. This task is left to the upper-layer applications. The platform

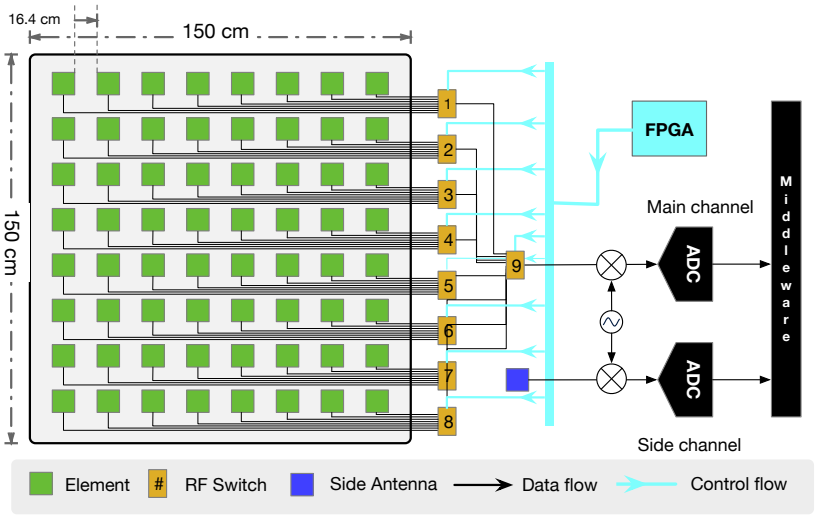


Figure 7-6: Schematic of the RF Frontend. The frontend comprises an 8×8 antenna array and a side antenna that form the main and side channels, respectively.

simply stores all location results over the phase matrices in the database with a timestamp. The transmitting device or any third-party (e.g., gateway) that is interested in the localization results should know when the device successfully transmits a packet. They can retrieve the locations through the timestamps. Thus, the platform can support any number of devices as long as they could successfully communicate with each other. Our design is essentially scalable for any kind of IoT network with a large number of devices. Such a loose-coupling architecture design avoids packet recognition, which benefits fast deployment and provides a measure for protecting privacy.

7.3 Hardware: Acquiring Signals in A non-intrusive Manner

At the heart of the hardware is the RF frontend, which is used to acquire RF signals from the air. Fig. 7-6 shows the schematic of our frontend. From a high level perspective, the frontend consists of an antenna array and a side channel. The frontend continues to listen at the 915 MHz wireless carrier with a 32.8 cm wavelength.

7.3.1 Switched Antenna Array

One of the key components of the frontend is an 8×8 antenna array. A total of 64 elements are deployed in the form of a grid with a spacing of 16.4 cm. Each array element consists of a directional square pitch with a side length of 12 cm (i.e., half wavelength). The array is 150×150 cm^2 in area. The ADC, which is an expensive component in the RF communication system, is used to convert an analog signal into a digital signal. The high-end antenna array used in ArrayTrack [180] and Argos [182] equips each element with a stand-alone ADC, which increases the entire cost of the antenna array by orders of magnitude. A few bits of the incoming packet are sufficient for each element to estimate the signal phase. Acquiring the entire packet is unnecessary for each element. To save cost, we allow all elements to share a single ADC by using 9 single-pole-eight-throws (SP8T) RF switches. Specifically, these switches constitute a two-stage switching system. The first stage is composed of eight switches (switches #1 $\sim$ #8), each of which is connected to eight array elements in a row. In the second stage, the ninth switch (switch #9) further bridges the previous eight RF switches to an ADC. This design can fully utilize the 64 ports of the switches. These switches are manipulated by an FPGA to connect one of the 64 elements to the ADC instantly (with a delay of approximately $3\mu s$). At any moment, only a single antenna element can be put through. In this manner, each element can receive a segment of the entire packet. The switching delay is negligible compared with the ms-level packet duration.

The 64 elements are scheduled to connect to the ADC in sequence. The scheduling cycle starts from the first element on the top-left corner and then goes through the others from left to right, top to bottom, and finally ends at the 64^{th} element. The scheduling cycle is repeated. In each cycle, every element is exclusively connected to the ADC over $30\mu s$. In total, each scheduling cycle takes $(30 + 3) \times 64 = 2112$ μs. Regarding the 6 MS/s sampling, we can obtain $30\mu s \times 6$MS/s $= 180$ samples from each element during a scheduled cycle. The

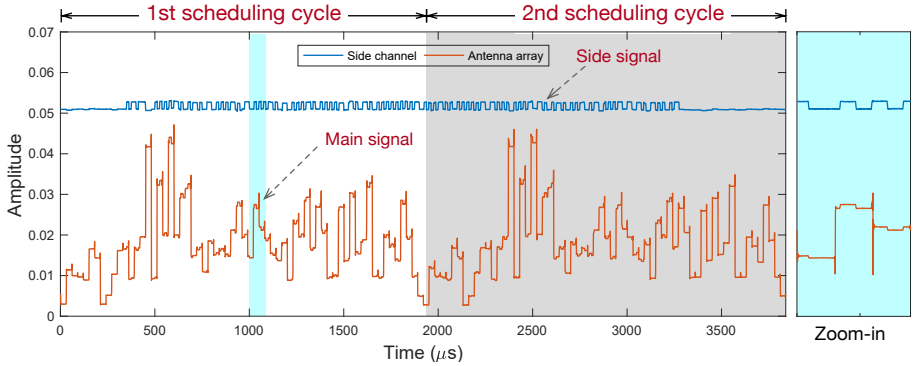


Figure 7-7: RF Signals Acquired via the Main and Side Channels. The signal is transmitted from an RFID tag. The red curve shows the main signal, which merges the segments acquired by the elements. Side channel continuously acquires the RF signal as shown in the blue curve.

64 array elements and the ADC form the *main channel*.

7.3.2 Side-Channel

Unlike prior works [180,181,183–187], our unique design is the *side channel*, which is a second channel parallel to the main channel. The side channel is composed of an additional side antenna and a standalone ADC. It acquires RF signals persistently and continuously. Fig. 7-7 compares the signals acquired from the main and side channels. The signal contains many discontinuous "stairs", each of which represents the signal acquired via one array element. The elements on the opposite corners are up to 163 cm apart and span over nearly 5 propagation cycles, which makes the signal segments differ in amplitude and phase. On the contrary, the signal through the side channel is considerably smoother and more continuous because it is acquired by the same antenna uninterruptedly. The merits of the side channel are discussed in §7.4.

7.4 Middleware: Generating Protocol Free Spatial Spectrum

The fundamental of AoA computation is the phase estimation of the RF signal as mentioned in §7.1. However, estimation is non-trivial, especially when handling

numerous IoT devices that communicate with different protocols. In this section, we first investigate the current estimation algorithms to motivate our design and then describe our estimation algorithm.

7.4.1 Background of Phase Estimation

A wireless signal is typically represented by a stream of discrete complex numbers [188]. The digital bits cannot be directly propagated into the air. The transmitter should initially convert the digital bits into complex symbols. This step is called shift-keying. For example, Zigbee and NB-IoT map a "0" bit to $e^{\mathbf{J}\pi}$ and "1" bit to $e^{-\mathbf{J}\pi}$ (i.e., BPSK), and LoRa maps two bits into different chirp signals (i.e., chirp spectrum spread (CSS)). For simplicity, we use $X[t]$ to denote the transmitted symbols after shift-keying. Then, the transmitter moves these symbols onto the carrier at frequency f_c by multiplying the symbols and the carrier signal, in a process called modulation. The transmitting carrier can also be represented by complex numbers, that is, $C_T[t] = e^{\mathbf{J}(2\pi f_c t+\Phi)}$, where Φ is a constant (i.e., the initial phase shift caused by the hardware). Finally, the transmitted signal (i.e., $S_T[t]$) propagated into the air can be modeled as:

$$S_T[t] = X[t]C_T[t] = X[t]e^{\mathbf{J}(2\pi f_c t+\Phi)} \tag{7.6}$$

At the receiver side, the received signal differs from the transmitted signal due to propagation. Specifically, the received signal (denoted by $S_R[t]$) turns into:

$$S_R[t] = HS_T[t] = aX[t]e^{\mathbf{J}(2\pi f_c t+\theta+\Phi)} \tag{7.7}$$

where $H = ae^{\mathbf{J}\theta}$ is the channel parameter. Amplitude a refers to the channel attenuation and its angle θ is a phase shift that depends on the distance between the transmitter and the receiver. To decode the transmitted symbols, the received signal is multiplied by the conjugate of a similar carrier signal (i.e., $C_R[t] = e^{\mathbf{J}(2\pi f_c' t+\Delta\phi)}$) in a process called demodulation. The received symbol (i.e., $Y[t]$) is

given by:

$$Y[t] = S_R[t]C_R^*[t] = aX[t]e^{\mathbf{J}(2\pi\Delta f \cdot t + \theta + \Phi - \Delta\phi)} \tag{7.8}$$

where $\Delta f = (f_c - f_c')$ is called the carrier frequency offset (CFO) and $\Delta\phi$ is the carrier phase offset (CPO) introduced by the up- and down-conversions. Essentially, CFO and CPO are results of the out-of-sync in frequency and clock between the transmitter and the receiver. The received symbol $Y[t]$ differs from the transmitted symbol $X[t]$ in amplitude and phase. If the transmitter and receiver are well-synchronized in frequency and clock, then $\Delta f = 0$ and $\Delta\phi = 0$. The received symbol can be further simplified as follow:

$$Y[t] = aX[t]e^{\mathbf{J}(\theta + \Phi)} \tag{7.9}$$

Let $\widetilde{\theta} = \angle\frac{Y[t]}{X[t]}$ where $\angle(\cdot)$ takes the angle of the complex number. Then,

$$\widetilde{\theta} = \angle\frac{Y[t]}{X[t]} = \theta + \Phi \tag{7.10}$$

θ is the *true phase* that rotates over the distance. We call $\widetilde{\theta}$ *pseudo-phase*, which is the sum of the true phase and the constant phase shift, for distinction. From the equation, we can estimate the pseudo-phase of the RF signal by comparing the transmitted with the received symbols. We call this approach *direct phase estimation* (DPE), which is widely adopted in existing systems.

7.4.2 Motivations and Challenges

Unfortunately, DPE fails to work for a large number of protocol-diverse IoT devices for four reasons:

• **Preamble dependence**: Eqn. 7.10 suggests that the pseudo-phase is the angle of the ratio of received symbols (i.e., $Y[t]$) to the transmitted symbols (i.e., $X[t]$). Thus, the receiver must know what symbols are transmitted from the device. To address this issue, the receivers use the preambles that are fixed and known to each other for the DPE. However, preambles are diverse in different

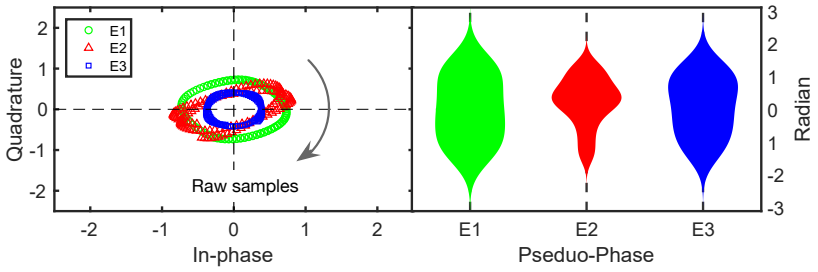
protocols. To meet this condition, the platform must attempt to decode each packet using all kinds of protocols one by one, which is a time-consuming and cumbersome procedure.

• **Modulation dependence**: Some modulation techniques, such as phase-shift keying or chirp spread spectrum, represent data by changing the phase of the signal, which leads to extra uncontrollable variables in the pseudo-phase. For example, the LoRa protocol adopts CSS as the shift-keying scheme, which increases the frequency or equivalent phase linearly. Consequently, the pseudo-phase may span over the entire domain and result in an uncertain AoA.

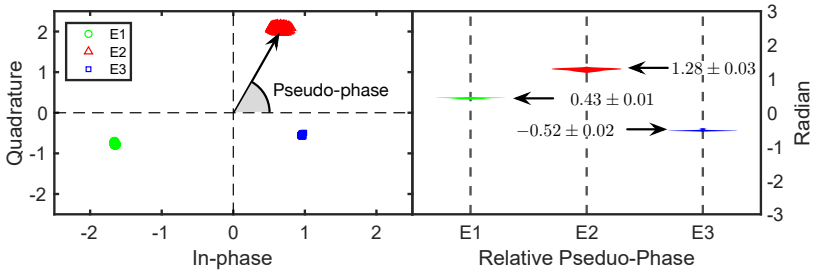
• **Channel dependence**: We assume that CFO and CPO do not exist in Eqn. 7.10. The CFO not only exists in iArk but also behaves worse than the existing IoT systems because our frontend is fixed to listen at 915 MHz, whereas IoT devices may choose a sub-channel that deviates from the center frequency to reduce signal collisions. Moreover, the RFID Gen 2 protocol requires RFID systems to hop randomly among the 52 sub-channels every 500 ms [97]. Thus, the practical pseudo-phase $\widetilde{\theta} = 2\pi\Delta f \cdot t + \theta + \Phi - \Delta\phi$, is derived from Eqn. 7.8. Thus, the pseudo-phase increases as a function of time, that is, $\widetilde{\theta} \sim \Delta f \cdot t$.

• **Device dependence**: True phase θ is hidden inside the pseudo-phase $\widetilde{\theta}$ (Eqn. 7.10). The true phase is the only phase variable related to distance and the key for the AoA estimation. If the pseudo-phase is directly used, then the estimated AoA will contain a large error. To extract θ from $\widetilde{\theta}$, transitional solutions usually require a pre-process step similar to that in previous work [189] for estimating Φ in the calibration process. However, this approach is unscalable. We consider a large-sized warehouse where thousands of IoT devices are deployed; thus, measuring Φ for each device is a nearly impossible mission.

The four factors above are protocol-specific. Any ambiguity in one of them may lead to false phase estimation. Thus, past localization solutions must be specifically bound to a device type with only one protocol. To better understand this issue, we show a LoRa signal captured by three array elements, in Fig. 7-8a. Ideally, these samples should distribute within a small cluster because the device



(a) Raw samples



(b) Relative samples

Figure 7-8: An RF Signal Transmitted from a Static LoRa Device. The signal is captured by three array elements (i.e., E1 $\sim$ E3), each of which acquires 180 samples. (a) On the left, the raw samples are shown in the constellation. These samples rotate as a function of time because the LoRa protocol adopts CSS as its shift-keying scheme, resulting in three loops. On the right, we show the distribution of the samples' pseudo-phases (i.e., the angles of samples). (b) Similarly, we show the relative samples in the constellations and their distribution in pseudo-phase on the left and right respectively. Compared with raw samples, the relative samples collapse into three extremely small clusters in the constellation and remain consistent in the distribution of pseudo-phase.

remains static and the corresponding true phase and pseudo-phase remain constant (see Eqn. 7.10). In fact, the samples rotate as a function of time, resulting in three loops, due to the presence of CFO and CSS based shift-keying. This physical-layer constraint severely weakens the generalization of past solutions.

7.4.3 Protocol-free Estimation Algorithm

To address the issue, we develop a novel protocol-free phase estimation algorithm that can address all the dependencies mentioned above [8,190,191]. Our intuition is that the main and side channels are driven by the same oscillation such that their signals are well-synchronized and contain identical symbols, the CFO, the CPO, and hardware diversity. Thus, by comparing their signals, we can remove

all negative effects. Let $Y_0[t]$ and $Y_{i,j}[t]$ be the received symbol from the side channel and the array element $A_{i,j}$ respectively. $i, j = 0, 1, 2, \ldots, 63$. We define the relative sample (i.e., $Z_{i,j}[t]$) at time t as follows:

$$\begin{aligned} Z_{i,j}[t] = \frac{Y_{i,j}[t]}{Y_0[t]} &= \frac{a_{i,j}\cancel{X[t]}e^{\mathbf{J}(\cancel{2\pi\Delta f t}+\theta_{i,j}+\cancel{\Phi}-\cancel{\Delta\phi})}}{a_0\cancel{X[t]}e^{\mathbf{J}(\cancel{2\pi\Delta f t}+\theta_0+\cancel{\Phi}-\cancel{\Delta\phi})}} \\ &= (\frac{a_{i,j}}{a_0})e^{\mathbf{J}(\theta_{i,j}-\theta_0)} \end{aligned} \tag{7.11}$$

where $\theta_{i,j}$ and θ_0 are the true phases of the RF signals acquired by the array element $A_{i,j}$ and the side channel, respectively. $Y_k[t]$ and $Y_0[t]$ are defined in Eqn. 7.8. Correspondingly, the *relative pseudo-phase* of the RF signal received at $A_{i,j}$, which is denoted by $\widehat{\theta}_k$, is computed as follows:

$$\widehat{\theta}_{i,j} = \angle Z_{i,j}[t] = \theta_{i,j} - \theta_0 \tag{7.12}$$

Compared with the traditional pseudo-phase (i.e., $\tilde{\theta}$ in Eqn. 7.10), the relative pseudo-phase holds the true phase but provides the following key points:

- **Preamble-free**: The transmitted symbol (i.e., $X[t]$) is canceled, which implies that the algorithm does not need to know what symbols are transmitted from the device. The algorithm can estimate the phase using any received symbols more than preambles.
- **Modulation-free**: Similarly, the means of encoding the data on the baseband and modulating onto the carrier becomes irrelevant to the estimation results because $X[t]$ is cancelled.
- **Channel-free**: Δf and $\Delta\phi$ are canceled. The platform does not need to know at which frequency the RF source operats and how the data is up- or down-converted. The relative pseudo-phase is totally free of the out-of-sync frequency or phase.
- **Device-free**: The initial phase shift from the device (i.e., Φ) is canceled. Thus, transmitter calibration (i.e., on IoT devices) is not required.

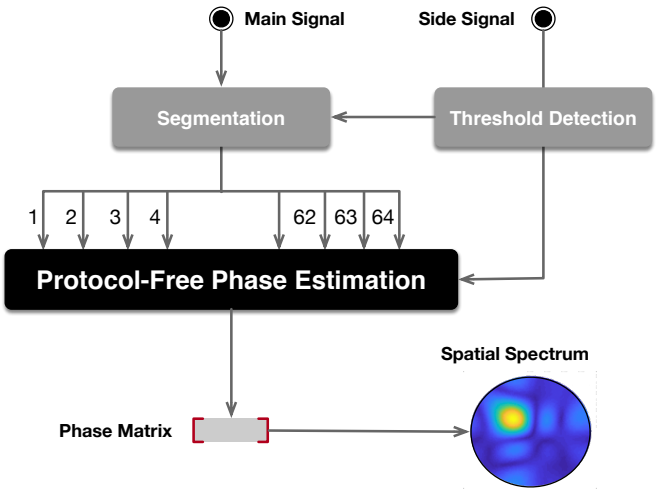


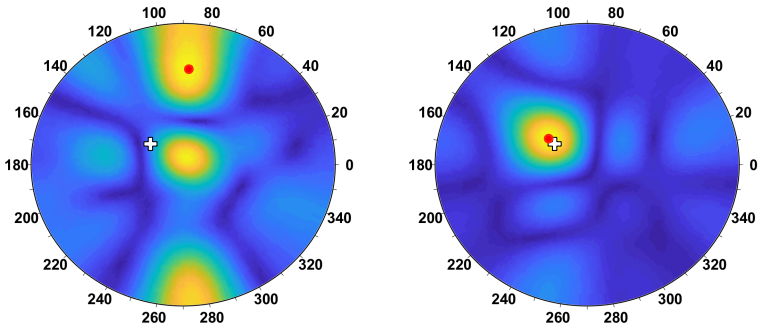
Figure 7-9: Middleware Workflow

In summary, the algorithm spontaneously eliminates the heterogeneity of protocols regardless of which type of the preamble is used, which bits are received, and which channel the transmitters operate at. Thus, the algorithm is applicable to any kind of protocol, that is, protocol-free. Importantly, the algorithm can scale up to any number of IoT devices because pre-calibration is unnecessary. To visually understand the protocol-free feature, we show the relative samples of the same LoRa signal in Fig. 7-8b for comparison. As opposed to raw samples, a total of 180 relative samples collected by each element collapse into an extremely small cluster with a stable and the equal angle (i.e., pseudo-phase) instead of a loop. This condition demonstrates that the relative pseudo-phase maintains stability over all the samples without negative affects.

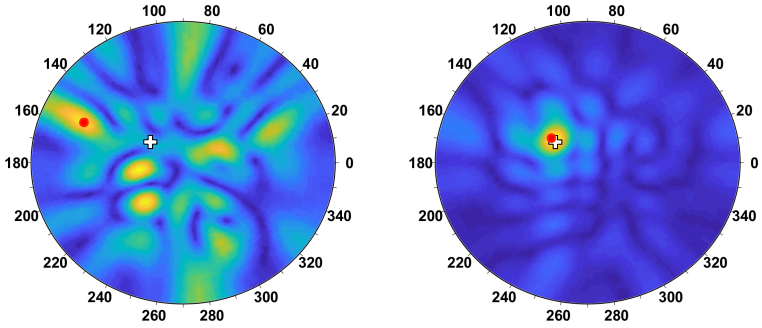
7.4.4 Put Things Together

The discussion so far focuses on the protocol-free phase estimation (PPE) by virtual of the side channel. Fig. 7-9 sketches the entire workflow. When gathering 180×64 sample from the hardware (i.e., each of the 64 elements collects 180 samples.), the middleware repeats the following steps.

- **Step 1**: The middleware decides the presence of a packet by comparing the average amplitude of the samples from the side-channel with a user-defined threshold.



(a) Raw 4×4 **array** **(b) Relative** 4×4 **array**



(c) Raw 8×8 **array** **(d) Relative** 8×8 **array**

Figure 7-10: Comparisons of SS Generated with DPE and PPE. The symbol of white cross (+) denotes the ground truth, while the symbol of red circle (•) denotes the peak of the spectrum. The spectrums are computed by using raw pseudo-phase and relative pseudo-phase, which are estimated by DEP and PPE respectively. (a) and (b) are generated by a 4×4 antenna array; (c) and (d) are generated by an 8×8 antenna array.

• **Step 2**: If the average value is higher than the threshold, then the main and side signals are divided into 64 segments, each of which contains 180 samples; otherwise, the algorithm skips the following steps and goes back to Step 1.

• **Step 3**: The average relative pseudo-phase across the 180 samples is computed using Eqn. 7.12 for each array element.

• **Step 4**: The direct AoA of the RF through the spatial spectrum is computed by substituting the relative pseudo-phase into Eqn. 7.5.

We compare the effectiveness of DPE and PPE in Fig. 7-10. In the figure, the spatial spectrums are generated with the raw and relative pseudo-phase, estimated by DPE and PPE respectively, as a function of the array size. The two spatial spectrums above are generated by a 4×4 antenna array, while the two

below are generated by an 8×8 antenna array. When using the raw pseudo-phase, the direct AoA errors are $(58.7°, 37.2°)$ and $(9.3°, 40.2°)$ for the 4×4 and 8×8 arrays. In contrast, the AoA errors are reduced to $(2.7°, 2.2°)$ and $(1.7°, 0.8°)$ when using the relative pseudo-phase. These results fully demonstrate that the PPE can truly improve the accuracy of AoA estimation significantly.

Discussion. Finally, two points are worth-noting:

• The five IoT protocol has independent preambles. The platform uses the five preamble templates to correlate the incoming signal in turn for the packet detection. If one correlation results exceed a threshold, it confirms the presence of an IoT packet. This is also a very common way in wireless communication to determine the arrival of a packet. The threshold is a user-specific parameter.

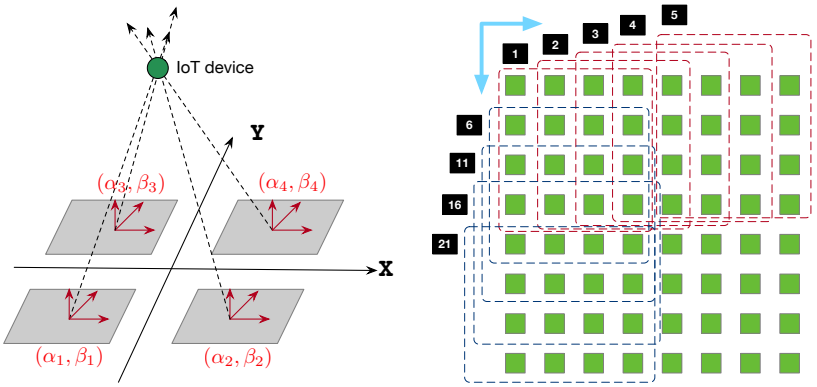
• The shortest packet is transmitted from the RFID tags and has an interval of 3300 μs among all the IoT protocols, which is $1.5\times$ longer than the time consumed on the scheduling cycle (i.e., $2112\mu s$). Thus, our algorithm can ensure that the device is located at least once during transmission of any kind of packet.

7.5 Learnware: a framework for deep tracking

This section demonstrates an AI-boosted deep-tracking framework.

7.5.1 Convolutional Multi-view Stereo

A high-precision AoA can be estimated through the SS across 64 array elements. However, a single AoA result is inadequate to compute the position of an IoT device. In practice, we must obtain at least two AoA results from two different antenna arrays. This process is similar to the binocular stereo (or the multi-view stereo) in computer vision, where two (or multiple) cameras displaced horizontally from one another are used to obtain two differing views on a scene. Triangulation is the underlying mechanism behind the multi-view stereo, that is, the transmitter is located at the intersection of two (or multiple) lines along the AoAs. Similarly, we can *logically* divide the entire antenna array into several sub arrays, each of which can calculate an AoA of the IoT device. The device is finally located at



(a) Four-view Stereo **(b) Convolutional Stereo**

Figure 7-11: Multi-view Stereo. (a) the antenna array is divided into four 4×4 subarrays; (b) the array is divided into 25 overlapping subarrays.

the intersection of the computed AoAs. Fig. 7-11a shows an example wherein the entire array is logically divided into four 4×4 subarrays.

When the incoming signals are phase-synchronized with one another due to multipath, the system perceives distinct incoming signals as one superimposed signal, which results in distortion in the SS and false peaks in $P(\alpha, \beta)$ [180, 192]. To address this issue, we use the *convolutional multi-view stereo* to handle the distortion caused by the multipath. The convolutional multi-view stereo averages the results of multiple overlapping sub arrays. Fig. 7-11b shows a toy example where a 4×4 window is moved from right to left and from top to bottom with one step. We compute the direction by using the 16 antenna elements in each window. As a result, we can obtain 25 AoAs, each of which is assigned a weight. The position of the device is determined by the weighted average. All weights are learned using a neural network, which is described subsequently.

7.5.2 Deep Tracking Networks

AoA-based localization is widely used in radar and acoustics. However, realizing this task indoors is challenging due to the presence of strong multipath RF propagations. We notice that the final received signal is a weighted superimposition of the signals traveling from multiple paths. The AoA correlation should achieve a peak in the line-of-sight direction. In contrast, the essence of a neural network is

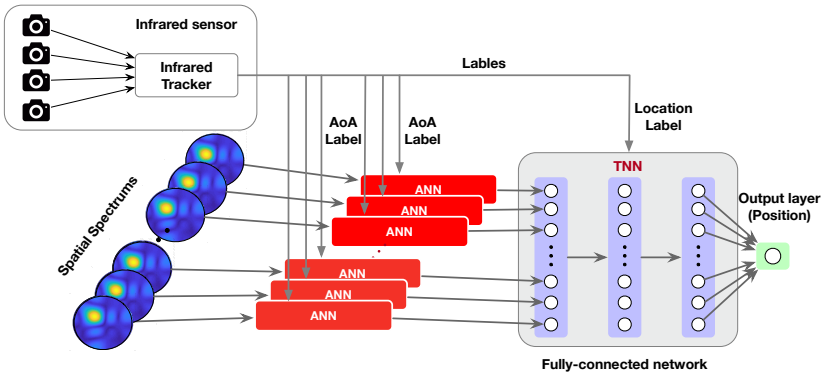


Figure 7-12: Deep Neural Networks. The overall model can be divided into two main neural networks: AoA neural network (`ANN`) and triangulation neural network (`TNN`). `ANN` is a typical CNN-based deep residual network, whereas `TNN` is a fully connected network with three layers. The top graph shows the generation of the true AoAs and positions of the transmitter by using infrared motion sensors. They are used to train the two networks.

to decompose the input signal (e.g. an image) through the approach of weighted averaging where each neuron is assigned a weight. Inspired by this insight, we would like to use the deep neural network to identify the LOS propagation from the final combined signal. Fig. 7-12 illustrates the structure of the framework. Our framework contains two neural networks, namely, *AoA neural network* (`ANN`) and *Triangulation neural network* (`TNN`).

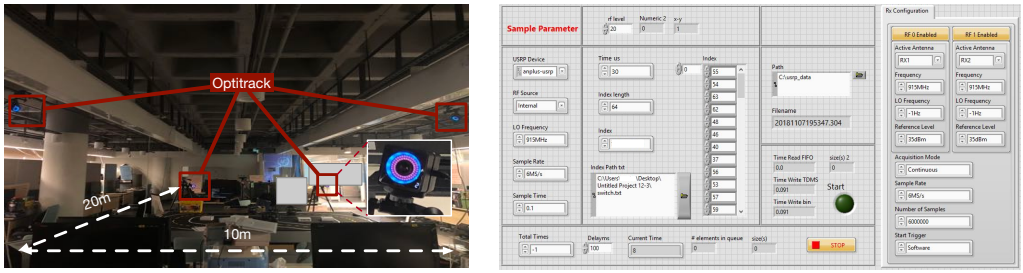
`ANN`: We leverage the recent success of convolutional neural networks (CNNs), which have demonstrated a major development in object tracking for images and videos [193]. The value of neurons in a CNN can be computed as the convolution of a weighted kernel with the neurons in the previous layer. Our intuition here is that the linear superposition of multipath propagations is extremely similar to the weighted overlay of neurons[2], whereas the weight kernel is desired to depict such combinations in the SS. The CNN takes the SS as input and outputs the AoA. A few standard CNN designs are widely used across tasks. We choose to use the 18-layer ResNet [194], which uses residual connections across different layers. The final layer is fully connected for regression. The `ANN` is a general neural network used to identify the AoA from an SS regardless of the spectrum

[2]The in-phase and quadrature components of the complex signal can be viewed as the values from two color channels.

generated by using a logical or a physical antenna array. In iArk, we have 25 logical subarrays. Thus, 25 ANNs are constructed and trained separately. In the example shown in Fig. 7-10, the AoA errors are further decreased by ANN to $(2.05°, 0.8°)$ and $(1.3°, 0.4°)$ when using the 4×4 and 8×8 antenna arrays respectively. The ANN provides over 50% extra improvements compared with the naive peak based approach.

TNN: We construct another fully-connected neural network including three layers to perform the triangulation. The TNN takes the 25 AoAs estimated by the previous ANNs as the input. Essentially, the network determines the weighted averages of the input AoA values, whereas the weights are automatically learned by training. Two neural networks are not combined into one because a single neural network may cause learning to fall into signature-based recognition. Consequently, the position cannot be correctly predicted if the device is located at a new or distant position that is not sampled in the training phase. Instead, we build the connection between the RF signal and the AoA rather than the position in the ANN. We only need a relatively small number of training data around the antenna array; however, the proposed framework can predict the distant transmitter. We aim to let the network learn how to identify the LOS AoA in the presence of severe multipath effects through the ANN.

Training data. We need to collect two kinds of labels (i.e., AoA and position) for the ANN and the TNN respectively. Thus, we gather the true positions of transmitters by using OptiTrack [195]. The OptiTrack system can track the center of any object with an accuracy of 20μm using infrared (IR) cameras. We attach an infrared marker on the antenna of each target device and measure its location via OptiTrack 120 times every second. The AoA relative to the subarray can be computed by the collected position.



(a) Scenario **(b) Debug tool**

Figure 7-13: Experimental Setup. (a) Evaluation scenario and OptiTrack system and (b) Screenshot of our debug tool.

7.6 Implementation

After a two-year effort, we successfully built an entire prototype of iArk with the hardware, middleware, and learnware.

Hardware Settings: Fig. 7-2 shows the RF frontend. We use the microstrip technique to fabricate the antenna array on a printed circuit board (PCB) for avoiding signal attenuation on the board. The PCB is composed of the substrate of RT/duroid 5880. Each array element has 0.5 dBi gain, 0.5 dB flatness, and 52° beamwidth. The model of high-speed RF switches is BGS18GA14 [196] from Infineon Technologies. We use a USRP 2950 software defined radio (SDR) from NI [197] to build the frontend. The SDR contains two stand-alone I/O interfaces, which are used to build the main and side channels. The RF frontend has a bandwidth of 140 MHz, which covers the entire UHF of 860 $\sim$ 960 MHz. We also develop a debugging tool to facilitate hardware testing (Fig. 7-13b). The backend runs at a high-performance PC equipped with an Intel CPU Xeon E5-2620 and an NVIDIA GTX 1080Ti GPU.

Middleware Settings: We test five types of IoT devices, namely, ImpinJ Monza QT4 RFID tags [198], iGi XBee Pro 900 HP (Zigbee) [199], Dragino LoRa Shields (LoRa) [200], DFRobot SIMC7000C shields (NB-IoT) [201] and PyCom Sipy chips (Sigfox) [202]. They are shown in Fig. 7-2. We use a Zigbee device for the following evaluation by default unless otherwise noted. In addition, an

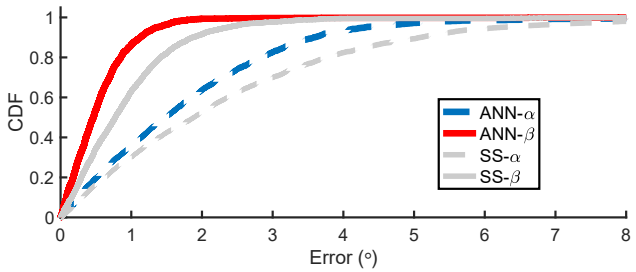


Figure 7-14: AoA Accuracy. AoA errors by using `ANN` and the peak of SS.

ImpinJ 420 reader is utilized to activate RFID tags.

Learnware Settings: As mentioned, we use the OptiTrack system [195] to collect the ground truth. We collect an entirely diverse dataset engaged in the five types of IoT devices. The dataset contains $270,938$ samples, of which $84,280$ samples are used to train the two networks, and the remaining are utilized for testing. We release iArk code and dataset in [203]. The samples are nearly uniformly distributed in the 3D space. The longest duration of continuous testing spans across 48 h. The `ANN` and the `TNN` are trained using the ADAM optimizer with a learning rate of 0.01 and an epoch size of 200. Residual connection and batch normalization are adopted in the `ANN` to benefit the training. The `ANN` configuration refers to the 18-layer setting shown in Table 1 in [194]. The training is performed in the GPU and takes approximately 1 h, while the testing can be accomplished using the CPU. To save on cost and time, the training workload is advised to be performed in the cloud in practice.

7.7 Results

We empirically evaluate the performance of the iArk prototype by conducting experiments in our office of 10×20 m^2, which currently accommodates 20 people. Fig. 7-13a shows the scenario, which is full of different types of indoor reflectors including tables, chairs, computers, and metal obstacles.

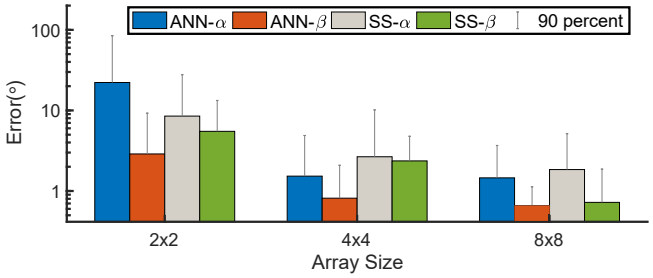


Figure 7-15: Accuracy vs Array Size

7.7.1 Accuracy of AoA Estimation

First, we evaluate the capability of iArk to estimate AoA by using the antenna array. The transmitting device is placed 8 m away from the antenna array. We place the device at different angles and estimate the AoA along the direct path. The ground truth is calculated on the basis of the physical locations of the device and the antenna array.

■ **Angular Accuracy**. Fig. 7-14 plots the CDFs of the AoA estimation errors as determined by the peak of the SS and the `ANN`. We observe the following findings. (a) If the SS is used, then the median errors of the azimuthal (i.e., α) and elevation (i.e., β) angles are 1.87° and 0.77°, and the 90 percentile errors are 5.1° and 1.87°, respectively. (b) If the `ANN` is used, then the median errors are 1.5° and 0.46°, and the 90 percentile errors are 3.57° and 1.13° in the two angles, respectively. We see 20% and 40% improvements in the estimation. In either case, the azimuthal angle is better than the elevation angle. This phenomenon may be attributed to the smaller searching space of the elevation angle (i.e., $0° \sim 90°$) than that of the azimuthal angle (i.e., $0° \sim 360°$).

■ **Impact of Array Size**. We examine the impact of array size on accuracy. We select the signals from the adjacent 2×2, 4×4, and 8×8 antenna elements to estimate the AoA. Fig. 7-15 plots the errors of AoA estimation for the three cases. When the SS is used directly, the median errors of the azimuthal angle are 8.5°, 2.6° and 1.8°; those of the elevation angle are 5.5°, 2.3° and 0.7°. By contrast, when `ANN` is used, the median errors of the azimuthal angle are 22.1°, 1.5° and 1.4°; those of the elevation angle are 2.8°, 0.5° and 0.4°. Accuracy is

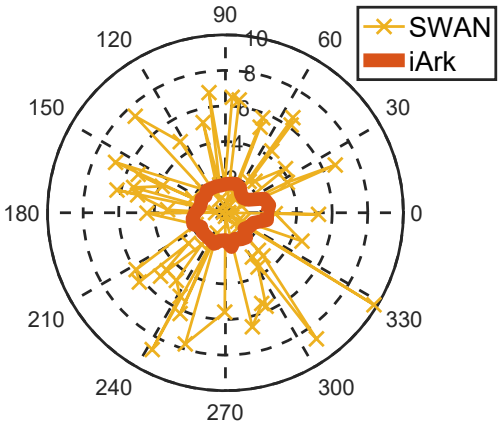
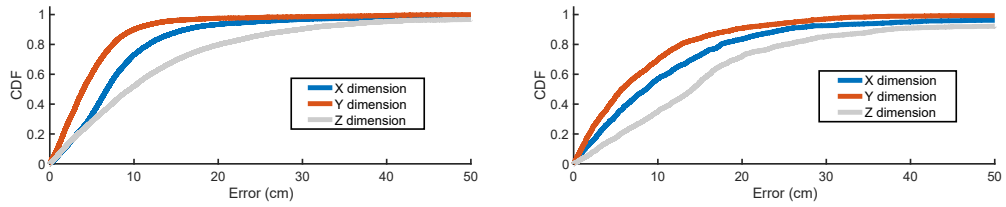


Figure 7-16: Comparison to SWAN

evidently improved when additional elements are used because the LOS paths are consistent with the ground truth in different positions, but the NLOS paths are different. This condition is equivalent to increasing the spatial diversity. Only 6% improvement is obtained by the 8×8 array compared with the 4×4 array, but the manufacturing cost is nearly quadruple. A trade-off between accuracy and cost should be made with respect to the practical demand. Here, we choose to use logical 4×4 arrays as described above.

■ **Comparison with State of the Art**. We compare the results of iArk and the recent work (i.e., SWAN) in Fig. 7-16. SWAN extends the commercial Wi-Fi system to an antenna array through RF switches. Given that SWAN works for 2D localization, the figure only plots the comparison of the azimuthal results. [3] The average error of SWAN is 3.98° with a standard deviation (SD) of 2.45°, whereas the error of iArk is $1.68 \pm 0.28°$. iArk performs significantly better than SWAN in all directions. In particular, iArk is uniformly accurate across all the test directions and has a maximum error of 2.51°. It can provide a 360° field of view at azimuth with high accuracy. The SWAN also adopts a switched antenna array, which is manipulated by an Arduino with a scheduling delay of $29.5\mu s$. By contrast, our system uses the FPGA for scheduling with a delay of $3\mu s$. Thus, iArk can acquire more samples. The superiority of iArk can also be attributed to

[3]Note that the results of SWAN shown in Fig. 7-16 are from Y. Xie [181].



(a) LOS **(b) NLOS**

Figure 7-17: Accuracy in 3D Localization. Localization errors in the (a) LOS and (b) NLOS settings.

the precise PPE.

7.7.2 Accuracy of 3D Tracking

We then focus on the developed 3D tracking primitive and evaluate its accuracy across three dimensions. The OptiTrack systems are set to track the target accurately only when it moves within an 8×8 m^2 area in the room. Given that OptiTrack provides the ground truth in our experiment, we move the target in an area that is 1 m away from the antenna array. Thus, the minimum separation between the antenna array and the device in this experiment is 1 m and the maximum separation is approximately 9 m. The influence of distance is discussed later.

■ **Dimensional Accuracy.** Fig. 7-17 plots the CDFs of the location error along the X, Y and Z coordinates in the two LOS and NLOS settings. For NLOS, we block the direct paths using a wooden obstacle. iArk median location errors for the LOS experiments are 6.8, 3.9, and 9.5 cm along the X, Y, and Z dimensions, respectively. By contrast, the median location errors in the NLOS experiments are 8.7, 5.8, and 14.2 cm along the three dimensions. As expected, the location accuracy in LOS is higher than when the device is behind an obstacle due to the additional attenuation, which reduces the SNR. However, the median error in both settings is small.

■ **Impact of Protocol.** We then evaluate the impact of the protocol on localization accuracy. Fig. 7-18 plots the location error in 3D. The median errors of NB-IoT, LoRa, RFID, Zigbee and Sigfox are 16.4, 16.0, 19.7, 13.9 and 16.9 cm,

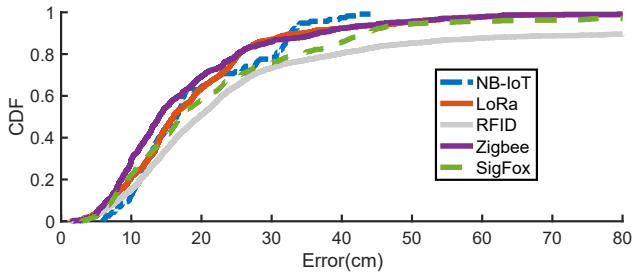


Figure 7-18: Impact of Protocols

respectively. The difference in localization error across protocols is less than 6 cm, which demonstrates that the performance of tracking algorithms is completely protocol-free. In particular, the RFID performs relatively poorly because the tags are battery-free, and their signals are weak.

■ **Comparison with State of the Art**. We compare iArk with several state-of-the-art indoor localization systems, listed in Table 7.2. The table shows that iArk can achieve the best performance in terms of accuracy in localization and AoA estimation for commercial devices. All past systems only support a single type of device (i.e., Wi-Fi receivers or RFID tags), whereas our system can serve all types of IoT devices concurrently. The superiority of our system can be attributed to two main reasons. First, stable RF phase measurements are crucial in the upper-layer algorithm. Our PPE powered by the dual-channel hardware design effectively removes all negatives that may affect the phase. It can provide additional uncontaminated and reliable measured data compared with

Table 7.2: Comparison with State of the Art

Systems	Model	Target	Location (3D)		AoA	
			50^{th}	90^{th}	50^{th}	90^{th}
WiTrack [187]	ToF	Passive	22 cm	62.9 cm	11.5°	37.9°
ArrayTrack [180]	AoA	Wi-Fi	23 cm	80 cm	-	-
LTEye [190]	SAR	LTE	61 cm	104 cm	6.9°	12.2°
SWAN [181]	AoA	Wi-Fi	45 cm	65 cm	2.6°	5.7°
mD-Track [204]	Hybrid	Wi-Fi	36 cm	71 cm	3.3°	12°
xArray [205]	Beamform	RFID	-	150 cm	-	-
RFind [96]	UWB	RFID	1.9 cm	4.9 cm	-	-
TurboTrack [206]	UWB	RFID	0.5 cm	1.1 cm	-	-
iArk	AoA+DL	IoT	16.2 cm	35 cm	1.5°	3.5°

[1] The above results are directly from the reported papers respectively except RFind and iArk due to the lack of essential hardwares.

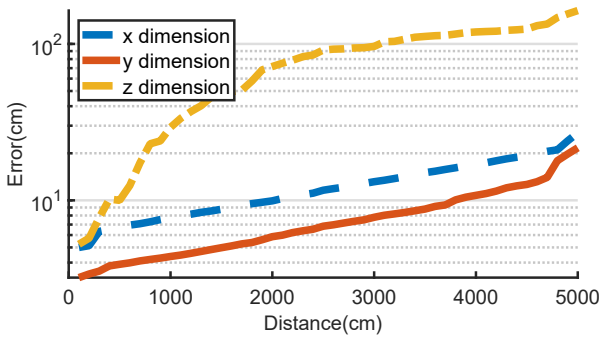


Figure 7-19: Accuracy vs. Distance

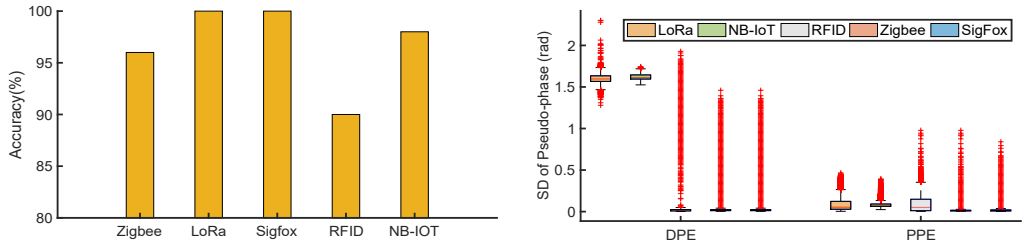
the previous work. Second, most past systems still use the geometric model, and thus fall short of real-life indoor conditions, especially in the presence of strong multipath propagations. We introduce the deep learning technique to identify the LOS path, which demonstrates its effectiveness.

7.7.3 Accuracy vs. Distance

We are interested in evaluating the accuracy of iArk as the device moves farther away from the array. We repeat the above-mentioned experiments as a function of distance. As mentioned, OptiTrack requires the device to move within a certain area that is in the LOS of IR cameras. Thus, we move the device and the OptiTrack systems away from the antenna array. The positions with distances to the array greater than 25 m are tested at the adjacent room. Fig. 7-19 illustrates the localization error of iArk as a function of its distance to the antenna array. The median errors of the estimation for the X, Y, and Z dimensions are shown.

■ **X/Y Dimension.** The accuracy along the X-axis (and Y-axis) changes from 5.0 cm (and 3.2 cm) to 27.1 cm (and 21.6 cm) for distances of 1 m to 50 m away from the device. The distance does not affect the accuracy along X and Y coordinates significantly. The decrease is caused by the range-dependent signal attenuation.

■ **Z Dimension.** By contrast, the error along the Z dimension is not as good as the other two dimensions. The error increases from 5.2 cm to 163.4 cm. This is because the Z dimension indicates the depth – the distance between the antenna



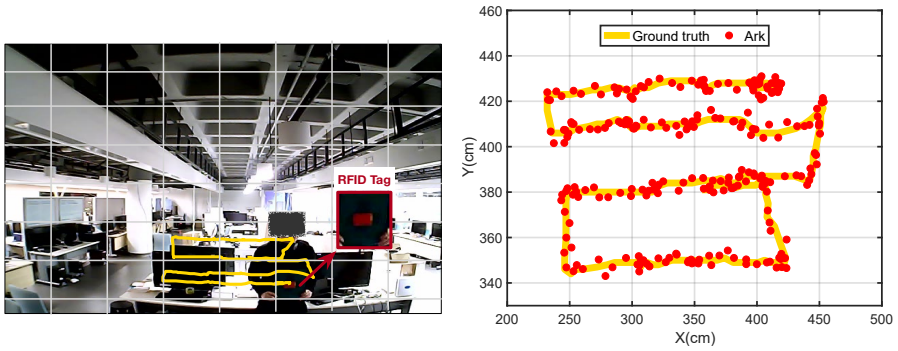
(a) Packet Detection **(b) Phase Estimation**

Figure 7-20: Accuracy vs Protocol.

array and the device (Fig. 7-3). Our 25 logical antenna arrays are deployed on the same side. As a result, the intersection of lines along the determined AoAs is a sharp and long zone along the Z dimension. We argue that this phenomenon is similar to that of the GPS wherein all satellites fly above the Earth's surface (i.e., one-sided). Consequently, the longitude and latitude results of the GPS are significantly better than the elevation result. To reduce the impact of depth, the best deployment approach is to deploy the antenna array at the ceiling. Note that we only present the results of up to 50 m distance, which works for most of the large-sized warehouses. In our future work, we will continue to extend iArk to larger spaces.

7.7.4 Accuracy vs. Protocol

A key feature of iArk is its capability to remove protocol diversity. In this study, we examine the performance of packet detection and phase estimation with respect to different wireless protocols. We place five types of devices in fixed positions and run iArk. The average packet detection accuracy among five protocols is above 90% as shown in Fig. 7-20a. Such accuracy is quite enough for most applications. Next, we check the effectiveness of PPE compared with the DPE. Ideally, the estimated pseudo-phase should remain stationary because the device locations do not change. Fig. 7-20b presents the SD of the pseudo-phase over the 180 samples estimated using the two approaches. From the figure, the SD of PPE follows within 0.15 radians, whereas those of DPE vary in 1.6 radians. In



(a) Moving RFID Tags **(b) Zoomed-in Results**

Figure 7-21: iArk in Real-World Applications. (a) shows the scenario where a person moves an RFID tag attached on a black board and the yellow line indicates the trajectory of the tag. (b) shows the comparison of ground truth and the tracking results.

the 32.8 cm wavelength, the two estimation approaches produce potential ranging errors of $0.15/2\pi * 32.8 = 0.78$ cm and $1.6/2\pi * 32.8 = 8.35$ cm. The phase stability of PPE is 90.6% higher than that of DPE. This feature is the key to the high-precision AoA estimation and localization in iArk.

7.7.5 Case Study

Finally, we qualitatively demonstrate iArk in real-world applications. Fig. 7-21 presents a scenario of iArk for tracking an RFID tag. The left figure shows that the tag is moved freely in front of the antenna array, while the right figure exhibits the tracking results. iArk can achieve the same level of accuracy reported in the quantitative results above.

7.8 Related Work

In this section we survey prior works in location estimation and phase calibration, and describe how these systems relate to iArk.

■ **RF Localization.** Many studies have been conducted on indoor localization. They can be grouped into three main categories.

(1) Trilateration: In this category, the receiver estimates the distances to at least three anchors, and then locates itself at the intersection of the multiple

spheres. Studies in this category usually require a GHz bandwidth (e.g., $>$ 2 GHz [21, 207–212]), to estimate the time of flight. For example, RFind [96] and TurboTrack [206] observe that the RFID tags can respond in a wider band in practice than they claim in the Gen 2 protocol. Inspired by this, these two works emulate hundreds of MHz bandwidth on tags to locate tags via UWB. Unfortunately, the majority of IoT devices still work at a narrow band of less than 4 MHz. Thus, the trilateration or UWB techniques cannot be used as a general approach for localizing IoT devices in practice.

(2) Signature: The second category collects the signal signatures in the entire surveillance region in advance and locates a running IoT device at the position where the signal of the device most matches the stored signature [126, 213–219]. For example, PinIt [126] captures and extracts the multipath profiles as signatures by using an antenna array, with the intuition that nearby RFIDs experience a similar multipath environment. As a general-purpose platform, iArk can also work for these signature based localization solutions. Particularly, iArk can increase the dimension of signature because it can acquire the signals of IoT devices from different angles, and thereby improve the matching accuracy potentially. In our future work, we will study how could iArk benefits the signature based localization.

(3) Triangulation. Similar to ours, the final category leverages an antenna array to estimate the AoA of the wireless signal [180, 181, 183–187, 191]. For example, DOF [191] identifies unknown radios through the hidden and repeating patterns in their signals and furhter locates them through AoAs. A recent work [220] constructed a phased array on commodity Wi-Fi with three antennas for localization. ArrayTrack [180] and SWAN [181] are the most similar systems to ours. However, iArk is superior to the two works in terms of hardware and middleware. First, the array sizes of ArrayTrack and SWAN are limited to 16 and 12 respectively. On the contrary, our antenna array is 4$\times$ and 5$\times$ larger than them in size. As demonstrated, the array size is a key factor of accuracy. To the best of our knowledge, ours is the first real large-sized antenna array designed

for studying the localization of IoT devices in the community. In particular, ArrayTrack uses 16 standalone ADCs to construct the RF frontend, with a cost that is 8× higher than ours. Second, ArrayTrack and SWAN target the localization of Wi-Fi devices without the challenge of protocol diversity, while one of our contributions is in the pervasiveness derived from the cross-protocol design. Phaser [220] constructs an antenna array by using multiple receiving Wi-Fi radios. The most challenging issue that Phaser addresses is how to overcome the phase difference among the receiving multi-radios caused by the unsynchronized and stand-alone oscillators. iArk contains two receiving channels (or radios), both of which share a single oscillator, thus this issue does not exist in our system. As opposed to Phaser, we take advantage of the phase difference between two channels caused by the difference in locations of two antennas (rather than the unsynchronization) to eliminate the protocol diversity. LTEye [190] can monitor the 3-dimensional physical locations of LET transmitters. Our design is similar to this work but differs in three aspects. First, from the view of hardware, iArk is 250× efficient than LTEye in signal acquisition. iArk uses a real antenna array while LTEye adopts a movable antenna. As a result, iArk only takes 2 ms on the signal acquisitions, but LTEye requires 500 ms at least. Second, from the view of middleware, LTEye also utilizes a static antenna as a reference to overcome the CFO. Beyond this, iArk extends the relative signal to eliminate the protocol diversity including dependences of preamble, modulation, channel and devices. We attempt to resolve the newly emerging localization issues in IoT even if using a similar technique. Third, from the view of learnware, the localization results of iArk are much more accurate, outperforming LTEye by about 4×, which benefits from the well-designed hardware and the deep learning.

■ **Coexistence.** Many previous works [8, 26, 28, 29, 39, 40, 42, 43] study the coexistence and/or mutual communications of multiple diverse wireless technologies (e.g., Zigbee, WiFi) by designing more efficient protocols. The main goal of these work is to improve the utilization of the spectrum. In contrast, iArk focuses on locating diverse IoT devices regardless which protocols they are using. Thus,

iArk can work with these previous works.

7.9 Discussion

■ **Limitation.** iArk leaves room for further investigations as discussed below:

(1) Dynamic environment. Neural networks assume that the learned models are static over time. However, in a complex indoor building, the environment is always dynamic, due to the unpredictable movements of people and radio interference. Thus, the propagation paths and distributions of phase values at training and test periods may be significantly different. This difference causes the learned models to fail in the test period if using the data collected at the trained period. Prior work in machine learning has demonstrated that information learned at the early layers of a neural network is used to model the input and mostly independent of the output, whereas the later layers are more specialized. This phenomenon inspires the community to transfer the learning to accommodate dynamic [219, 221]. The basic idea is to recollect a few new training data in the new environment to update the last few layers while maintaining the previous layers unchanged [222]. We plan to adopt the transfer learning to deal with the dynamics in our future work.

(2) Scalability. Acting as a sniffer, the platform does not need to know which device transmits the signal or whether a signal collision occurs. The platform neither introduces additional traffics or collisions nor requires cooperation from other devices. The platform simply stores all location results over the phase matrices in the database with a timestamp. Thus, the platform can support any number of devices as long as they could successfully communicate with each other. Thus, our design is essentially scalable for any kind of IoT network with a large number of devices.

■ **Conclusion.** In this chapter, we presents iArk, which is a general-purpose platform for tracking IoT devices across protocols. To the best of our knowledge, this study is the first to propose an all-in-one platform that can track all types

of IoT devices. The platform puts the cross-technology sensing into practice. The design of iArk introduces three key innovations. First, it presents a practical hardware design of an $8 \times 8 + 1$ antenna array. Second, a middleware is designed to remove device diversity and protocol varieties. Third, a deep tracking framework is developed to improve the tracking accuracy and stability. The platform provides a wide range of exciting opportunities for developing tracking systems.

Chapter 8

Conclusions and Future Work

In this chapter, we conclude the whole thesis and look ahead to the future work on cross-technology mutualism in IoT.

8.1 Conclusions

In conclusion, this dissertation proposes a new direction for the development of IoT, that is, building a Cross-Technology Mutualism ecosystem. The core idea is to create new D2D physical layer communication channels between heterogeneous networks. In the envisioned future network, the physical boundaries between devices are broken, and any two IoT devices can communicate with each other freely. Similar to the neural connections in human brains, an increasing number of smart devices become connected together. Unfortunately, past works devoted to this area only consider the way to connect devices working on the same frequency band. This dissertation breaks the stereotype by presenting four CTM systems that bridge the different wireless technologies that even work on different frequencies. Specifically, we make the following contributions:

- *Identifying UHF RFID with WiFi*: We present TiFi, which is the first system that gives commercial WiFi devices the ability to identify the UHF RFID tags. It breaks the huge frequency gap and protocol gaps between WiFi and UHF RFID by exploring the nonlinearity of the RFID tags. We

show that CTM can give rebirth to UHF RFID in the customer market.

- *Activating Wireless Voice for ETC Systems*: We describe Tagcaster, which is a technique that first demonstrates that UHF ETC signal can be captured and interpreted by AM radio receiver with the only software update. The heart of Tagcaster is a new approach to leverage the nonlinearity of the AM radio to naturally downconvert the UHF signal. We demonstrate the improvement brought by CTM in the user experience of machine-to-machine communication such as the ETC system.

- *Rebooting Ultrasonic Positioning for Ultrasound-incapable Devices*: We design UPS+, which is a totally new ultrasonic positioning system that can provide centimeter-level localization service for ultrasound-incapable smart devices. The heart of this system is a new scheme for ultrasound speakers to talk directly to ordinary microphones. UPS+ also designs a concise physical layer protocol for building a localization network. We first demonstrate the extension of the service scope of ultrasonic localization infrastructure brought by CTM.

- *General-purpose Deep Tracking Platform across Protocols*: We present a new system named iArk that can localize multiple IoT devices with different protocols using one set of RF links. We increased the range of applicability of radar systems by CTM design. We use nearly 2 years to build a vast UHF band antenna array with 64 elements and a whole series of RF controlling modules to push the limits of indoor positioning. On the basis of the antenna array, we also develop a deep network for array signal processing and further improve the performance in complex indoor environments. This work is the first use a high accurate array radar to simultaneously service various IoT devices.

8.2 Lessons Learned

This dissertation presents our efforts to achieve CTM across multiple IoT protocols and provide novel cross-technology communication and localization service. Through this process, besides making important research contributions, we have learned several lessons on how to build a better CTM system.

■ **Cooperate, not divide.** We are currently transforming the world with the artificial intelligence of things. Numerous IoT devices are needed to be fully aware of the world for empowering the IoT networks with strong intelligence. This task can only be conducted with the cooperation of multiple IoT technologies. Especially, D2D communication and cooperation will become an essential part of the future hyper-connected world given its low latency and informative sensing. For example, when building an intelligent supermarket, we have WiFi for personal networking and RFID for managing products. Let us build a bridge between RFID and WiFi, similar to what we did in TiFi. Then we can use their D2D communication to collect the products that consumers are interested in and make customized recommendations. Such cooperation between devices is significant. The intelligence comes when they can exchange data and sense each other through this channel. Past fragmented development of IoT can not follow up the growing demand for intelligence. More cohesive cooperation between devices must be conducted in the future IoT world, and we believe the whole IoT ecosystem will finally become a whole intelligent agent.

■ **Generalization and virtualization.** Some fundamental network services such as localization should be generalized. Although different IoT devices work with different protocols, the way we localize them is very similar. One device can be extended to provide many services from the aspects of the hardware. Hence, network designers could consider more about how to make the most of the existing wireless infrastructures with the software update rather than add more. According to our experience in iArk, such generalization can reduce redeployment, save costs, and improve system efficiency. Beyond the generalization, we could do

the network function virtualization further to improve the scalability and agility of the system. Users do not need to equip dedicated hardware for every network function, and service providers can deliver new network services and applications on demand without additional hardware resources.

■ **Considering hardware and software as a whole.** The one who is really serious about wireless network design should immerse himself in communication hardware rather than just focus on the upper layer software. TiFi, Tagcaster, and UPS+ show that the nonlinearity of hardware can be used to achieve CFC. Of course, such idea only comes when we have a deep understanding of the hardware. Current communication hardware is highly integrated and modular. We can easily build our own communication system by combining existing modules. This approach does not mean we really understand our hardware. Nonlinearity is a good start for exploring hardware characteristics for novel communication. We believe many other interesting hardware phenomena need to be discovered. If we can design the matching software, then novel communication comes.

■ **Communication is sensing.** Wireless communication and sensing are born together from the perspective of utilizing channels. In wireless communication, channel estimation is conducted to eliminate wireless channel interference for the maximum data transmission rate. Meanwhile, in wireless sensing, channel estimation is performed to explore the information hidden in the channel. Thus, when we talk about communication, sensing also comes. An example can be found in UPS+. At first, the researchers only consider using the connection between microphone and ultrasound for cross-frequency acoustic communication and inaudible acoustic attack. Most works in this area focus on conducting cross-technology communication. We go in another direction by embracing cross-technology communication as a chance for cross-technology localization. Such cross-technology sensing is found to have a significant benefit for the usability of the traditional localization system. We believe our works are only the tip of the iceberg, and broad opportunities for deep cross-technology sensing are need to be explored.

8.3 Future Work

IoT is at an exciting time. The rapid development of wireless communication and computing hardware has facilitated the Internet of Everything. This process introduces new opportunities and challenges. As an increasing number of diverse devices and platforms connect to the Internet, building a more efficient, harmonic, and smart wireless network is becoming much more challenging. One important direction is to build an IoT ecosystem where technologies are closely connected and mutually beneficial. In this dissertation, we made some attempts on limited CTM. However, the way to connect a large-scale world without limits is only starting. We can expand the boundaries of IoT and build a truly intelligent network in the following areas.

8.3.1 Zero-Limit connectivity

The wireless connection at present still has many constraints, which limit our progress in building an intelligent IoT. First, the seamless connection between IoTs is very difficult. As discussed in the above chapters, diverse IoT technologies are point-specific solutions that target particular applications. The huge difference in hardware and software architectures causes difficulty in bridging them together. Thus, a new agglomerate architecture that can integrate different wireless platforms across layers should be built for the IoT network. Software-defined radio is a very promising solution to achieve such grand unification. It enables communication tasks as much as possible to be performed by flexible software. The design of iArk partly originates from this idea. This area can still be explored. For example, there are numerous existing backscatter platforms targeting to different connection demands (e.g., WiFi [49, 88], Bluetooth [20, 48], Zigbee [223], LoRa [87, 224], and RFID). Building software-defined general-purpose backscatter nodes, which can switch the supporting communication protocols with only software updates, will dramatically improve the ultra-low-power network adaptivity. TinySDR [225] is a good attempt in this direction, which builds a low-

power software-defined radio with over-the-air programmability. We could move forward by designing a software define backscatter RF frontend.

Second, the IoT network is very crowded in the time and frequency domain. The increase in IoT network capacity can not keep pace with the fast-growing demand being put on it. Especially, the low-power IoT network (e.g., RFID) has very poor computing and network schedule ability, which results in the low utilization of resources. The throughput of a low-power IoT network is too low to adjust the large-scale deployment. We are thirsty for a new resource scheduling scheme to develop the enormous potentials of wireless resources further for achieving optimal resource allocation regardless of situation. Such an architecture should automatically adapt to the demanding of user and be globally optimizied at all levels of the networks. My previous works [78,226,227] make some attempts in this area, but this effort is far from our vision.

8.3.2 Cooperative Wireless Sensing

In future, every transceiver should participate in sensing for a truly smart IoT network. The deep collaboration of the whole wireless network can be very powerful for sensing. Take the antenna array as an example. We use an antenna array to locate by leveraging the spatial sampling to discern the direction of the signal. The positioning accuracy is better when the array is larger. Thus, iArk builds an antenna array with 64 elements to achieve a higher positioning performance. Building a large antenna array is costly and space-consuming. A better idea is to combine the numerous existing low-cost wireless devices around us to form a large virtual antenna array. In the past, this devices are used individually given that they may have different protocols and operate in different frequency bands. If we can let the heterogeneous wireless devices sense the signal collaboratively, then obtaining very high accurate sensing ability with the large array aperture will be much easier. To achieve such a goal, we should further break the boundaries between the devices and empower wireless devices to sense the signals of heterogeneous devices. Additionally, unlike the shape of traditional centralized

antenna array, that of the virtual array is irregular, and the carrier frequency for each element may also be different. We need a new array signal processing theory to adapt it.

8.3.3 Wireless Network Function Virtualization

Network function virtualization (NFV) is a future direction for maximizing the resource utilization, agility, and scalability of network infrastructures. With the generalization of wireless devices, we could achieve the NFV for wireless communication and sensing. The RF hardware and baseband computing resources would be virtualized and uniformly scheduled by network service providers. By then, the connection between devices would be dynamically configured by software according to the communication need and channel state.

8.3.4 Security and Privacy Inherent in Everything

Wireless security is an eternal topic, and it should be embedded in all areas. However, building a secure and reliable IoT ecosystem is challenging due to the huge gap in computing ability between devices. Especially, low-power IoT devices can not support complex security methods and are easy to be attacked. Along with the CTM idea, a good way to solve this problem is to let the powerful IoT devices, such as a gateway, help authenticate the low-power device and defense against the attack. For example, we could extract the hardware fingerprints [228, 229] and physical position information as a reliable authentication scheme. In the future, we need to design an asymmetric security scheme for adjusting the miscellaneous IoT devices.

Bibliography

[1] J.-M. Laheurte, C. Ripoll, D. Paret, and C. Loussert, *UHF RFID technologies for identification and traceability.* John Wiley & Sons, 2014.

[2] P. Lazik and A. Rowe, "Indoor pseudo-ranging of mobile devices using ultrasonic chirps," in *Proc. of ACM Sensys*, 2012.

[3] "Cisco Annual Internet Report (2018–2023) White Paper," https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html.

[4] E. Elbasani, P. Siriporn, and J. S. Choi, "A survey on rfid in industry 4.0," in *Internet of Things for Industry 4.0.* Springer, 2020, pp. 1–16.

[5] S. S. I. Samuel, "A review of connectivity challenges in iot-smart home," in *2016 3rd MEC International conference on big data and smart city (ICBDSC).* IEEE, 2016, pp. 1–4.

[6] W. Zhao, S. Lin, J. Han, R. Xu, and L. Hou, "Design and implementation of smart irrigation system based on lora," in *2017 IEEE Globecom Workshops (GC Wkshps).* IEEE, 2017, pp. 1–6.

[7] "What is LoRa?" https://www.semtech.com/lora/what-is-lora, 2021.

[8] S. Gollakota, F. Adib, D. Katabi, and S. Seshan, "Clearing the rf smog: making 802.11 n robust to cross-technology interference," in *Proc. of ACM SIGCOMM*, 2011, pp. 170–181.

[9] A. Amiruddin, A. A. P. Ratna, R. Harwahyu, and R. F. Sari, "Secure multi-protocol gateway for internet of things," in *2018 Wireless Telecommunications Symposium (WTS).* IEEE, 2018, pp. 1–8.

[10] K. Cui, Y. Wang, Y. Zheng, and J. Han, "Shakereader:'read'uhf rfid using smartphone," *IEEE Transactions on Mobile Computing*, 2021.

[11] M. Aly, F. Khomh, Y.-G. Guéhéneuc, H. Washizaki, and S. Yacout, "Is fragmentation a threat to the success of the internet of things?" *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 472–487, 2018.

[12] Z. Yang, Z. Zhou, and Y. Liu, "From rssi to csi: Indoor localization via channel response," *ACM Computing Surveys (CSUR)*, vol. 46, no. 2, pp. 1–32, 2013.

[13] J. Holland and J. L. Bronstein, "Mutualism," in *Encyclopedia of Ecology, Five-Volume Set.* Elsevier Inc., 2008, pp. 2485–2491.

[14] "What is beyond Hyper-Connectivity?" http://www.6gsummit.com/2019/wp-content/uploads/2019/04/Day3_Session3_Dong_Seung_ETRI.pdf, 2019.

[15] "ImpinJ, Inc," http://www.impinj.com/, 2019.

[16] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, "The cricket location-support system," in *Proc. of ACM MobiCom*, 2000.

[17] J. Yang, S. Sidhom, G. Chandrasekaran, T. Vu, H. Liu, N. Cecan, Y. Chen, M. Gruteser, and R. P. Martin, "Detecting driver phone use leveraging car speakers," in *Proc. of ACM MobiCom*, 2011.

[18] P. Lazik, N. Rajagopal, B. Sinopoli, and A. Rowe, "Ultrasonic time synchronization and ranging on smartphones," in *Proc. of IEEE RTAS*, 2015.

[19] C. Peng, G. Shen, Y. Zhang, Y. Li, and K. Tan, "Beepbeep: a high accuracy acoustic ranging system using cots mobile devices," in *Proc. of ACM Sensys*, 2007.

[20] V. Iyer, V. Talla, B. Kellogg, S. Gollakota, and J. Smith, "Inter-technology backscatter: Towards internet connectivity for implanted devices," in *Proceedings of the 2016 ACM SIGCOMM Conference*, 2016, pp. 356–369.

[21] X. Hui and E. C. Kan, "Radio ranging with ultrahigh resolution using a harmonic radio-frequency identification system," *Nature Electronics*, vol. 2, no. 3, pp. 125–131, 2019.

[22] A. Hithnawi, H. Shafagh, and S. Duquennoy, "Understanding the impact of cross technology interference on ieee 802.15. 4," in *Proceedings of the 9th ACM international workshop on Wireless network testbeds, experimental evaluation and characterization*, 2014, pp. 49–56.

[23] K. Lakshminarayanan, S. Sapra, S. Seshan, and P. Steenkiste, "Rfdump: an architecture for monitoring the wireless ether," in *Proceedings of the 5th international conference on Emerging networking experiments and technologies*, 2009, pp. 253–264.

[24] X. Zhang and K. G. Shin, "Enabling coexistence of heterogeneous wireless systems: Case for zigbee and wifi," in *Proceedings of the Twelfth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, 2011, pp. 1–11.

[25] J. Huang, G. Xing, G. Zhou, and R. Zhou, "Beyond co-existence: Exploiting wifi white space for zigbee performance assurance," in *The 18th IEEE International Conference on Network Protocols*. IEEE, 2010, pp. 305–314.

[26] X. Zhang and K. G. Shin, "Gap sense: Lightweight coordination of heterogeneous wireless devices," in *Proc. of IEEE INFOCOM*, 2013.

[27] C. Won, J. Youn, H. Ali, H. Sharif, and J. Deogun, "Adaptive radio channel allocation for supporting coexistence of 802.15. 4 and 802.11 b," in *IEEE Vehicular Technology Conference*, vol. 62, no. 4. IEEE; 1999, 2005, p. 2522.

[28] S. M. Mishra, R. W. Brodersen, S. ten Brink, and R. Mahadevappa, "Detect and avoid: an ultra-wideband/wimax coexistence mechanism," *IEEE Communications Magazine*, vol. 45, no. 6, pp. 68–75, 2007.

[29] H. Rahul, N. Kushman, D. Katabi, C. Sodini, and F. Edalat, "Learning to share: narrowband-friendly wideband networks," *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 4, pp. 147–158, 2008.

[30] L. Cao, L. Yang, and H. Zheng, "The impact of frequency-agility on dynamic spectrum sharing," in *2010 IEEE Symposium on New Frontiers in Dynamic Spectrum (DySPAN)*. IEEE, 2010, pp. 1–12.

[31] L. Yang, W. Hou, L. Cao, B. Y. Zhao, and H. Zheng, "Supporting demanding wireless applications with frequency-agile radios." in *NSDI*, 2010, pp. 65–80.

[32] R. Chandra, R. Mahajan, T. Moscibroda, R. Raghavendra, and P. Bahl, "A case for adapting channel width in wireless networks," *ACM SIGCOMM computer communication review*, vol. 38, no. 4, pp. 135–146, 2008.

[33] T. Moscibroda, R. Chandra, Y. Wu, S. Sengupta, P. Bahl, and Y. Yuan, "Load-aware spectrum distribution in wireless lans," in *2008 IEEE International Conference on Network Protocols*. IEEE, 2008, pp. 137–146.

[34] Y. Yubo, Y. Panlong, L. Xiangyang, T. Yue, Z. Lan, and Y. Lizhao, "Zimo: Building cross-technology mimo to harmonize zigbee smog with wifi flash without intervention," in *Proceedings of the 19th annual international conference on Mobile computing & networking*, 2013, pp. 465–476.

[35] S. Katti, S. Gollakota, and D. Katabi, "Embracing wireless interference: Analog network coding," *ACM SIGCOMM Computer Communication Review*, vol. 37, no. 4, pp. 397–408, 2007.

[36] S. Katti, D. Katabi, H. Balakrishnan, and M. Medard, "Symbol-level network coding for wireless mesh networks," *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 4, pp. 401–412, 2008.

[37] J. Ou, M. Li, and Y. Zheng, "Come and be served: Parallel decoding for cots rfid tags," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, 2015, pp. 500–511.

[38] M. Jin, Y. He, X. Meng, Y. Zheng, D. Fang, and X. Chen, "Fliptracer: Practical parallel decoding for backscatter communication," *IEEE/ACM Transactions on Networking*, vol. 27, no. 1, pp. 330–343, 2019.

[39] Z. Li and T. He, "Webee: Physical-layer cross-technology communication via emulation," in *Proc. of ACM MobiCom*, 2017.

[40] S. M. Kim and T. He, "Freebee: Cross-technology communication via free side-channel," in *Proc. of ACM MobiCom*, 2015.

[41] Z. Chi, Y. Li, H. Sun, Y. Yao, Z. Lu, and T. Zhu, "B2w2: N-way concurrent communication for iot devices," in *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems*, 2016, pp. 245–258.

[42] K. Chebrolu and A. Dhekne, "Esense: Communication through energy sensing," in *Proc. of ACM MobiCom*, 2009.

[43] Y. Zhang and Q. Li, "Howies: A holistic approach to zigbee assisted wifi energy savings in mobile devices," in *Proc. of IEEE INFOCOM*, 2013.

[44] Z. Yin, W. Jiang, S. M. Kim, and T. He, "C-morse: Cross-technology communication with transparent morse coding," in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE, 2017, pp. 1–9.

[45] W. Jiang, Z. Yin, S. M. Kim, and T. He, "Transparent cross-technology communication over data traffic," in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE, 2017, pp. 1–9.

[46] W. Jiang, Z. Yin, R. Liu, Z. Li, S. M. Kim, and T. He, "Bluebee: a 10,000 x faster cross-technology communication via phy emulation," in *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*, 2017, pp. 1–13.

[47] J. Shi, D. Mu, and M. Sha, "Lorabee: Cross-technology communication from lora to zigbee via payload encoding," in *2019 IEEE 27th International Conference on Network Protocols (ICNP)*. IEEE, 2019, pp. 1–11.

[48] P. Zhang, M. Rostami, P. Hu, and D. Ganesan, "Enabling practical backscatter communication for on-body sensors," in *Proceedings of the 2016 ACM SIGCOMM Conference*, 2016, pp. 370–383.

[49] D. Bharadia, K. R. Joshi, M. Kotaru, and S. Katti, "Backfi: High throughput wifi backscatter," *ACM SIGCOMM Computer Communication Review*, vol. 45, no. 4, pp. 283–296, 2015.

[50] V. Liu, A. Parks, V. Talla, S. Gollakota, D. Wetherall, and J. R. Smith, "Ambient backscatter: wireless communication out of thin air," in *Proc. of ACM SIGCOMM*, 2013.

[51] B. Kellogg, A. Parks, S. Gollakota, J. R. Smith, and D. Wetherall, "Wi-fi backscatter: Internet connectivity for rf-powered devices," in *Proceedings of the 2014 ACM conference on SIGCOMM*, 2014, pp. 607–618.

[52] A. Wang, V. Iyer, V. Talla, J. R. Smith, and S. Gollakota, "Fm backscatter: Enabling connected cities and smart fabrics." in *Proc. of ACM NSDI*, 2017.

[53] A. Varshney, C. P. Penichet, C. Rohner, and T. Voigt, "Towards wide-area backscatter networks," in *Proceedings of the 4th ACM Workshop on Hot Topics in Wireless*, 2017, pp. 49–53.

[54] S. Thomas and M. S. Reynolds, "Qam backscatter for passive uhf rfid tags," in *2010 IEEE International Conference on RFID (IEEE RFID 2010)*. IEEE, 2010, pp. 210–214.

[55] R. Zhao, F. Zhu, Y. Feng, S. Peng, X. Tian, H. Yu, and X. Wang, "Ofdma-enabled wi-fi backscatter," in *The 25th Annual International Conference on Mobile Computing and Networking*, 2019, pp. 1–15.

[56] B. Kellogg, V. Talla, S. Gollakota, and J. R. Smith, "Passive wi-fi: Bringing low power to wi-fi transmissions." in *Proc. of ACM NSDI*, vol. 16, 2016, pp. 151–164.

[57] V. Talla, J. Smith, and S. Gollakota, "Advances and open problems in backscatter networking," *arXiv preprint arXiv:2011.03242*, 2020.

[58] G. A. Vera, Y. Duroc, and S. Tedjini, "Third harmonic exploitation in passive uhf rfid," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 9, pp. 2991–3004, 2015.

[59] N. Roy, H. Hassanieh, and R. Roy Choudhury, "Backdoor: Making microphones hear inaudible sounds," in *Proc. of ACM MobiSys*, 2017.

[60] T. Eriksson, W. Cao, and C. Fager, "Nonlinear effects of wireless transceivers," *Wiley 5G Ref: The Essential 5G Reference Online*, pp. 1–30, 2019.

[61] J. C. Pedro and N. B. Carvalho, *Intermodulation distortion in microwave and wireless circuits*. Artech House, 2003.

[62] H. Sakran, M. Shokair, and A. Abou Elazm, "An efficient technique for reducing papr of ofdm system in the presence of nonlinear high power amplifier," in *2008 9th International Conference on Signal Processing*. IEEE, 2008, pp. 1749–1752.

[63] S. Vasudevan, D. Towsley, D. Goeckel, and R. Khalili, "Neighbor discovery in wireless networks and the coupon collector's problem," in *Proc. of ACM MobiCom*, 2009.

[64] A. Collado and A. Georgiadis, "Optimal waveforms for efficient wireless power transmission," *IEEE Microwave and Wireless Components Letters*, vol. 24, no. 5, pp. 354–356, 2014.

[65] A. Boaventura, A. Collado, N. B. Carvalho, and A. Georgiadis, "Optimum behavior: Wireless power transmission system design through behavioral models and efficient synthesis techniques," *IEEE Microwave Magazine*, vol. 14, no. 2, pp. 26–35, 2013.

[66] N. B. Carvalho, A. Georgiadis, A. Costanzo, H. Rogier, A. Collado, J. A. García, S. Lucyszyn, P. Mezzanotte, J. Kracek, D. Masotti *et al.*, "Wireless power transmission: R&d activities within europe," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 4, pp. 1031–1045, 2014.

[67] G. A. Vera, Y. Duroc, and S. Tedjini, "Analysis and exploitation of harmonics in wireless power transfer (h-wpt): passive uhf rfid case," *Wireless Power Transfer*, vol. 1, no. 2, pp. 65–74, 2014.

[68] Z. Li, Y. Xie, M. Li, and K. Jamieson, "Recitation: Rehearsing wireless packet reception in software," in *in Proc. of ACM MobiCom*, 2015, pp. 291–303.

[69] Y. Ma, X. Hui, and E. C. Kan, "3d real-time indoor localization via broadband nonlinear backscatter in passive devices with centimeter precision," in *Proc. of ACM MobiCom*, 2016.

[70] G. Zhang, C. Yan, X. Ji, T. Zhang, T. Zhang, and W. Xu, "Dolphinattack: Inaudible voice commands," in *Proc. of ACM SIGSAC*, 2017.

[71] N. Roy, S. Shen, H. Hassanieh, and R. R. Choudhury, "Inaudible voice commands: The long-range attack and defense," in *Proc. of USENIX NSDI*, 2018.

[72] R. Liu, Z. Yin, W. Jiang, and T. He, "Wibeacon: expanding ble location-based services via wifi," in *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, 2021, pp. 83–96.

[73] "Stone–Weierstrass theorem," https://en.wikipedia.org/wiki/Stone-Weierstrass_theorem, 2021.

[74] "Intermodulation," https://en.wikipedia.org/wiki/Intermodulation, 2021.

[75] W. Shockley, "The theory of p-n junctions in semiconductors and p-n junction transistors," *Bell System Technical Journal*, vol. 28, no. 3, pp. 435–489, 1949.

[76] M. Lundstrom, "An ebers-moll model for the heterostructure bipolar transistor," *Solid-state electronics*, vol. 29, no. 11, pp. 1173–1179, 1986.

[77] B. Clerckx, "Wireless information and power transfer: Nonlinearity, waveform design, and rate-energy tradeoff," *IEEE Transactions on Signal Processing*, vol. 66, no. 4, pp. 847–862, 2017.

[78] Z. An, Q. Lin, L. Yang, and W. Lou, "Embracing tag collisions: Acquiring bloom filters across rfids in physical layer," in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 2019, pp. 1531–1539.

[79] L. Yang, Y. Chen, X.-Y. Li, C. Xiao, M. Li, and Y. Liu, "Tagoram: Real-time tracking of mobile rfid tags to high precision using cots devices," in *Proc. of ACM MobiCom*, 2014, pp. 237–248.

[80] L. Yang, Y. Li, Q. Lin, X.-Y. Li, and Y. Liu, "Making sense of mechanical vibration period with sub-millisecond accuracy using backscatter signals," in *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, 2016, pp. 16–28.

[81] Frost and Sullivan, "Global rfid healthcare and pharmaceutical marke," *Industry Report*, 2011.

[82] "Phychips Technologies," http://www.phychips.com/applications-main/.

[83] "TSL-1128 Handled Reader," https://www.impinj.com/platform/connectivity/tsl-1128/.

[84] "ALR-S350 Handled Reader," http://www.alientechnology.com/products/readers/alr-s350/.

[85] "ThingMagic M6," https://www.atlasrfidstore.com/thingmagic-m6-uhf-rfid-reader-4-port/.

[86] "Alien," http://www.alientechnology.com, 2017.

[87] V. Talla, M. Hessar, B. Kellogg, A. Najafi, J. R. Smith, and S. Gollakota, "Lora backscatter: Enabling the vision of ubiquitous connectivity," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, pp. 1–24, 2017.

[88] P. Zhang, D. Bharadia, K. Joshi, and S. Katti, "Hitchhike: Practical backscatter using commodity wifi," in *Proc. of ACM SenSys*, 2016.

[89] D. Allane, G. A. Vera, Y. Duroc, R. Touhami, and S. Tedjini, "Harmonic power harvesting system for passive rfid sensor tags," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 7, pp. 2347–2356, 2016.

[90] T.-W. Yoo and K. Chang, "Theoretical and experimental development of 10 and 35 ghz rectennas," *IEEE Transactions on Microwave Theory and Techniques*, vol. 40, no. 6, pp. 1259–1266, 1992.

[91] G. A. Vera, Y. Duroc, and S. Tedjini, "Analysis of harmonics in uhf rfid signals," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 6, pp. 2481–2490, 2013.

[92] P. V. Nikitin and K. Rao, "Harmonic scattering from passive uhf rfid tags," in *Antennas and Propagation Society International Symposium, 2009. APSURSI'09. IEEE*. IEEE, 2009, pp. 1–4.

[93] G. A. Vera, Y. Duroc, and S. Tedjini, "Redundant backscattering modulation of passive uhf rfid tags," in *Microwave Symposium Digest (IMS), 2013 IEEE MTT-S International*. IEEE, 2013, pp. 1–3.

[94] "KYEINSIGHT Technologies," https://www.keysight.com.

[95] "Packaging Material Implication in RFID Operations," http://www.spstj.jp/english/publication/thesis/vol14_No5_40-56.pdf.

[96] Y. Ma, N. Selby, and F. Adib, "Minding the billions: Ultra-wideband localization for deployed rfid tags," in *Proc. of ACM MobiCom*, 2017.

[97] "EPCglobal Gen2 Specification," www.gs1.org/epcglobal, 2004.

[98] D. M. Dobkin, *The RF in RFID: UHF RFID in Practice*. Newnes, 2012.

[99] Q. Lin, L. Yang, H. Jia, C. Duan, and Y. Liu, "Revisiting reading rate with mobility: Rate-adaptive reading in cots rfid systems," in *Proc. of ACM CoNEXT*, ser. CoNEXT '17, 2017.

[100] "Wi-Fi Analyzer," https://play.google.com/store/apps/details?id=com.vrem.wifianalyzer.

[101] "USRP Reader," https://github.com/nkargas/Gen2-UHF-RFID-Reader.

[102] "ImpinJ R420," https://www.impinj.com/platform/connectivity/speedway-r420/.

[103] "Lenovo Moto X Phone," https://www3.lenovo.com/ae/en/phone/moto-phone/phones/Moto-X-Style/p/WMD00000239.

[104] R. Zitouni, S. Ataman, M. Mathian, and L. George, "Radio frequency measurements on a sbx daughter board using gnu radio and usrp n-210," in *Measurements & Networking (M&N), 2015 IEEE International Workshop on*. IEEE, 2015, pp. 1–5.

[105] L. Yang, J. Han, Y. Qi, C. Wang, T. Gu, and Y. Liu, "Season: Shelving interference and joint identification in large-scale rfid systems," in *2011 Proceedings IEEE INFOCOM*. IEEE, 2011, pp. 3092–3100.

[106] M. S. Trotter, J. D. Griffin, and G. D. Durgin, "Power-optimized waveforms for improving the range and reliability of rfid systems," in *RFID, 2009 IEEE International Conference on*. IEEE, 2009, pp. 80–87.

[107] "Comparison of open-source wireless drivers," https://en.wikipedia.org/wiki/Comparison_of_open-source_wireless_drivers.

[108] "Electronic toll collection," https://en.wikipedia.org/wiki/Electronic_toll_collection.

[109] O. Abari, D. Vasisht, D. Katabi, and A. Chandrakasan, "Caraoke: An e-toll transponder network for smart cities," in *ACM SIGCOMM Computer Communication Review*, vol. 45, no. 4. ACM, 2015, pp. 297–310.

[110] "Demo Audio," https://www.youtube.com/watch?v=-0w0sdJlThM.

[111] "ImpinJ R2000 RFID reader chip," https://support.impinj.com/hc/en-us/articles/202755828-Indy-R2000-Datasheet.

[112] "Audio Sampling," https://en.wikipedia.org/wiki/Sampling_(signal_processing)#Audio_sampling.

[113] "SONY Radio," https://www.sony.co.th/en/electronics/radios/icf-p36.

[114] H. Lee, T. H. Kim, J. W. Choi, and S. Choi, "Chirp signal-based aerial acoustic communication for smart devices," in *Proc. of IEEE INFOCOM*, 2015.

[115] Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Transactions on audio, speech, and language processing*, vol. 16, no. 1, pp. 229–238, 2008.

[116] "Official PESQ Tool," https://www.itu.int/rec/T-REC-P.862-200102-I/en.

[117] S. Merchan, A. G. Armada, and J. Garcia, "Ofdm performance in amplifier nonlinearity," *IEEE Transactions on Broadcasting*, vol. 44, no. 1, pp. 106–114, 1998.

[118] D. Vasisht, G. Zhang, O. Abari, H.-M. Lu, J. Flanz, and D. Katabi, "In-body backscatter communication and localization," in *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*, 2018, pp. 132–146.

[119] Z. An, Q. Lin, and L. Yang, "Cross-frequency communication: Near-field identification of uhf rfids with wifi!" in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, 2018, pp. 623–638.

[120] J. Gummeson, J. Mccann, C. Yang, D. Ranasinghe, S. Hudson, and A. Sample, "Rfid light bulb: Enabling ubiquitous deployment of interactive rfid systems," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 2, pp. 1–16, 2017.

[121] Y. Shu, K. G. Shin, T. He, and J. Chen, "Last-mile navigation using smartphones," in *Proc. of ACM MobiCom*, 2015.

[122] J. Haverinen and A. Kemppainen, "Global indoor self-localization based on the ambient magnetic field," *Robotics and Autonomous Systems*, vol. 57, no. 10, pp. 1028–1035, 2009.

[123] C. Zhang and X. Zhang, "Pulsar: Towards ubiquitous visible light localization," in *Proc. of ACM MobiCom*, 2017.

[124] S. Zhu and X. Zhang, "Enabling high-precision visible light localization in today's buildings," in *Proc. of ACM MobiSys*, 2017.

[125] C. Wang, H. Wu, and N.-F. Tzeng, "Rfid-based 3-d positioning schemes," in *Proc. of IEEE INFOCOM*, 2007.

[126] J. Wang and D. Katabi, "Dude, where's my card?: Rfid positioning that works with multipath and non-line of sight," in *Proc. of ACM SIGCOMM*, 2013.

[127] J. Wang, F. Adib, R. Knepper, D. Katabi, and D. Rus, "Rf-compass: robot object manipulation using rfids," in *Proc. ACM MOBICOM*, 2013.

[128] Y. Chen, D. Lymberopoulos, J. Liu, and B. Priyantha, "Fm-based indoor localization," in *Proc. of ACM MobiSys*, 2012.

[129] P. Bahl and V. N. Padmanabhan, "Radar: An in-building rf-based user location and tracking system," in *Proc. of IEEE INFOCOM*, 2000.

[130] K. Chintalapudi, A. Padmanabha Iyer, and V. N. Padmanabhan, "Indoor localization without the pain," in *Proceedings of the sixteenth annual international conference on Mobile computing and networking*. ACM, 2010, pp. 173–184.

[131] T. A. Johnson and P. Seeling, "Localization using bluetooth device names," in *Proc. of ACM MobiHoc*, 2012.

[132] M. Azizyan, I. Constandache, and R. Roy Choudhury, "Surroundsense: mobile phone localization via ambience fingerprinting," in *Proc. of ACM MobiCom*, 2009.

[133] W.-T. Tan, M. Baker, B. Lee, and R. Samadani, "The sound of silence," in *Proc. of ACM Sensys*, 2013.

[134] Y. Gao, J. Niu, R. Zhou, and G. Xing, "Zifind: Exploiting cross-technology interference signatures for energy-efficient indoor localization," in *Proc. of IEEE INFOCOM*, 2013.

[135] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller, "3d tracking via body radio reflections." in *NSDI*, vol. 14, 2014, pp. 317–329.

[136] "MIT Project Oxygen," http://oxygen.csail.mit.edu.

[137] R. Want, A. Hopper, V. Falcao, and J. Gibbons, "The active badge location system," *ACM Transactions on Information Systems*, vol. 10, no. 1, pp. 91–102, 1992.

[138] N. B. Priyantha, A. K. Miu, H. Balakrishnan, and S. Teller, "The cricket compass for context-aware mobile applications," in *Proc. of ACM MobiCom*, 2001.

[139] A. Smith, H. Balakrishnan, M. Goraczko, and N. Priyantha, "Tracking moving devices with the cricket location system," in *Proc. of ACM MobiSys*, 2004.

[140] G. Borriello, A. Liu, T. Offer, C. Palistrant, and R. Sharp, "Walrus: wireless acoustic location with room-level resolution using ultrasound," in *Proc. of ACM MobiSys*, 2005.

[141] S. P. Tarzia, P. A. Dinda, R. P. Dick, and G. Memik, "Indoor localization without infrastructure using the acoustic background spectrum," in *Proc. of ACM MobiSys*, 2011.

[142] M. Hazas and A. Ward, "A novel broadband ultrasonic location system," in *Proc. of ACM Ubicomp*, 2002.

[143] M. McCarthy, P. Duff, H. L. Muller, and C. Randell, "Accessible ultrasonic positioning," *IEEE Pervasive Computing*, vol. 5, no. 4, pp. 86–93, 2006.

[144] A. Mandal, C. V. Lopes, T. Givargis, A. Haghighat, R. Jurdak, and P. Baldi, "Beep: 3d indoor positioning using audible sound," in *Proc. of IEEE CCNC*, 2005.

[145] Q. Lin, L. Yang, and Y. Liu, "Tagscreen: Synchronizing social televisions through hidden sound markers," in *Proc. of IEEE INFOCOM*, 2017.

[146] P. Horowitz and W. Hill, *The art of electronics*. Cambridge Univ. Press, 1989.

[147] "Transducer Datasheet," https://www.murata.com/products/productdetail?partno=MA40S4S.

[148] "Ultrasonic Dynamic Speaker Vifa," http://www.avisoft.com/usg/vifa.htm.

[149] "TI PTP Application Report," http://www.ti.com/lit/an/snla098a/snla098a.pdf.

[150] V. Lalitha and S. Kathiravan, "A review of manchester, miller, and fm0 encoding techniques," *SmartCR*, vol. 4, no. 6, pp. 481–490, 2014.

[151] R. G. Brown, P. Y. Hwang *et al.*, *Introduction to random signals and applied Kalman filtering*. Wiley New York, 1992, vol. 3.

[152] M. Jeub, C. Nelke, H. Krüger, C. Beaugeant, and P. Vary, "Robust dual-channel noise power spectral density estimation," in *2011 19th European Signal Processing Conference*. IEEE, 2011, pp. 2304–2308.

[153] "ESP32-DevKitC," https://www.espressif.com/en/products/hardware/esp32-devkitc/overview.

[154] R. Nandakumar, S. Gollakota, and N. Watson, "Contactless sleep apnea detection on smartphones," in *Proc. of ACM MobiSys*, 2015.

[155] P. Lazik, N. Rajagopal, O. Shih, B. Sinopoli, and A. Rowe, "Alps: A bluetooth and ultrasound platform for mapping and localization," in *Proc. of ACM Sensys*, 2015.

[156] J. Xiong and K. Jamieson, "Arraytrack: a fine-grained indoor location system," in *Proc. of USENIX NSDI*, 2012.

[157] "Bluetooth iBeacon," https://estimote.com/.

[158] M. Hazas and A. Ward, "A high performance privacy-oriented location system," in *Proc. of IEEE PerCom*, 2003.

[159] A. K. L. Miu, "Design and implementation of an indoor mobile navigation system," Ph.D. dissertation, Massachusetts Institute of Technology, 2002.

[160] "Smartphone market shares," https://www.idc.com/promo/smartphone-market-share/vendor.

[161] A. Harter, A. Hopper, P. Steggles, A. Ward, and P. Webster, "The anatomy of a context-aware application," *Wireless Networks*, vol. 8, no. 2/3, pp. 187–197, 2002.

[162] V. Filonenko, C. Cullen, and J. Carswell, "Investigating ultrasonic positioning on mobile phones," in *Proc. of IEEE IPIN*, 2010.

[163] M. Rossi, J. Seiter, O. Amft, S. Buchmeier, and G. Tröster, "Roomsense: an indoor positioning system for smartphones using active sound probing," in *Proc. of ACM AH*, 2013.

[164] K. Lorincz and M. Welsh, "Motetrack: A robust, decentralized approach to rf-based location tracking," in *International Symposium on Location-and Context-Awareness*. Springer, 2005, pp. 63–82.

[165] H. Lu, W. Pan, N. D. Lane, T. Choudhury, and A. T. Campbell, "Soundsense: scalable sound sensing for people-centric applications on mobile phones," in *Proc. of ACM MobiSys*, 2009.

[166] Y.-C. Tung and K. G. Shin, "Echotag: Accurate infrastructure-free indoor location tagging with smartphones," in *Proc. of ACM MobiCom*, 2015.

[167] A. A. Panchpor, S. Shue, and J. M. Conrad, "A survey of methods for mobile robot localization and mapping in dynamic indoor environments," in *Proc. of IEEE SPACES*, 2018.

[168] M. Jeub, C. Herglotz, C. Nelke, C. Beaugeant, and P. Vary, "Noise reduction for dual-microphone mobile phones exploiting power level differences," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 1693–1696.

[169] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Transactions on speech and audio processing*, vol. 9, no. 5, pp. 504–512, 2001.

[170] S. W. Smith *et al.*, "The scientist and engineer's guide to digital signal processing," 1997.

[171] Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Transactions on audio, speech, and language processing*, vol. 16, no. 1, pp. 229–238, 2007.

[172] W. Wang, V. Srinivasan, B. Wang, and K.-C. Chua, "Coverage for target localization in wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 2, pp. 667–676, 2008.

[173] P. Li, L. Sun, Q. Fang, J. Xie, W. Yang, and K. Ma, "An study of indoor localization algorithm based on imperfect signal coverage in wireless networks," in *International Conference in Swarm Intelligence*, 2013, pp. 325–333. [Online]. Available: https://academic.microsoft.com/paper/2280349607

[174] C. Wu, Z. Yang, Y. Liu, and W. Xi, "Will: Wireless indoor localization without site survey," in *2012 Proceedings IEEE INFOCOM*, 2012, pp. 64–72. [Online]. Available: https://academic.microsoft.com/paper/2077287651

[175] Q. Chen, B. Wang, X. Deng, Y. Mo, and L. T. Yang, "Placement of access points for indoor wireless coverage and fingerprint-based localization," in *International Conference on High Performance Computing*

and Communications, 2013, pp. 2253–2257. [Online]. Available: https://academic.microsoft.com/paper/2102739771

[176] H. Rizk and M. Youssef, "Increasing coverage of indoor localization systems for eee112 support," *arXiv preprint arXiv:1902.01707*, 2019. [Online]. Available: https://academic.microsoft.com/paper/2914278951

[177] G. L. Scalia, G. Aiello, R. Micale, and M. Enea, "Coverage analysis of rfid indoor localization system for refrigerated warehouses based on 2d-ray tracing," *International Journal of Rf Technologies: Research and Applications*, vol. 3, no. 2, pp. 85–99, 2012. [Online]. Available: https://academic.microsoft.com/paper/1663503187

[178] M. Gai, A. Azadmanesh, and A. Rezaeian, "A hybrid approach to indoor sensor area localization and coverage," *Journal of Networks*, vol. 10, no. 4, pp. 209–221, 2015. [Online]. Available: https://academic.microsoft.com/paper/2056880622

[179] K.-H. Lam, C.-C. Cheung, and W.-C. Lee, "Lora-based localization systems for noisy outdoor environment," in *Proc. of IEEE WiMob*. IEEE, 2017, pp. 278–284.

[180] J. Xiong and K. Jamieson, "Arraytrack: A fine-grained indoor location system," in *Proc. of USENIX NSDI*, 2013, pp. 71–84.

[181] Y. Xie, Y. Zhang, J. C. Liando, and M. Li, "Swan: Stitched wi-fi antennas," in *Proc. of ACM MobiCom*, 2018.

[182] C. Shepard, H. Yu, N. Anand, E. Li, T. Marzetta, R. Yang, and L. Zhong, "Argos: Practical many-antenna base stations," in *Proc. of ACM MobiCom*. ACM, 2012, pp. 53–64.

[183] S. Kumar, S. Gil, D. Katabi, and D. Rus, "Accurate indoor localization with zero start-up cost," in *Proc. of ACM MOBICOM*, 2014.

[184] S. Sen, J. Lee, K.-H. Kim, and P. Congdon, "Avoiding multipath to revive inbuilding wifi localization," in *Proc. of ACM MobiSys*, 2013, pp. 249–262.

[185] K. R. Joshi, S. S. Hong, and S. Katti, "Pinpoint: Localizing interfering radios." in *Proc. of USENIX NSDI*, 2013, pp. 241–253.

[186] D. Niculescu and B. Nath, "Vor base stations for indoor 802.11 positioning," in *Proc. of ACM MobiCom*, 2004.

[187] F. Adib, Z. Kabelac, D. Katabi, and R. Miller, "Witrack: motion tracking via radio reflections off the body," in *Proc. of USENIX NSDI*, 2014.

[188] H. Meyr, M. Moeneclaey, and S. Fechtel, *Digital communication receivers: synchronization, channel estimation, and signal processing*. John Wiley & Sons, Inc., 1997.

[189] J. Xiong, K. Sundaresan, and K. Jamieson, "Tonetrack: Leveraging frequency-agile radios for time-based indoor wireless localization," in *Proc. of ACM MobiCom*, 2015, pp. 537–549.

[190] S. Kumar, E. Hamed, D. Katabi, and L. Erran Li, "Lte radio analytics made easy and accessible," *ACM SIGCOMM Computer Communication Review*, vol. 44, no. 4, pp. 211–222, 2014.

[191] S. S. Hong and S. R. Katti, "Dof: a local wireless information plane," in *Proc. of ACM SIGCOMM*, 2011, pp. 230–241.

[192] T.-J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 4, pp. 806–811, 1985.

[193] Y. Qi, S. Zhang, L. Qin, H. Yao, Q. Huang, J. Lim, and M.-H. Yang, "Hedged deep tracking," in *Proc. of IEEE CVPR*, 2016, pp. 4303–4311.

[194] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. of the IEEE CVPR*, 2016, pp. 770–778.

[195] "OptiTrack," https://optitrack.com/, 2020.

[196] "RF Switch," https://www.infineon.com/cms/cn/product/rf-wireless-control/rf-switches-spxt-dpxt/bgs18ga14/, 2016.

[197] "USRP 2950," http://www.ni.com/en-us/support/model.usrp-2950.html, 2018.

[198] "RFID Tag," https://support.impinj.com/hc/en-us/articles/202756908-Monza-4-RFID-Tag-Chip-Datasheet, 2019.

[199] "Zigbee node," https://www.digi.com/products/embedded-systems/rf-modules/sub-1-ghz-modules/xbee-pro-900hp, 2016.

[200] "LoRa," https://www.dragino.com/products/module/item/102-lora-shield.html, 2019.

[201] "NB-IOT," https://wiki.dfrobot.com/SIM7000_Arduino_NB-IoT_LTE_GPRS_Expansion_Shield_SKU___DFR0505_DFR0572, 2017.

[202] "Sigfox," https://pycom.io/product/sipy/, 2018.

[203] "Source Code," https://github.com/Anplus/iArk, 2020.

[204] Y. Xie, J. Xiong, M. Li, and K. Jamieson, "md-track: Leveraging multi-dimensionality in passive indoor wi-fi tracking," *arXiv preprint arXiv:1812.03103*, 2018.

[205] "Impinj xarray," https://www.impinj.com/platform/connectivity/xarray/, 2019.

[206] Z. Luo, Q. Zhang, Y. Ma, M. Singh, and F. Adib, "3d backscatter localization for fine-grained robotics," in *Proc. of USENIX NSDI*, 2019, pp. 765–782.

[207] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller, "3d tracking via body radio reflections," in *Proc. of USENIX NSDI*, vol. 14, 2013.

[208] F. Adib, Z. Kabelac, and D. Katabi, "Multi-person motion tracking via rf body reflections," in *Proc. of USENIX NSDI*, 2015.

[209] M. Zhao, Y. Tian, H. Zhao, M. A. Alsheikh, T. Li, R. Hristov, Z. Kabelac, D. Katabi, and A. Torralba, "Rf-based 3d skeletons," in *Proc. of ACM SIGCOMM*, 2018, pp. 267–281.

[210] M. Zhao, T. Li, M. Abu Alsheikh, Y. Tian, H. Zhao, A. Torralba, and D. Katabi, "Through-wall human pose estimation using radio signals," in *Proc. of IEEE CVPR*, 2018, pp. 7356–7365.

[211] Y. Ma and E. C. Kan, "Accurate indoor ranging by broadband harmonic generation in passive nltl backscatter tags," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 5, pp. 1249–1261, 2014.

[212] A. Haniz, G. K. Tran, K. Saito, K. Sakaguchi, J.-i. Takada, D. Hayashi, T. Yamaguchi, and S. Arata, "A novel phase-difference fingerprinting technique for localization of unknown emitters," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 9, pp. 8445–8457, 2017.

[213] M. Youssef and A. Agrawala, "The horus wlan location determination system," in *Proc. of ACM MobiSys*, 2005, pp. 205–218.

[214] S. Sen, B. Radunovic, R. R. Choudhury, and T. Minka, "You are facing the mona lisa: Spot localization using phy layer information," in *Proc. of ACM MobiSys*, 2012, pp. 183–196.

[215] Z. Yang, C. Wu, and Y. Liu, "Locating in fingerprint space: wireless indoor localization with little human intervention," in *Proc. of ACM MobiCom*, 2012, pp. 269–280.

[216] H. Liu, Y. Gan, J. Yang, S. Sidhom, Y. Wang, Y. Chen, and F. Ye, "Push the limit of wifi based localization for smartphones," in *Proc. of ACM MobiCom*, 2012, pp. 305–316.

[217] H. Wang, S. Sen, A. Elgohary, M. Farid, M. Youssef, and R. R. Choudhury, "No need to war-drive: Unsupervised indoor localization," in *Proc. of ACM MobiSys*, 2012, pp. 197–210.

[218] L. Ni, Y. Liu, Y. Lau, and A. Patil, "Landmarc: Indoor location sensing using active rfid," *Wireless networks*, 2004.

[219] S. J. Pan, V. W. Zheng, Q. Yang, and D. H. Hu, "Transfer learning for wifi-based indoor localization," in *Proc. of ACM AAAI workshop*, vol. 6, 2008.

[220] J. Gjengset, J. Xiong, G. McPhillips, and K. Jamieson, "Phaser: Enabling phased array signal processing on commodity wifi access points," in *Proc. of ACM MobiCom*, 2014, pp. 153–164.

[221] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.

[222] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Advances in neural information processing systems*, 2014, pp. 3320–3328.

[223] Y. Li, Z. Chi, X. Liu, and T. Zhu, "Passive-zigbee: enabling zigbee communication in iot networks with 1000x+ less power consumption," in *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems*, 2018, pp. 159–171.

[224] Y. Peng, L. Shangguan, Y. Hu, Y. Qian, X. Lin, X. Chen, D. Fang, and K. Jamieson, "Plora: A passive long-range data network from ambient lora transmissions," in *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*, 2018, pp. 147–160.

[225] M. Hessar, A. Najafi, V. Iyer, and S. Gollakota, "Tinysdr: Low-power {SDR} platform for over-the-air programmable iot testbeds," in *17th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 20)*, 2020, pp. 1031–1046.

[226] L. Yang, Q. Lin, C. Duan, and Z. An, "Analog on-tag hashing: Towards selective reading as hash primitives in gen2 rfid systems," in *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, 2017, pp. 301–314.

[227] Q. Lin, L. Yang, C. Duan, and Z. An, "Tash: Toward selective reading as hash primitives for gen2 rfids," *IEEE/ACM Transactions on Networking*, vol. 27, no. 2, pp. 819–834, 2019.

[228] X. Chen, J. Liu, X. Wang, H. Liu, D. Jiang, and L. Chen, "Eingerprint: Robust energy-related fingerprinting for passive {RFID} tags," in *17th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 20)*, 2020, pp. 1101–1113.

[229] D. Zanetti, B. Danev, and S. Capkun, "Physical-layer identification of uhf rfid tags," in *Proceedings of the sixteenth annual international conference on Mobile computing and networking*, 2010, pp. 353–364.