Title
Embracing Mmwave for XG in the Sky: A Cross-Domain Approach

Author
Zhang, Zhehui

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Embracing Mmwave for XG in the Sky:
A Cross-Domain Approach

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Computer Science

by

Zhehui Zhang

2022

ABSTRACT OF THE DISSERTATION

Embracing Mmwave for XG in the Sky:
A Cross-Domain Approach

by

Zhehui Zhang
Doctor of Philosophy in Computer Science
University of California, Los Angeles, 2022
Professor Songwu Lu, Chair

Drone-based applications are increasingly prevalent in surveillance, delivery, automation, and many more usage scenarios. They all require seamless mobility support and high-throughput data access. The mmWave communication in current cellular networks offers the physical-layer technology candidate, but cannot deliver the required services. The root cause lies in the fundamental limit of the legacy time-frequency domain design, where time-frequency-based multiplexing amplifies Doppler effects and interference in the drone setting, thus incurring packet losses and connectivity failures.

This dissertation tackles the problem from a novel cross-domain approach to embracing mmWave for 6G and beyond cellular networks. The cross-domain design synthesizes the delay-Doppler domain and time-frequency domain. The insight is that we may achieve the best of both worlds if properly multiplexing signaling in both domains. Along this new direction, we seek to achieve three goals: high-fidelity channel estimation under aerial mobility, effective interference cancellation without pre-transmission negotiation, and readily-deployable practical solutions in 6G and beyond.

Our contributions can be summarized as follows: 1) We design a novel concurrent channel state inference algorithm for drones by modeling the shared multipathing features between cells. Our key insight is that the delay-Doppler domain reveals frequency-decoupled physical path features. We devise the delay-Doppler signaling that is embedded into the legacy OFDM signals to enable Doppler-independent inference. Our design characterizes and decouples path delay, and infers multipath propagation. 2) We design a novel cross-domain interference cancellation scheme by multiplexing signals and data across neighboring cells. Unlike prior schemes that require pre-transmission negotiation of power and data rate, we transform transmission-induced interference between cells into meaningful data, boosting throughput among dense, interfering cells in the sky. Our solution is resilient to multiple interferer scenarios under dense cell deployment. 3) We integrate cross-domain signals in 6G and beyond with practical system designs. Our design has one key objective: retain the current data transmission pipeline within 3GPP standardized framework to maximize the deployability. We devise an overlay to embed new delay-Doppler signals into the current physical layer in the time-frequency domain. We then propose adaptive scheduling so that cross-domain signals can be readily integrated into current standards. We assess the feasibility of our approaches by prototyping and evaluating with an mmWave testbed.

The dissertation of Zhehui Zhang is approved.

Yuval Tamir

George Varghese

Lixia Zhang

Songwu Lu, Committee Chair

University of California, Los Angeles

2022

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am highly indebted to my academic advisor Prof. Songwu Lu for his extensive support, and for guiding me through the journey of Ph.D. training. I am fortunate to have him as a supervisor, who has such great research taste and cares about his students' success. He is sharp and critical when critiquing my research, setting a high benchmark for the research. At the same time, He has always encouraged me to go the extra mile, provided valuable insights, and cultivated inspiring discussions when I get stuck. Thanks to him, I attained unparalleled intellectual growth and mental maturity through the ups and downs.

I am also grateful to my dissertation committee at UCLA: Prof. Lixia Zhang, Prof. George Varghese, and Prof. Yuval Tamir, for their invaluable feedback and insightful suggestions that helped shape this dissertation. My dissertation would not be completed without their help and support. I will always be grateful for their guidance and support.

I also want to thank my collaborators for their valuable support. I learned a lot from Prof. Yuanjie Li, Prof. Lili Qiu, and Prof. Chunyi Peng. I had a great time during the summer interns. I would like to thank Dr. Xenofon Foukas, Dr. Bozidar Radunovic, Dr. Guohui Wang, Dr. Shu Shi, Dr. Varun Gupta, and Dr. Rittwik Jana for their help. I also want to thank my collaborators on RePaC, Haiyang Zheng, Xiangning Yu, Jiayao Hu, for their efforts in pushing the paper into ATC. I am thankful for that.

I also owe gratitude to the very exciting research group that I have been fortunate to be a part of. It was my honor and fortune to be surrounded by my genius peers: Dr. Qianru Li, Dr. Zhaowei Tan, Yunqi Guo, Jinghao Zhao, Boyan Ding, and Yifei Xu. I enjoyed those valuable moments with them, from discussing research problems to having fun together. All of them are super supportive and always bring new ideas during our discussion. It's a great pleasure to work with such an amazing group of talented people.

As I complete my Ph.D. journey, when I look back, I cannot help but to thank the people who supported me in getting into UCLA. I am very grateful to Prof. Xinbing Wang and

Prof. Xiaohua Tian at SJTU, who intrigued my interests in system research, mentored my undergraduate research thesis, helped me write my first paper, and backed me up when I was a research novice. I also owe deep gratitude to all the teachers who taught me, believed in me, and prepared me for the next steps in my journey.

Outside of work, I am blessed to have a fantastic set of friends. I will not speak out their names but the gratitude goes deeply in my heart. Many friends provided warmth and companionship even though they were located far away. I will forever cherish the fun we have when we are together. Thanks for their friendship, warmth, and support.

Finally, and more importantly, I can't find words to describe my gratitude to my family. My parents, Liying and Jinbang, taught me the values of confidence, dedication, and positivity. They are the reasons behind my accomplishments. My boyfriend, for the happiness and pleasure that he brought into my life. Imaging my life without him seems impossible now. My furry friend, Gogo, for comforting me with his fluffy tail. This thesis would not have been possible without them.

VITA

2012–2016	B.S., Computer Science and Technology, SJTU, Shanghai, China
2016–2022	Graduate Student Researcher, Computer Science, UCLA.
2017–2022	Teaching Assistant, Computer Science, UCLA.
2016 spring	Software Engineer Intern, Intel, Shanghai, China
2017 summer	Visiting Scholar, Ohio State University, Columbus, OH
2018 summer	Student Technician, AT & T Labs, Bedminster, NJ
2019 summer	Research Intern, Alibaba Group, Sunnyvale, CA
2021 summer	Research Intern, Microsoft Azure for Operators, Remote, CA
2021 & 2022 summer	Program Coordinator of Summer Institute (CSSI), Computer Science, UCLA

PUBLICATIONS

Zhehui Zhang, Yuanjie Li, Qianru Li, Jinghao Zhao, Ghufran Baig, Lili Qiu, Songwu Lu. Movement-Based Reliable Mobility Management for Beyond 5G Cellular Networks, IEEE/ACM Transactions on Networking (TON) 2022.

Qianru Li*, **Zhehui Zhang*** (co-primary), Yanbing Liu, Zhaowei Tan, Chunyi Peng, Songwu

Lu. "CA++: Enhancing Carrier Aggregation Beyond 5G", Under review, ACM MobiCom 2023.

Jinghao Zhao, Zhaowei Tan, Yifei Xu, **Zhehui Zhang**, Songwu Lu. SEED: A SIM-Based Solution to 5G Failures, ACM SIGCOMM 2022.

Zhehui Zhang, Haiyang Zheng, Jiayao Hu, Xiangning Yu, Chenchen Qi, Xuemei Shi, Guohui Wang. "Hashing Linearity Enables Relative Path Control in Data Centers", USENIX ATC 2021.

Ghufran Baig, Changhan Ge, Lili Qiu, Yuanjie Li, Wangyang Li, Wei Sun, Jian He, **Zhehui Zhang**, Songwu Lu, "Extracting and Predicting Multipath Profiles under High Mobility", ACM MobiHoc 2022

Yuanjie Li, Chunyi Peng, **Zhehui Zhang**, Zhaowei Tan, Haotian Deng, Jinghao Zhao, Qianru Li, Yunqi Guo, Kai Ling, Boyan Ding, Hewu Li, Songwu Lu. "Experience: A Five-Year Retrospective of MobileInsight", ACM MobiCom 2021.

Yuanjie Li, Qianru Li, **Zhehui Zhang**, Ghufran Baig, Lili Qiu, Songwu Lu. "Beyond 5G: Reliable Extreme Mobility Management", ACM SIGCOMM 2020.

Zhehui Zhang, Shu Shi, Varun Gupta, Rittwik Jana. "Analysis of Cellular Network Latency for Edge-Based Remote Rendering Streaming Applications", ACM SIGCOMM 2019 Workshop on NEAT.

Zhaowei Tan, Yuanjie Li, Qianru Li, **Zhehui Zhang**, Zhehan Li, Songwu Lu. "Enabling Mobile VR in LTE Networks: How Close Are We?", ACM SIGMETRICS 2018.

CHAPTER 1

Introduction

Drones are increasingly valuable in surveillance, delivery, and many more emerging applications [24]. The global commercial drone market was valued at 2.72 billion US dollars in 2020, and is projected to reach 21.69 billion by 2030 [35]. At the same time, drones pose extra challenges to the supporting network. These critical applications need seamless network support to deliver high-throughput and real-time data in the sky. As required by always-on surveillance or automated delivery, the network should always be available on demand to deliver high-throughput traffic and offer low end-to-end latency.

The only promising network to enable drone-based applications is cellular networks, which provide anywhere, anytime access to ubiquitous devices. With the advances of 5G mmWave technology, above-24GHz frequencies provide multi-Gbps data rates with nearly 100x greater channel bandwidths than what is available in the sub-6 GHz [25]. According to recent studies, the ground devices witness boosted data rate of close to 3Gbps with the support of mmWave technology [92]. However, it is not the case for drones. Although mmWaves communication provides the physical-layer technology candidate with a wide channel bandwidth, it cannot deliver the required services for highly mobile drones in the sky due to several challenges.

This dissertation tackles the challenges of embracing mmWaves for 6G and beyond from a novel cross-domain approach. We next present the challenges of supporting highly-mobile drones with mmWave bands in §1.1. We then detail our contributions in addressing the challenges with innovative cross-domain designs in §1.2. Many application scenarios can benefit from our designs. We introduce the application scenarios in §1.3. Finally, we present

the organization of this dissertation in §1.4.

1.1 Challenges

There are many challenges to supporting networked drones due to unique propagation features in the sky, especially with the emerging mmWave radio. While mmWave technology successfully boosts data for ground devices under low mobility, it is vulnerable to dramatic wireless dynamics from high mobility and complicated propagation in the sky. These challenges induce incorrect channel estimation and interference under mobility, resulting in high packet loss and failures. Here we summarize the common challenges in two use scenarios: channel estimation and interference management.

The first and foremost challenge is to tackle the time-frequency coupling under aerial mobility for channel estimation. Current cellular networks and recently proposed approaches [120, 29, 137] are all based on the time-frequency domain. In the time domain or frequency domain, all paths are mixed, and the resulting phase and amplitude change with the delay, Doppler, and attenuation of these paths. While existing approaches can formulate channel estimation as a non-convex optimization problem with a simplified model, such approaches can only find the multipath parameters that best fit the falsely measured channel response. Besides, due to the strong inter-path dependency and the non-convex objective, the problem is fundamentally challenging.

The second challenge is the interference in the sky. Signals are susceptible to high fading and frequency coupling in the sky, which causes inaccurate channel state inference. Propagation in the sky brings more interferences, which deteriorates data speed. Although the operator attempts to relieve the problem by deploying more cells, overlapping cells only worsens the problem. Interference-free transmission is impossible in the sky. The interference amplifies inaccurate cell quality indicators and triggers poor cell selection.

The final challenge is to integrate the designs in the 6G and beyond cellular networks.

While wireless access technologies are advanced with a wide spectrum, higher modulation, etc., the system infrastructure remains unchanged. The current 5G system has designed a complicated protocol stack to provide high performance for common application scenarios. To support more challenging application scenarios, current 5G networks need to incorporate new radio designs. However, replacing the protocol stack would incur prohibitive costs. We desire a standard-compliant approach to enable fast deployment of our designs.

1.2 Our Contributions

In this dissertation, we focus on two goals that are critical to high-quality service.

• For mmWave communication from highly-mobile drones, *how to perform accurate and efficient channel estimation under wide spectrum?*

• With multiple interfering cells in the proximity of the drones, *how can the drones cancel interferences for aerial transmission?*

This dissertation takes a novel cross-domain approach to embrace mmWave for drones on both goals. The cross-domain design synthesizes the delay-Doppler domain and time-frequency domain. The insight is that we may achieve the best of both worlds if properly multiplexing signaling in both domains. We address the issues of cross-domain multiplexing with innovative transformation algorithms and practical system designs.

With our design, the system can utilize the delay-Doppler domain for explicit Doppler modeling and other physical path features while keeping the time-Frequency domain for data plane efficiency and deployability. Such a cross-domain approach also enables natural interference separation between two domains. Along this new direction, we seek to achieve three goals: high-fidelity channel estimation under aerial mobility, effective interference cancellation without pre-transmission negotiation, and readily-deployable practical solutions in 6G and beyond.

1.2.1 Concurrent Channel Estimation for MmWaves

We first present our design CCE on high-fidelity channel state inference for drones under mmWaves. The key innovation is the concurrent channel estimation algorithm. The algorithm operates in two domains, the legacy time-frequency domain and the delay-Doppler domain. We find that the delay-Doppler representation projects signals onto the 2D delay-Doppler plane to separate out paths different in either delay or Doppler. Therefore, only the paths that have a similar delay and similar Doppler can interfere, which are much fewer in number. The algorithm extracts common features for cells that are located at different cell towers in the spatial proximity of the mobile device. Therefore a single-cell measurement from a given tower can reveal the channels for all other colocated cells.

Our channel inference algorithm exploits the benefits of both domains. The delay-Doppler domain model makes inference accurate under mobility, since the model uses Doppler shifts that are not explicitly tailored to time. The delay-Doppler path features separate path delays from frequency domain coupling. By characterizing path delays, we can decouple multiple path propagation with adaptive scheduling of delay-Doppler signals. High-fidelity channel inference is further made possible over a wide spectrum by embracing fractional Doppler shifts. With this algorithm, we can infer all cells residing on a tower via a single-cell measurement.

We implement CCE on an SDR-based 5G NR testbed, and evaluate it with real-world experiments and trace-driven emulations. We assess the concurrent channel inference algorithm under a widespan of radio spectrum (2.45GHz - 5.55GHz for sub-6G channels and 58GHz - 62GHz over mmWave bands). The empirical validation shows that, CCE achieves an SNR inference error of 0.37 dB to 1.03 dB at low mobility ($\sim$10 km/h) and 0.39 dB to 0.79 dB at high mobility (>200 km/h). CCE outperforms all state-of-the-art schemes by reducing the error by more than 63%.

1.2.2 Embracing Interference with X-Domain Cancellation

We then show a novel cross (X)-domain interference cancellation scheme. We propose CODIM, which takes a new perspective on interference cancellation and overcomes the limitations of legacy time-frequency domain-based solutions. To handle mmWave scenarios, we address the challenges of intra-domain interference by strategically interleaving interfering signals across two domains. CODIM works well under both low/moderate and high mobility. It remains effective over a wide span of the radio spectrum.

The key insight is to multiplex signal and data at both domains to cancel interference. We discover the natural interference separation between two domains, which enables interference estimation with cross-domain transformation. Unlike previous interference cancellation schemes that require pre-transmission negotiation, we avoid heavy coordination between all interfering cells. We further devise a cooperative decoding algorithm to utilize the high-speed backhaul network between cells. CODIM transforms transmission-induced interference between cells into meaningful data, boosting throughput among dense, interfering cells in the sky.

We prototype CODIM in commodity software-defined radio and evaluate it with both testbed-based experiments and trace-driven emulation. Compared to solutions today, CODIM reduces packet losses by 63.2% on average. Even in high mobility, CODIM achieves comparable failure ratios to static and low mobility scenarios. We also evaluate CODIM's benefits for real-world applications, including mobile VR/AR and data transfer. Our experiments show that CODIM reduces disruption for mobile VR applications by 47.1%, improves object recognition accuracy by 88.3% for mobile AR, and reduces stalls by 31.8%-46.2%. Meanwhile, CODIM retains marginal overhead of signaling traffic and latency without hurting data transfer.

1.2.3 Integrating X-Domain Designs in 6G and Beyond

Finally, we enable cross-domain designs within 5G PHY to enable integration in XG. Although the cross-domain designs are attractive, it remains an open question on how to integrate delay-Doppler signals with current OFDM-based PHY. We make three contributions. We first devise an overlay to multiplex the OTFS-based reference signals in the standardized OFDM PHY. With the dual-domain PHY, we propose adaptive scheduling, such that cross-domain signals can be readily integrated into current 3GPP standards for PHY-above radio layers. Other than the radio layers, we leverage a novel transform and inter-cell collaboration to profile interference for control layers. Our proposed cell selection metrics solve the issue of current OFDM-based cell quality indicators by capturing the impact of interference at both uplink and downlink. We assess the feasibility of our approaches by prototyping and evaluating them with an mmWave testbed.

We then extend our testbed into an open-radio platform, `NEMONET`, enabling researchers to perform the trace-driven evaluation of new radio designs. The platform provides mmWave support and offers an end-to-end evaluation pipeline. While the existing mmWave testbed has been successful in supporting primitive data transmission, they are solely based on time-frequency signals. `NEMONET` enables open research on more innovations in the synergy of delay-Doppler domain and time-frequency domain.

1.3 Applications scenarios

The proposed fundamental technology achieves both wireless efficiency and mobility reliability, enabling diverse application scenarios, including the following:

Drone applications Drones are increasingly valuable in surveillance, delivery, and many more emerging applications [24]. These applications all require *always-on* network to enable real-time control over the drones. For example, for surveillance, drones are required to

monitor the real-time video of the site following the user's commands. These applications typically handle various environments, even extreme propagation conditions. Our proposed technologies facilitate communication between drones.

In-the-sky networks The drones are not only clients, but can be a part of network infrastructure. Recent work has been discussing the potential of drone-based mmWave networks [74], which facilitate line-of-sight group transmission with mmWaves. However, the instability of wireless remains the main bottleneck to applying mmWaves as the transmission medium. This dissertation can provide the technology for fundamental data transmission and interference mitigation.

Extreme mobility With the advances in transport technology, clients are moving at a faster speed. High mobility use scenarios become a norm rather than an exception, such as the high-speed rails (with speed up to 350 km/h [129]), vehicle-to-everything (e.g., autonomous driving and fleet management, with 6 million vehicles by 2022 [34]), drones, and many more. We propose advanced wireless channel estimation to enable reliable mobility management in this dissertation.

1.4 Organization

This dissertation is organized into the following chapters. §2 introduces the primer of cellular network on PHY basics and critical cell selection procedures. §3 studies the limitations of current channel estimation approaches and proposes two stage of solutions to enabled concurrent channel estimation with cross-domain channel transformation. §4 investigates how Doppler shifts under mmWave and mobility impede the effectiveness of cell selection. Then a cross-domain-based interference management scheme is presented. §5 solves the critical issues to enable cross-domain designs in current standardized framework. We propose to embed OTFS signals into OFDM PHY in a standard-compliant fashion. Finally, we summarize our works and insights, and discuss future works in §6.

CHAPTER 2

State of the Art and Overview

We have witnessed major technical evolutions in cellular networks in the past decades. While wireless access technologies are advanced with a wide spectrum, higher modulation, etc., the system infrastructure remains unchanged. The misalignment between two key technologies from the wireless and the infrastructure, mmWave and mobility management, induces bad performance and even failures. This dissertation aims to achieve the synergy between efficient wireless and reliable mobility.

This chapter discusses the state-of-arts and provides an overview of how our designs complement the gap between wireless designs and mobility management technologies. We introduce an overview of current 5G/6G architecture and fundamental technologies in §2.1. In §2.3 we discuss the existing efforts to improve the synergy between wireless and mobility. We give an overview of our designs in §2.4.

2.1 5G/6G Architecture

In cellular networks, devices connect to the network via local access stations and remote core networks, forming a hierarchical mobility management model. Figure2.1 illustrates an overview of the cellular network. In general, there are three parts: clients, network infrastructure (including the access network and core network), and the wireless channel in between. To provide "anywhere, anytime" network access, cellular networks deploy base stations at a large scale. Each base station may manage multiple cells (antennas), each of

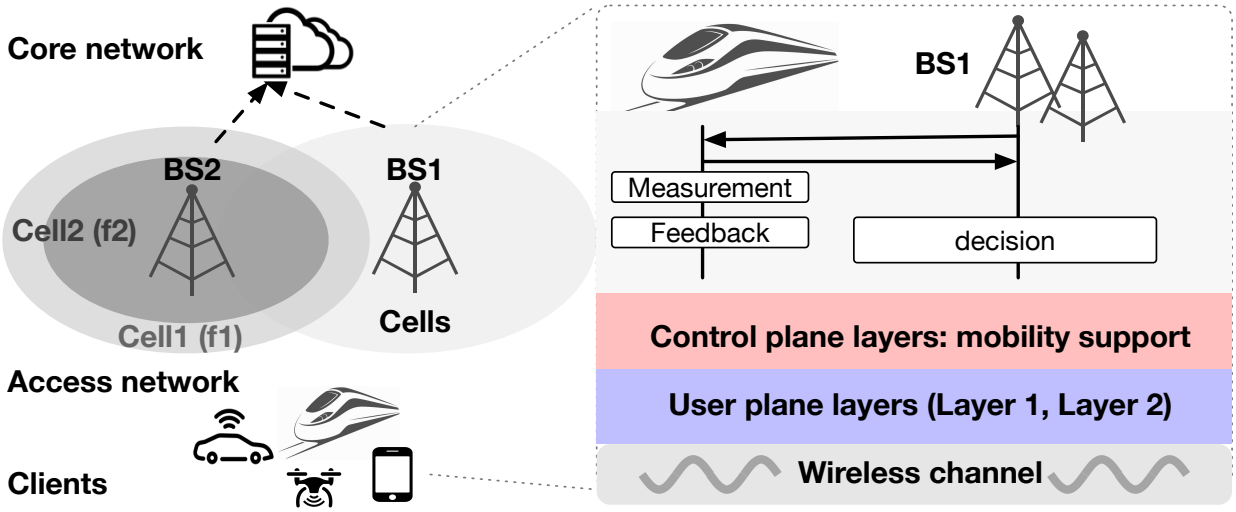


Figure 2.1: Illustration of 5G architecture.

which runs at various frequency bands with different coverage. We focus on wireless access and the infrastructure in this dissertation. We next describe the state-of-the-art and recent advances.

The wireless provides last-hop radio access to the clients. The data rate that the device can achieve is mainly decided by the wireless link. Although the spectrum was the main bottleneck before 5G, recent advances in mmWave technology opened a wide radio spectrum for high-speed transmission. However, more radio spectrum does not mean a higher quality of service. The wireless channel is dynamic and tends to diminish the gain of a wide spectrum. In this dissertation, we discuss what innovation is needed to utilize the wide radio spectrum robustly and efficiently.

The infrastructure bridges the wireless and the internet, providing seamless service to devices. 3GPP standards regulate the user plane layers to deliver data and the control plane layers to manage radio resources and mobility. Given the major scenarios in cellular networks involve mobility, mobility support is paramount to user experiences. In this dissertation, we revisit the legacy mobility management scheme and its deficiency in providing reliable services under new wireless technologies.

2.2 Key Technologies

In this section, we introduce the key technologies in the wireless link and the infrastructure. Wireless access technologies evolve to enable a wider radio spectrum and improve access efficiency. We focus on the key enabler of 5G, mmWave technology in §2.2.1. On the network side, the key technology to enable seamless service is mobility management. We summarize the current 5G mobility support in use and recent advances in §2.2.2.

2.2.1 MmWave communication

The mmWave communication is the key enabler for current 5G networks. The available radio spectrum has been greatly expanded with 30–300 GHz frequency bands. above-30GHz frequencies provide multi-Gbps data rates with nearly 100x greater channel bandwidths than what is available in the sub-6 GHz [25]. These frequency bands are highly available but under-utilized. According to recent studies, the ground devices witness boosted data rate of close to 3Gbps with the support of mmWave technology [92]. However, such supremacy to sub-6G bands diminishes upon mobility [92].

MmWave bands post high requirements on both channel estimation and interference management. Channel estimation under mmWave is more challenging due to the fast fading features of mmWave frequencies. As the center frequency increases, the time that a channel remains stable (i.e., coherence time) reduces, leaving less time for channel estimation and feedback. Interference management is the same challenging, if not more. Although existing approaches attempt to leverage the high-speed backhaul networks to improve the efficiency of interference cancellation, time-frequency domain multiplexing fails to handle Doppler. Doppler mixes data and interferences, making it hard to separate them in the time-frequency domain.

The main challenge is the applicability to mobility scenarios. Due to the ultra-short wavelength, MmWave communication is susceptible to propagation fading and blockage,

thus requiring highly directional links. However, the constant motion and relative position changes to the cells during mobility introduces challenges in maintaining the orientation of directional links. To this end, many recent works develop beamforming and nulling to mitigate the interference [53, 74, 120, 86, 136].

2.2.2 Mobility Management

Mobility management is provided by two critical procedures, handoff and carrier aggregation, in the current 5G. We next introduce how the current procedure works, challenges, and recent advances.

Mobility support centers on the network side control. At a given location, a mobile device accesses the network with one primary cell (aka primary component carrier). When the device moves, the base station migrates its service to a cell that the device is heading to. The migration procedure is called handoff. In order to boost throughput, the base station would enable one or more secondary cells to deliver data to the device, so-called carrier aggregation (CA).

Both handoff and carrier aggregation involve the selection of cells to serve the device. Mobility support replies on the handoff procedure, during which the device switches from the current primary cell to another. The primary cell is used as an anchor to manage the service and signaling between the device and the core network. As a device leaves the current primary cell's coverage, it will be migrated to another one. Such procedures follow the configurations and the handoff target decided by the base station. Based on patents [84, 99, 135] and manuals [40, 39, 41], the base station make decision based on device reports. Throughput boosting rely on another procedure called carrier aggregation, during which the device aggregates a few secondary cells at a different frequency from primary cells to boost throughput. Both procedures involve the decision of which primary cell to handoff to and which secondary cells to aggregate with.

The main challenge of mobility support is to decide which cells to connect to and when/where to switch cells. Several recent studies seek to improve cell selection using the standardized mechanisms in 4G/5G [73, 43, 64, 50, 79]. [73, 43] focus on better cell selections with either network or device-centric solutions. Other prior efforts [64, 50, 79] use resource scheduling or joint carrier selection to maximize utilization.

2.3 Limitations of the State of the Art

Recent advances push the limits of wireless efficiency and mobility reliability, respectively. Nevertheless, the limitation of the coupling between these two is rarely discussed. This dissertation tackles the disastrous coupling between wireless and mobility.

Inefficient wireless under mobility. The main limitation is the inefficiency under mobility. Mobility introduces Doppler shifts to signals, harming wireless efficiency. The key to high-speed wireless access is high-order channel coding, which needs to adapt to the dynamic channels in an ultra-low time span. Given that the minimal granularity to scheduling resources is sub-ms, the channel is expected to remain unchanged during this time slot, which is however unrealistic under mobility.

The key requirement for mmWave efficiency is the accuracy of channel state estimation. Without accurately synchronizing the channel states between the client and the network, the core functions cannot fully work to utilize the best of the wide spectrum, e.g., the precoding of data, beamforming management, and massive MIMO. This problem involves two sub-problems, to accurately estimate the channel state at one side, and to synchronize the state from one side to another. In this dissertation, we propose innovative channel estimation algorithms to facilitate mmWave communication.

Unreliable mobility under mmWaves. The problem of mobility lies in the wireless-strength-based design. Current mobility support centers on wireless-strength-based design. Thus wireless and mobility are deeply-coupled. As the device is moving at high speed, the

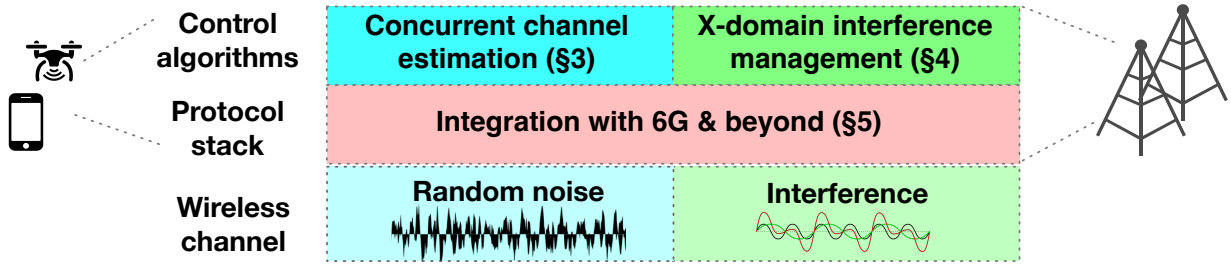


Figure 2.2: Research overview.

mobility management function relies on unreliable and inaccurate wireless metrics to perform cell selection, which results in bad performance and failures. Even when the device is moving at a low speed, the mobility reliability is affected by wireless interference.

The key requirement for reliable mobility is to decouple the instability of wireless with the mobility functions. What's more, the reliability should be guaranteed under various application scenarios, no matter under low or high mobility, and low or high radio frequency. In this dissertation, we take a new perspective. We unveil the fundamental limitations of wireless-centric mobility design. We explore a cross-domain approach to accommodate the expanding spectrum space (say, mmWave bands) and high mobility scenarios.

2.4 Research Statement and Overview

This dissertation aims to achieve both efficient wireless and reliable mobility support. We identify two key issues, channel estimation under mobile mmWave (under random noise) and interference under dense cell deployment, in supporting drones with mmWave technology in current cellular networks. We contribute with a novel cross-domain approach and .

We visualize the relationship between our designs in Figure 2.2. We propose two novel designs, concurrent channel estimation and interference management, to deal with frequency coupling. The first step is to combat the mobility and wide radio spectrum under random noise with *concurrent channel estimation*. The second step is to combat interferences, as well

as the same challenges as the random noise scenario, fulfilled with cross-domain interference management. With these two designs, we propose the readily deployable integration scheme to incorporate our designs with the current protocol stacks, enabling continuous deployment in 6G and beyond.

We propose three innovative systems, `CCE` for concurrent channel estimation under random noise, `CCE` for interference management under dense cell deployment, and `NEMONET` for practical integration in 6G and Beyond.

- `CCE`: Concurrent Channel Estimation for MmWaves (§3). We seek to parallelize channel measurements over the wide spectrum provided by mmWave bands. The goal is to enable efficient and high-fidelity channel estimation even under high mobility. The main idea is that we estimate signals in the delay-Doppler domain and transform them back to the conventional time-frequency domain.

- `CODIM`: Embracing Interference with X-Domain Cancellation (§4). We take a cross-domain approach to estimate and cancel interference to relieve drones from interfering transmission. Unlike the above system, the estimation is done under structural interferences rather than random noise. The key insight is that by properly aligning interferences at both domains, we can cancel interference from multiple interferers.

- `NEMONET`: Integrating X-Domain Designs in 6G and Beyond (§5). We propose practical system designs to integrate the cross-domain approaches within current 3GPP standards. We also design cross-domain channel metrics to enable interference-perceptible cell selection. The main innovation is the standard-compliant co-existence scheme of cross-domain signals.

CHAPTER 3

High-Fidelity Concurrent Channel Estimation for MmWaves

This chapter is organized as follows. §3.1 presents our motivation of supporting concurrent channel estimation and the challenges. After an overview in §3.2, we elaborate on the design of `CCE` 's core functions in two stages: single path (§3.3) and multi-path (§3.4). §3.5 presents our implementation and evaluation. We discuss the extension of this work in §3.6. Finally, we present related works on channel state estimation in §3.7.

3.1 Motivation

3.1.1 Primer and Issues

We first introduce the primer on 5G mobility management procedure.

As shown in Figure 3.1, the decision of primary cell and secondary cell for devices in the active state involves four steps: configure, monitor, report, and decide. The decision is controlled by the serving cell and assisted by the client. Different from active state devices, idle state devices make handoff decisions by themselves. The decision logic of CA management and active state handoff is not standardized. The network side measurement could also affect the decision, but it is not visible to that. Once the client connects to the base station, it starts to exchange signaling messages with the base station. In step *configure*, the base station configures the devices with a list of configurations to mandate

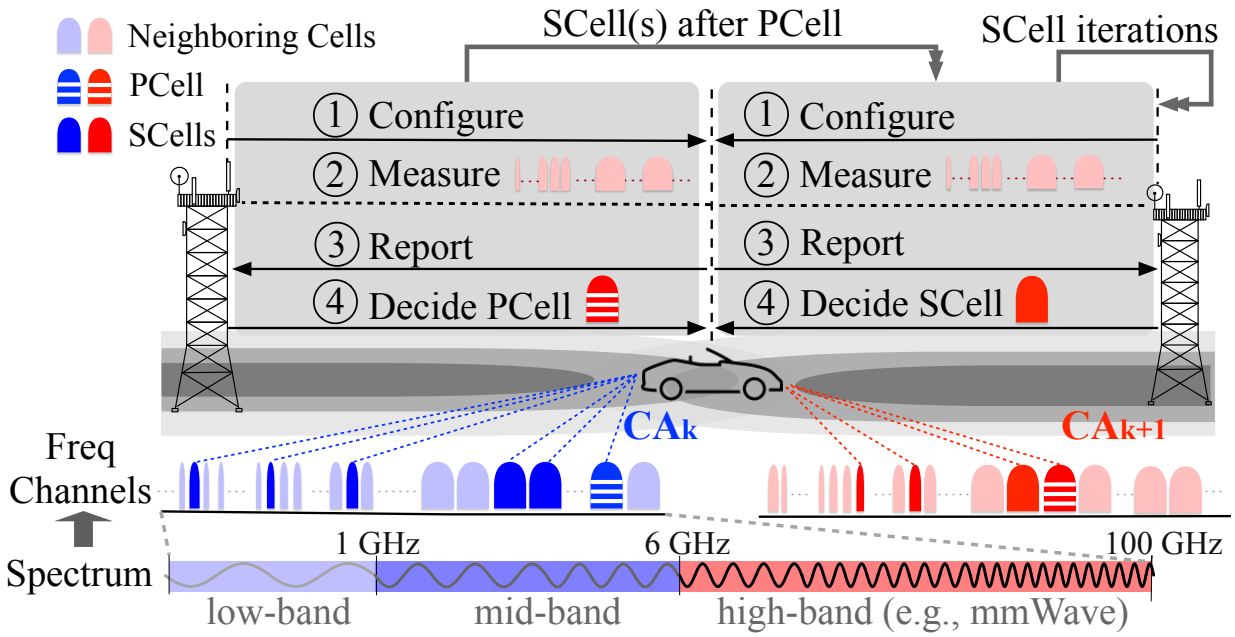


Figure 3.1: Illustration of handoff and carrier aggregation.

the monitoring and reporting procedure. The configurations ask the devices to monitor the serving/target cells' signal strengths and standard triggering criteria (Table 3.1). Then the device starts to monitor its connected cells and neighbor cells. This is the step *monitor* and reports the feedback if any criteria are satisfied. The step *report* happens when standard triggering conditions configured by the serving cell are satisfied. The device sends the report to the serving cell. Upon receiving the client's report that meets the criteria, the serving cell moves to the *decision* phase. It runs its local policy to decide if more feedback is needed (via reconfiguration), if handoff should start, and which cells to migrate to. In this process, it may reconfigure the device by adding or deleting the triggering criteria or updating the signal strength thresholds. If the serving cell decides on a handoff, it will send the handoff command to the device. Then the device would follow the handoff command to move to another cell.

5G standards [21, 16, 12] introduced many new technologies to improve reliability, such as new radio physical design, cloud-native deployment, etc. However, the mobility manage-

Table 3.1: Reporting triggering criterias. R stands for RSRP, RSRQ or SINR. O stands for cell offset configured for each cell [5]

Event	Criteria	Explanation
A1	$R_s > \Delta_{A1}$	Serving cell becomes better than a threshold
A2	$R_s < \Delta_{A2}$	Serving cell becomes worse than a threshold
A3	$R_n + O_n > R_s + O_s + \Delta_{A3}$	Neighbor cell becomes offset better than primary cell
A4	$R_n + O_n > \Delta_{A4}$	Neighbor cell becomes better than a threshold
A5	$R_s < \Delta^s_{A5}$ & $R_n + O_n > \Delta^n_{A5}$	Serving cell becomes worse than a threshold, and neighbor cell becomes better than a threshold
A6	$R_n + O_n > R_s + O_s + \Delta_{A3}$	Neighbor cell becomes offset better than secondary cell

ment mechanism stays the same. As the management is based on wireless signal strength, it still follows the three phases, triggering, decision, and execution. The feedback and signaling rely on OFDM-based transmission, which is unreliable for making handover decisions. The decision policy tightly couples with signaling reliability, motivating operators to adopt complex, even conflicting decision logic.

Challenges in the wireless channel. The wireless channel is dynamic by nature. The impact of mobility on higher-frequency channels (say, those mmWave ones) is more severe. Radio channels remain statistically coherent within the channel coherence time [119]: $T_c \propto \frac{1}{f_m} = \frac{c}{v \cdot f}$, where f_m is the maximum Doppler shift, decided by the speed of light c, velocity v, and frequency f. From 2.4 GHz to 39 GHz, higher-frequency channels vary 16× faster at the same velocity, thus making measurement less reliable. Faster measurements are needed to not miss good cells on higher frequency channels.

Challenges in cell selection. There are many challenges for the mobility management in 5G. First, the density of cells increases in 5G, which implies more frequent handovers. Especially in the non-standalone mode, each device is connected to a 4G cell and a 5G cell. Thus, the device will experience more cell switching. Recent research on 5G mobility [92] conforms to our statement that 5G handover is more frequent. Second, 5G adopts mmWave

bands (e.g., 29GHz), where the Doppler spread is more severe. The 5G also added new numerology with a shorter slot time [12], which is more susceptible to Doppler spread caused issues.

Another challenge is the complicated cell selection with carrier aggregation. Carrier aggregation (CA) aggregates multiple contiguous or non-contiguous frequency bands to boost user throughput. It is a technique critical to utilizing the expanded radio spectrum in the current 5G network and beyond [45, 25]. In 5G and beyond, the available radio spectrum is expanding with <6 GHz bands and mmWave bands (i.e., >24 GHz) [27]. Although current CA already boosted the data rate significantly (Figure 3.2), it remains an open question to fully utilize the potential of a wide spectrum.

Although the mobility mechanism does not change, one might wonder whether the infrastructure change in 5G brings more benefits to optimize mobility performance. The main infrastructure update from 4G to 5G is the adoption of cloud-native deployment. The control plane operations can be moved closer to the RAN to speed up the decision. However, triggering delay is the main bottleneck. The measurement and reporting are still constrained by the round-trip between the device and the base station. As the unstable OFDM-based modulation remains unchanged, the problems with signaling and policy remain.

3.1.2 Reality Check With 5G

We next use a real-world example to illustrate the above issues. In Trace 1, the mobile phone was looking for a new serving cell set to continue service. First, the phone was configured to search PCell candidates on four frequency channels. The phone performed sequential measurement for about 1.3s until finding a strong cell on frequency channel F9820. After receiving the report, the network immediately selected that cell to keep connectivity. Afterward, it took 3.5s to scan channels on 4G, 5G sub-6G and 5G mmWave bands; Finally, a 4G cell and a 5G sub-5G cell were added as SCells. mmWave cells were unavailable for

```
20:19:31.945 Configure measurement on freq. channel F850,F66486,F66661,F9820.
20:19:32.405 Measure cell #462 on F66486.
20:19:32.805 Measure cells #340,#162) on F2175.
20:19:33.045 Measure cell #61 on F9820.
20:19:33.685 Report cell #61 on F9820.
20:19:33.738 Select cell #61 on F9820 as PCell.
20:19:33.738-20:19:37.222 Search 4G SCell(s) on frequency channel F850,F5110,
    F66486,F66661, F66936; Finally add SCell #370 on F66661.
20:19:33.886-20:19:37.222 Search 5G SCell(s) on mmWave channels F2253331,
    F2251665,F2254997, and sub-6G channel F174270; Finally add SCell #561
    on F174270 (5G sub-6G).
```

Listing 1: An illustrative example of CA (Frequency channels are represented by F plus standardized ID).

the phone and data speed was about 15.7 – 56.3Mbps[1]. However, our experiments show that there exists a superior option, PCell candidate on channel F850, at the same location. It usually works with mmWave cells and achieves a data speed of 87.8 – 156.3Mbps. Due to sequential measurement, it was missed when an early-measured cell got selected. Given sequential, cell-by-cell selection, the network made a much worse choice without assessing available SCells.

We study 5G mobility management in an empirical study with 3866 handover samples. We run 45-hour driving experiments with 5G phones with low mobility. The experiments are performed under AT&T, one of the largest wireless operators in the US. During the experiment, the devices keep active connection by downloading files and sending heartbeat to the servers. The driving speed ranges from 10-40 km/h to keep 5G connections. We

[1]The range is from p25 to p75 of all throughput samples.

Speed (km/h)	$0-40$
Average handover interval	41.9 s
Total failure ratio	2.7%
Feedback delay/loss	1.02%
Missed cell	0.32%
Handover command loss	0.7%
Coverage holes	0.67%

Table 3.2: Mobility in 5G.

collect signaling events with MobileInsight [78] to check handover events. Figure 3.2 shows how the carrier aggregation improves the throughput.

We find that mobility reliability becomes even become worse in 5G since handover will be more frequent and the issues persist. As shown in Table 3.2. The average handover interval is 41.9s, which is even smaller than 50.2s in 4G with 0-100 km/h driving speed. The failure ratio due to feedback delay/loss and handover command loss are 1.02% and 0.7%, which are even higher than 0.78% and 0.61% in 4G. This proves that the unreliability of signaling under mobility persists. The failure ratio due to missed cell and coverage holes dropped in 5G, which is benefited from the denser deployment. In conclusion, the signaling-strength-based mobility management in 5G is still susceptible to signaling loss and unreliable channel feedback. The policy-based decision is expedited due to denser deployment but the handover frequency increases at the same time.

3.2 Design Goals and Overview

Our goal is concurrent channel estimation for all cells in the super set upon mobility. Ideally, the super set includes every candidate cell the CA seeks to evaluate and select from. Note that, these cells can be deployed on multiple cell towers in the spatial proximity of the mobile

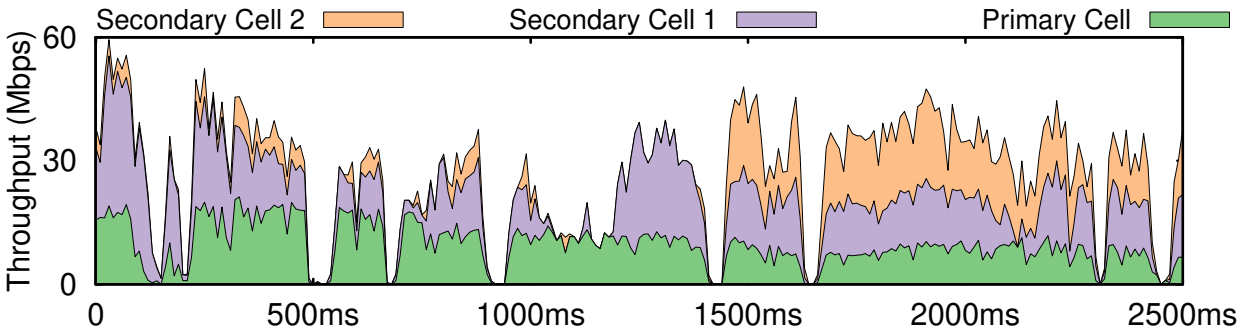


Figure 3.2: Carrier aggregation illustration.

device, with each tower having several colocated cells. To achieve the above goal, we address two issues:

• Given multiple, colocated cells on a single cell tower, *how to measure or infer the channel quality of all the cells quickly?*

• Given the superset of cells on multiple towers, *how to minimize the measurement cost to cover all the cells?*

To achieve both goals, `CCE` uses two components, which work in concert to address all two issues:

• **New algorithm for channel inference under mobility.** We adopt the delay-Doppler domain channel matrix, and devise a matrix-decomposition-based algorithm to infer the channel matrix. The algorithm is fast and robust under LoS scenarios. • **New algorithm for concurrent channel inference to decouple multipath.** Our algorithm accurately models the shared multipaths between the device and each colocated cell, and infers all cells from a single measurement. It thus measures one and infers all.

It exploits the channel representation in the delay-Doppler domain. Since all cells are colocated on the same tower, they generally share propagation paths. By modeling all shared paths (by delays) and mobility impact (by Doppler shifts), the channel model can be reused by all colocated cells. Moreover, By considering fractional Doppler shift, our

inference algorithm remains accurate over the wider spectrum space (say, both sub-6GHz and mmWave bands).

3.3 A Case with Single Path

3.3.1 Inference in Delay-Doppler Domain

We next introduce necessary details of wireless-mobility interplay for later sections. A time-varying wireless channel can be represented in multiple ways. 4G/5G measures its OFDM channel in the time-frequency domain. Equivalently, we can represent the same channel in *delay-Doppler domain* [32]:

$$h(\tau, \nu) = \sum_{p=1}^{P} h_p \delta(\tau - \tau_p)\delta(\nu - \nu_p), \tag{3.1}$$

where P is the number of physical paths, h_p, τ_p, ν_p are the p-th path's complex attenuation, path delay and Doppler frequency shift, respectively, and δ is the Dirac delta function. The delay-Doppler representation reflects the geometry of the multi-path in the client movement. Given $h(\tau, \nu)$ and the sent signal $s(t)$, a received signal can be obtained as

$$r(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\tau, \nu)s(t - \tau)e^{j2\pi\nu t}d\tau d\nu$$

The time-frequency channel response $H(t, f)$ in 4G/5G OFDM and delay-Doppler channel function $h(\tau, \nu)$ is related by

$$H(t, f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\tau, \nu)e^{j2\pi(t\nu - f\tau)}d\tau d\nu = \sum_{p=1}^{P} h_p e^{j2\pi(t\nu_p - f\tau_p)}$$

The time-frequency plane represents signals with a $M \times N$ grid by sampling frequency and time at intervals of subcarriers Δf (Hz) and symbols T (seconds). An OFDM symbol $x[m, n]$ locates at $(m\Delta f, nT)$, where $n = 0..N - 1$ and $m = 0..M - 1$. An OTFS frame at delay-Doppler plane is also represented with an $M \times N$ grid by M delay taps $\frac{1}{M\Delta f}$ and N Doppler taps $\frac{1}{NT}$. An OTFS symbol of $X[k, l]$ locates at $(\frac{k}{M\Delta f}, \frac{l}{NT})$, where $k = 0..M-1, l = 0..N-1$.

Table 3.3: Table of Notations

$X^i[k,l]$	transmitted OTFS symbol at index (k,l) for cell i
$x^i[m,n]$	transmitted OFDM symbol at index (m,n) for cell i
$Y^i[k,l]$	received OTFS symbol at index (k,l) for cell i
$y^i[m,n]$	received OFDM symbol at index (m,n) for cell i
k	indexes of Delay tap $\frac{1}{M\Delta f}$
l	indexes of Doppler tap $\frac{1}{NT}$
m	indexes of frequency tap Δf
n	indexes of time tap T

The transformation between OFDM and OTFS representation relies on ISFFT and SFFT. The OTFS modulator converts $X[k,l]$ to $x[m,n]$ with the inverse symplectic finite Fourier transform (ISFFT) and applies the Heisenberg transform to $x[m,n]$ using a transmit pulse to create a time domain signal. The signal is then transmitted over a wireless channel. The receiver applies the Wigner transform and discretization to the received signal. The resulting signal is $y[m,n]$ in the time-frequency domain. Then SFFT is applied to derive $Y[k,l]$ in the delay-Doppler domain.

$$x[m,n] = \frac{1}{NM}\sum_{k=0}^{M-1}\sum_{l=0}^{N-1} X[k,l]e^{j2\pi(\frac{mk}{M}-\frac{nl}{N})} \qquad \text{(ISFFT, DD}\rightarrow\text{TF)} \qquad (3.2)$$

$$Y[k,l] = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} y[m,n]e^{-j2\pi(\frac{mk}{M}-\frac{nl}{N})} \qquad \text{(SFFT, TF}\rightarrow\text{DD)} \qquad (3.3)$$

3.3.2 Limitations of Time-Frequency Domain

We first illustrate the limitations of current channel estimation in the time-frequency domain. In general, these algorithms can be categorized into two classes. One set of channel estima-

tions stems from the communication theory in the time-frequency domain. This model class is closely coupled with channel frequency. Specifically, the channel on the time-frequency domain is:

$$\hat{h}(t,f) = \sum_{i=1}^{P} a_i e^{j2\pi(t\upsilon_i - f\tau_i)}, \tag{3.4}$$

where P denotes the total number of propagation paths, and the i-th path is characterized by path attenuation a_i, delay τ_i, and Doppler shift υ_i upon mobility.

It is clear from (3.4) that the channel at time t is coupled with Doppler υ over each path. Without explicit decoupling of Doppler υ, the time-frequency model cannot be accurately modeled under mobility.

The second class of algorithms uses machine learning based modeling, and includes OptML [29] and R2F2 [120] for a sample of proposals in the literature. However, they do not explicitly take domain knowledge into account, and may suffer from under/over-fitting problems when seeking to improve the accuracy. Besides, they do not account for Doppler in the modeling. Their models simplify $\hat{h}(t,f)$ as $\hat{h}(t,f) = \sum_{i=1}^{P} a_i e^{-j2\pi\frac{d_i}{\lambda}+j\Phi_i}$, where d_i is the distance, λ is the wavelength, and Φ_i is the phase shift. Both OptML [29] and R2F2 [120] assume that only λ is different for different frequency although both a_i and Φ_i also vary across different frequencies under Doppler effects. Some approaches require external sensor/GPS assistance [58, 104], which limits their applicability.

3.3.3 Decomposition with Singularity

Sparsity for accuracy. We further leverage the sparsity of multi-path propagation to improve estimation accuracy and efficiency. To improve the estimation accuracy, we show that under the single path model, the SVD decomposition holds under no assumptions on Doppler and delay. To improve efficiency, the sparsity of propagation paths serves as a constraint to filter significant paths among possible $min(M,N)$ paths.

Propagation path sparsity in high-speed rails originates from the environmental setting

and operator deployments. Cellular networks for high-speed rails require dedicated antennas to serve fast-moving users [115]. The dedicated antenna guarantees that the line-of-sight (LOS) propagation path usually exists. The existence of the LOS path means a dominant path presents higher amplitude [121]. Thus a single path model captures the channel quality. This is why the single path model is adopted by the 3GPP high-speed-train propagation model [9].

We first relax the condition on delay and Doppler with the insight of propagation path sparsity. It relaxes the condition of SVD decomposition because the sparsity of propagation indicates the sparsity of the channel matrix. The prior work sets constraints on path delay and Doppler since the decomposed delay spread matrix $\mathbf{\Gamma}$ and Doppler spread matrix $\mathbf{\Phi}$ have to be semi-unitary. If the delay and Doppler are not integrals, the fractional delay and Doppler cause coupling between paths, destroying the unitarity of the matrix. However, there is no coupling between paths when there is a single path.

We prove that when there is a single path, SVD decomposition holds. Theorem 3.3.1 shows that the SVD decomposition holds under any delay and Doppler if the number of physical paths is one, as shown in Appendix C. To summarize, the proof relies on the insight that there is no coupling between paths. The prior work relies on the conditions on delay τ_p and Doppler ν_p to eliminate the coupling between paths. After we show that the delay spread matrix $\mathbf{\Gamma}$ and Doppler spread matrix $\mathbf{\Phi}$ are unit vectors, the decomposition is thus valid.

Theorem 3.3.1 (Single path decomposition)**.** *The compact SVD decomposition for a delay-Doppler channel matrix* $\mathbf{H} = \mathbf{U\Sigma V}$ *is equivalent to decomposition* $\mathbf{H} = \mathbf{\Gamma P \Phi}$ *with* $\mathbf{U} = \mathbf{\Gamma}, \mathbf{\Sigma} = \mathbf{P}, \mathbf{V} = \mathbf{G}$ *when* $P = 1$.

The condition of Theorem 3.3.1 holds when there is a single path. Theorem 3.3.1 indicates the correlation between the path amplitude and the decomposed singular value. Even when there are more paths, the dominant path corresponds to the major component of SVD

decomposition. During our evaluation, we find out that the decomposition error is small even when there are many propagation paths. This is because the dominant path is decomposed correctly.

Avoid over-fitting. We adapt the algorithm to filter out weak propagation paths. This is done by setting a threshold on the number of inferred paths to leverage path diversity. In the algorithm line No. 2, we constrain the number of paths as P_{max}. This design will not affect Theorem 1's validity since we perform path pruning after SVD decomposition. Specifically, we first rank all potential paths by their amplitude and select the P_{max} strongest ones. In practice, there are two ways to decide the threshold. The operators can decide the number of paths empirically based on 3GPP standards. For example, the number of paths is 5 for driving scenarios in the reference propagation model.

Another way to decide the threshold is to compare the estimated path loss to the strongest path. Inspired by the correlation between the path amplitude and the decomposed singular value, we rank the decomposed components based on the singular value. The largest singular value is mapped to the strongest path, which dominates estimation accuracy. Other weaker paths can be affected by interference and noises. We thus omit the paths that are lower than the strongest one. The offset to filter paths depends on the operators' experiences.

We further leverage the sparsity of multi-path propagation to avoid the overfitting of the SVD decomposition. Previous studies show that there is a limited number of paths in the environment [66, 126, 42]. To avoid overfitting too many paths, we constrain the number of paths. We thus set a threshold of maximum paths as P_{max}. Constraining the number of paths improves the efficiency as well. If $P_{max} = MIN(M, N)$, the optimization is reduced to the naive estimation. With $(M, N) = (1200, 560)$, there are 560 paths derived from SVD decomposition. In such a case, setting $P_{max} = 9$ reduces the computation of path by 62 times. The idea of constraining the number of paths adheres to the nature of SVD to reduce data dimension. With SVD extracting the path loss with the singular values in $\mathbf{P}$, the algorithm can easily rank the diagonal matrix and filter out weak ones.

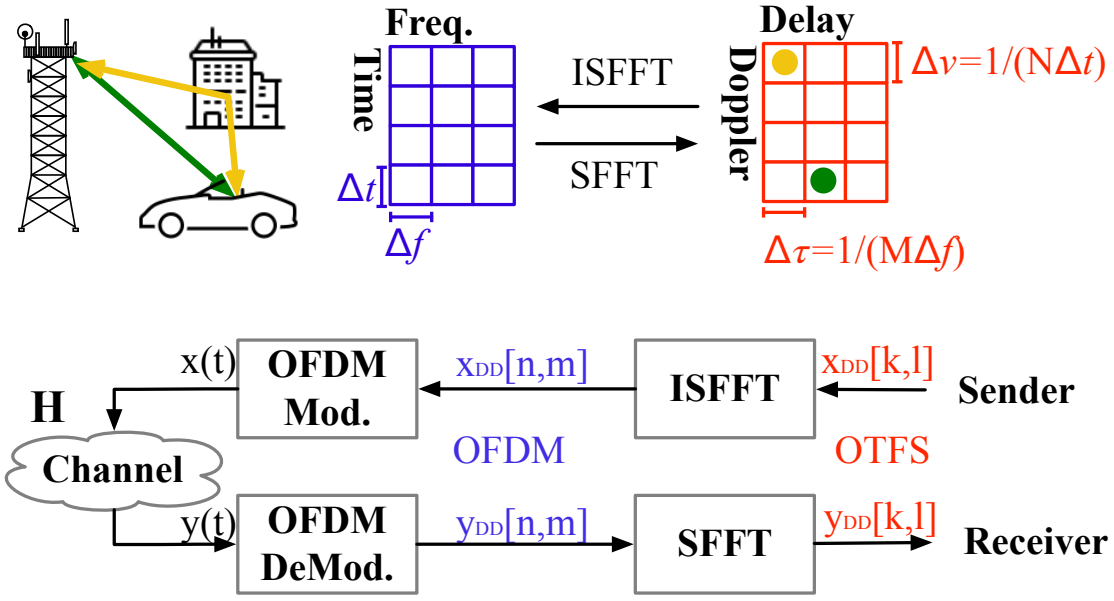


Figure 3.3: Illustration of cross-domain transformation.

3.4 Concurrent Channel Estimation

3.4.1 Multi-Path Modeling

We next show how to retrieve the shared multi-path characteristics; this is the most challenging step (i.e., Step 2). Our key finding is to use the delay spread to decouple the multiple paths in the delay-Doppler domain. To this end, we first show the limitations of the existing algorithm REM [77], and then elaborate on our design details.

Limitations of REM. REM is a recent algorithm that uses singular value decomposition (SVD) to decouple paths in the DD domain. However, its SVD based operation may only extract a single path in the presence of fractional Doppler as we elaborate next.

The main idea behind REM is to decompose the channel matrix $\mathbf{H}$ into the Doppler matrix $\mathcal{G}$ and the delay matrix $\mathcal{F}$. They may satisfy SVD by assuming that each path's delay and Doppler are multiples of the quantization steps of the DD grid. Equivalently, each path is located on the discrete DD grid with the integer coordinates $\kappa_i = \frac{\upsilon_i}{\Delta\upsilon}, l_i = \frac{\tau_i}{\Delta\tau}$.

Note that the quantization step of Doppler $\Delta\upsilon = \frac{1}{N\Delta t}$ depends on the grid duration $N\Delta t$ (Figure 3.3). In a typical 5G setting with carrier frequency $f_c = 38$ GHz, $N\Delta t = 5$ ms[2], we have the Doppler resolution $\Delta\upsilon = 200$ Hz, equivalent to a resolution of 5.7 km/h when mapped to the moving speed. Therefore, the assumption for REM does not hold in general; so as to the decomposition. Note that SVD relies on the singularity of the matrix. Since fractional Doppler violates the singularity condition, SVD will only yield a single path.

3.4.2 Delay-Based Decoupling

The foundation of `CCE` is to predict a single cell's quality *considering mmWave channels*. Wireless representation in delay-Doppler domain is applied to incorporate Doppler shift along with high frequency. We first introduce how to make cross-channel estimation by exploiting channel matrix decomposition. [77] has presented how to extract frequency-independent and frequency-dependent channel parameters, and project them to a different frequency. However, the prediction algorithm is not applicable to mmWave channels; the major reason is that assumptions for channel decomposition do not hold anymore. In our work, we explore new insights and techniques to overcome those challenges.

Challenges with mmWave channels. To speed up the measurement, the key is to simplify the measurement of physically co-located cells. To fully enable cross-channel prediction in 5G, it must be applicable to mmWave cells and new physical-layer settings in terms of grid size and numerology. New challenges have arisen from the new context. Unfortunately, the state-of-the-art algorithm on delay-Doppler domain [77] can tackle none of them. There are three main challenges:

◦ *C1. Channel estimation requires fractional Doppler to improve accuracy.* The existing mechanisms [77] assume the Doppler spread of each path is integral times of the Doppler resolution $\frac{1}{NT}$. However, the representation is over-simplified and inaccurate as the integral

[2]Based on the duration of reference signal in 5G (i.e. SSB burst set) [12].

resolution is quite low. We can translate the Doppler resolution to the speed resolution to make an intuitive explanation. Assume a practical set of parameters: the carrier frequency $f_c = 38$ GHz, sub-carrier spacing $\Delta f = 15$ kHz, $T = 66.7\mu s$, $M = 12, N = 15$. Therefore, delay resolution $\Delta\tau = \frac{1}{M\Delta f} = 5.5\mu s$ and it is sufficient to represent multiple paths. However, the Doppler resolution $\Delta v = \frac{1}{NT} = 1$ kHz, corresponding to speed resolution of 28.4 km/h. Such low resolution is hardly sufficient to represent the speed difference of paths. To make things worse, 5G would adopt radio measurement blocks (SSB and CSI-RS) with less period of time, which means smaller NT and larger spread of one integral Doppler. Therefore, it is essential to estimate the fractional Doppler together.

◦ *C2. Cross-channel prediction should be applicable to distant frequencies.* Given a much wider band, the channel spacing on mmWave bands is much higher. Predicting is much more sensitive to Doppler estimation error as the error is exaggerated when mapped to a distant frequency band. Unfortunately, the existing solutions like [77] could only work well with a small frequency gap (say, only 20 MHz difference). Under 4G/5G CA context, the device usually aggregates channels that are hundreds of MHz or even tens of GHz separated for mmWave bands. REM performs poorly (i.e., overestimate by 160%) and the error is as high as 5 dB. The estimation error affects both connectivity and CA efficiency. We thus need fine-grained cross-band estimation.

◦ *C3. Cross-channel prediction should take diverse numerology into account.* There are advanced physical-layer technologies to support variant sub-carrier spacing (Δf) and symbol duration (T) on different frequencies. As a result, the prediction algorithm should map channel parameters to another frequency by coordinating the grid size and resolution first.

Why can we predict channels? The intuition behind channel prediction is to retrieve the channel profile from one frequency and then map it to another frequency. Cells on the same base station share the physical paths of radio transmission to the user's device. The

wireless channel $h(\tau, \upsilon)$ is represented by

$$h(\tau, \upsilon) = \sum_{i=1}^{P} h_i \delta(\tau - \tau_i)\delta(\upsilon - \upsilon_i)$$

which consists of P propagation paths. Each path i is characterized by path coefficient h_i, path delay τ_i and path Doppler shift υ_i. Therefore, if we extract those channel parameters from the measurement of one cell, then we can transform them to another frequency and estimate channel quality there.

Fortunately, channel representation in delay-Doppler domain directly leads to the separation of frequency-dependent and frequency-independent parameters. Cellular network places grids on the physical channel along time and frequency dimensions. It consists of $M \times N$ cubes, and each cube spans T and Δf on time and frequency. One cube is the basic unit of a signal sample. Correspondingly, it maps to a $M \times N$ grid on the delay-Doppler domain, with delay resolution of $\frac{1}{M\Delta f}$ and Doppler resolution of $\frac{1}{NT}$. In delay-Doppler domain, the channel is

$$h_w(\tau, \upsilon) = \iint e^{-j2\pi\tau'\upsilon'} h_{\tau',\upsilon'} w(\tau - \tau', \upsilon - \upsilon')\, d\tau' d\upsilon'$$

It incorporates the convolution operation due to windowing effect of the $M \times N$ grid, and $w(\tau, \upsilon) = \sum_{l=0}^{N-1} \sum_{k=0}^{M-1} e^{-j2\pi(\upsilon l T - \tau k \Delta f)}$. Therefore, the channel estimation on delay-Doppler grid (k,l) has:

$$\frac{h_w(k,l)}{MN} = \sum_{p=1}^{P} \frac{\gamma(k,p)}{M} \cdot |h_p| \cdot e^{j\theta_p - j2\pi\tau_p\upsilon_p} \frac{\phi(l,p)}{N} \tag{3.5}$$

where $\gamma(k,p) = \sum_{k'=0}^{M-1} e^{-j2\pi(\tau_p - k\Delta\tau)k'\Delta f}$ and $\phi(l,p) = \sum_{l'=0}^{N-1} e^{-j2\pi(l\Delta\upsilon - \upsilon_p)l'T}$. Equation 3.5 can be represented in a form of matrix decomposition: $\mathbf{H} = \mathbf{\Gamma P \Phi}$ in which $\mathbf{H} \in \mathbf{C}^{M\times N}$, $\mathbf{\Gamma} \in \mathbf{C}^{M\times P}$, $\mathbf{P} \in \mathbf{R}^{P\times P}$ is a diagonal matrix, $\mathbf{\Phi} \in \mathbf{C}^{P\times N}$ and

$$\mathbf{H}(k,l) = \frac{h_w(k,l)}{MN}, \mathbf{\Gamma}(k,p) = \frac{\gamma(k,p)}{M}$$
$$\mathbf{P}(p,p) = |h_p|, \mathbf{\Phi}(p,l) = e^{j\theta_p - j2\pi\tau_p\upsilon_p} \frac{\phi(l,p)}{N}$$

The matrices $\mathbf{\Gamma}$ and $\mathbf{P}$ contain frequency-independent parameters only, say τ_p and h_p. While the matrix $\mathbf{\Phi}$ also has frequency-dependent parameters: the Doppler shift υ_p. We can calculate the Doppler shift of a different frequency with $\upsilon'_p = \frac{f'_c}{f_c}\upsilon_p$. [77] proposed to apply SVD for decomposition under the assumption of integer Doppler. Therefore, we can retrieve the accurate parameters, keep the independent part invariant, and map the dependent parameters to a different frequency channel. The start-of-the-art fails to do so because they cannot tackle challenges: fractional Doppler, wide frequency gap and heterogeneous numerology setting.

To our best knowledge, this is the first work to enable delay-Doppler-based channel prediction in 5G context. We make it happen by addressing aforementioned challenges with the following insights.

I1. Fractional Doppler can be extracted due to the deterministic relation with the observable channel matrix. The Doppler is composed of integer Doppler and fractional Doppler, and each manifest itself differently in the channel representation. The integer Doppler manifests as the peak index. The fractional Doppler manifests as inter-Doppler-interference between different Doppler taps. With the decomposition, we estimate each component separately. We find that the relation between fractional Doppler and max Doppler matrix $\mathbf{\Phi}$ intensity is deterministic (regardless of frequency). We exploit this finding in our algorithm to estimate the fractional Doppler, after the integral part is retrieved.

I2. Insights to deal with variant numerology in practice. By modeling the relationship between the interference amplitude and the fractional Doppler, we derive fractional Doppler after decomposing the channel matrix into frequency-dependent components and frequency-independent components. Multi-paths may share similar Doppler frequencies. Consider a common mobility case: $f_c = 38$ GHz, $v = 9$ m/s, then the Doppler spread is 1.14 kHz. If we assume $NT = 5 \times 66.7\ \mu s$ under the SSB setting, then the Doppler resolution would be 3kHz. As a result, all paths probably share the same integral Doppler, and have close fractional Doppler. Maybe this observation could explain the insights of using SVD

with fractional Doppler: $\mathbf{\Phi}$ becomes a one-rank matrix with a similar Doppler shift.

Algorithm. Our design `CCE` has addressed all challenges based on the insights. We first walk through how the wireless channel is composed in delay-Doppler domain. Then we derive the inverse function to get channel parameters, especially fractional Doppler from the wireless channel representation. Our main contributions are:

- A lightweight algorithm estimates fractional Doppler in addition to an integral part by transforming channel representation.
- A decomposition algorithm to estimate the number of paths to reduce complexity.
- Allow estimation across different numerology and grid settings for mmWave and low/mid bands.

From delay-Doppler channel representation, we derive the fractional Doppler by modeling the interference between different Doppler taps. Based on the channel representation with Φ, we derive the inversion to design the transform. We have $\mathbf{\Phi}(l,p) = \frac{1}{N}\sum_{n=0}^{N-1} e^{-j2\pi(l\Delta\nu-\nu_p)nT} e^{-j(\theta_p+2\pi\tau_p\nu_p)}$, for $l = 0, ..., N-1$. We have two observations: 1) the maximum of $\mathbf{\Phi}(l,p)$ is decided by fractional Doppler. 2) the phase of maximum $\mathbf{\Phi}(l,p)$ decided the sign of fractional Doppler.

We design the cross-band prediction algorithm as shown in Algorithm 1. The algorithm takes three steps. It first decomposes the channel matrix with SVD to separate frequency-independent and frequency-dependent parameters. Then it estimates the number of paths and calculates fractional Doppler for each path. Finally, the channel matrix is derived based estimated delay Doppler profile for each path. Our algorithm leverages the deterministic relationship between the Doppler matrix profile and fractional Doppler.

Heterogeneous bands affect delay-Doppler estimation. We consider the impact of both delay and Doppler. Delay is affected by different path delay sampling rates. If band 1 and 2 have different bandwidths (e.g., 15MHz V.S. 20MHz), we have $M_1 \neq M_2$. But M_i decides

Algorithm 1 Cross band estimation

Require: Band 1's channel matrix $\mathbf{H_1}$, $H_1(k,l)$

Ensure: Band 2's channel matrix $\mathbf{H_2}$

1: Decompose $\mathbf{H_1} = \mathbf{\Gamma P \Phi_1}$ using SVD matrix factorization;

2: Estimate the number of path $p \leftarrow rank(P)$;

3: **for** each path $p = 1, 2, ...p$ **do**

4: Derive $\nu_p^1(int), \nu_p^1(frac)$ based on $\Phi_1(max)$;

5: Estimate ν_p^1 based on $phase(\Phi_1(max))$;

6: $\nu_p^2 \leftarrow \nu_p^1 \frac{f_2}{f_1}$;

7: **end for**

8: Calculate Φ_2 based on ν_p^2;

9: $\mathbf{H_2} \leftarrow \mathbf{\Gamma P \Phi_2}$;

the sampling rate of path delay as $\Delta\tau_i = \frac{1}{M_i \Delta f}$. Therefore, estimating band 2 based on band 1 may have accuracy loss due to different path delay sampling rates.

For Doppler, it causes carrier-specific frequency shifts. For each propagation path, the Doppler frequency shift ν_i for band i depends on the carrier frequency f_i, client speed v, and path phase θ: $\nu_i = \frac{v \cos\theta}{c} f_i$. Given the same path, we have $\nu_2 = \frac{f_2}{f_1}\nu_1$ between band 1 and 2. Since $\nu_i = l_i \Delta\nu, \nu = 1/NT, i = 1, 2$, we have $l_2 = \lceil l_1 \frac{f_2}{f_1} \rceil$ That is, when mapping from band 1 to band 2, a *fractional* Doppler axis shift occurs for $\{y_p^2[k,l]\}$. This results in accuracy loss and interference between Doppler axes.

We take a two-step approach. We first assume two bands' carrier frequencies are identical ($f_1 = f_2$) but with different bandwidths ($M_1 \neq M_2$). In this situation, we adopt re-sampling for cross-band estimation, and prove the estimation relationship between two bands. Then we generalize the result to different carrier frequencies $f_1 \neq f_2$.

Resampling theorem in Fourier transform. When $f_1 = f_2$, mapping from band 1 to 2 can be viewed as a resampling process (with sampling rate $\tau_1 \rightarrow \tau_2, \tau_i = \frac{1}{M_i \Delta f}$) along the

delay axis. We start from the integer upsampling/downsampling process, and generalize it to fractional resampling.

In the integer downsampling (e.g., estimating 5MHz band 2 from 20MHz band 1), let $M_1 = TM_2$. Then we have

Proposition 1 (Integer downsampling)**.** *If $x_2[k,l] = x_1[kT,l]$ and $M_2 = TM_1, T \in \mathbb{N}$, then band 2's time-frequency signal after OTFS modulation can be derived from band 1 as*

$$X_2[n,m] = \frac{1}{T}\sum_{q=0}^{T-1} X_1[n,[m+qM_2]_{M_1}] \tag{3.6}$$

and the received delay-Doppler signal in band 2 can be inferred from band 1 as

$$y_2[k,l] = \frac{1}{NM_2}\sum_{k'=0}^{M_1-1}\sum_{l'=0}^{N_1-1} h(k'\Delta\tau_1, l'\Delta\nu)c[k'T]x_1[[kT-k']_{M_1}, l-l'] + n[k,l] \tag{3.7}$$

where $[\cdot]_{M_1}$ denotes modulo M_1 operations, and

$$c[k] = \begin{cases} 1 & \textit{if } k\textit{=0,T,2T,...} \\ 0 & \textit{otherwise} \end{cases} = \frac{1}{T}\sum_{q=0}^{T-1} e^{-j\frac{2\pi}{T}qk}$$

This is proved in Appendix A.3. The key result is that, *without directly measuring band 2*, we can get $\{y_2[k,l]\}$ from band 1, and then infer band 2's SNR with given $\{x_2[k,l]\}$ pilots.

3.4.3 Combating Interference and Noise

The algorithm assumes a perfect channel response matrix. However, this might not be true due to channel interference and noise. Channel interference and noise are common in wireless networks due to the shared channel. This will lead to an inaccurate channel responses matrix **H** and harm the accuracy of the Algorithm.

`CCE` takes two steps to combat channel interference and noise. First, we leverage existing interference cancellation algorithms to get an accurate channel response **H**. There are

established channel response estimation schemes [106, 108] to get **H** for the measured channel. `CCE` leverages existing channel response estimation algorithms to calculate the channel response matrix. `CCE` then applies the algorithm to estimate other unmeasured channels.

Second, `CCE` enhances the robustness of Algorithm 1 by calibration. The calibration is done based on the computation with a row of the matrix **H** instead of a single element $|H[l_p, 0]|$. Specifically, the calculation is based on Theorem 3.4.1. For each row in **H**, `CCE` extracts $|H[l_p, k]|$ and uses the square sum to derive path attenuation h_p. Note the matrix **H** is inaccurate and all rows might be non-zero due to errors. We thus count all rows as possible paths.

Theorem 3.4.1. The path attenuation h_p and channel response matrix **H** satisfy $h_p = \frac{1}{M^2N^2}\sqrt{\sum_{k=0}^{N-1}|H[l_p,k]|^2}$

The calibration benefits cross-channel estimation in two aspects. By combining multiple elements, the absolute error is constrained. Suppose $|H'[l_p, k]| = |H[l_p, k]|+n$, where $|H[l_p, k]|$ is the ground truth and $|H'[l_p, k]|$ is estimated. If n is randomly distributed, the error in h'_p calculated from $|H'[l_p, k]|$ is no more than the most inaccurate $|H'[l_p, k]|$, indicated by $\frac{h'_p}{h_p} <= max_k(\frac{|H'[l_p,k]|^2}{|H[l_p,k]|})$. Besides, the calibration reduces the impact of a single false positive path. The calculation of path attenuation for each path is independent. Even if noises cause false positive paths, the estimated false paths are weak as long as n is small. The overall estimation is dominated by paths with strong h_p. In practice, `CCE` drops weak paths if the estimated h_p of the path is 50 dB lower than the strongest path.

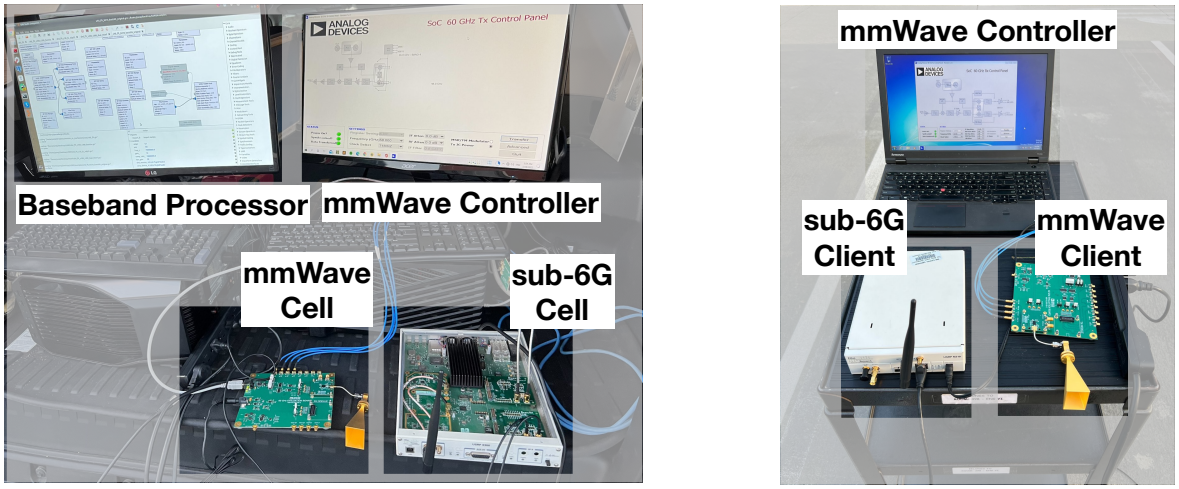


Figure 3.4: Experiments settings.

3.5 Evaluation

3.5.1 Methodology

Our evaluation uses both testbed-based experiments and trace-driven emulations. We gauge the accuracy and efficiency of concurrent channel inference with the SDR testbed under real-world channels. We further quantify the throughput boost by `CCE` and its efficiency using real-world traces.

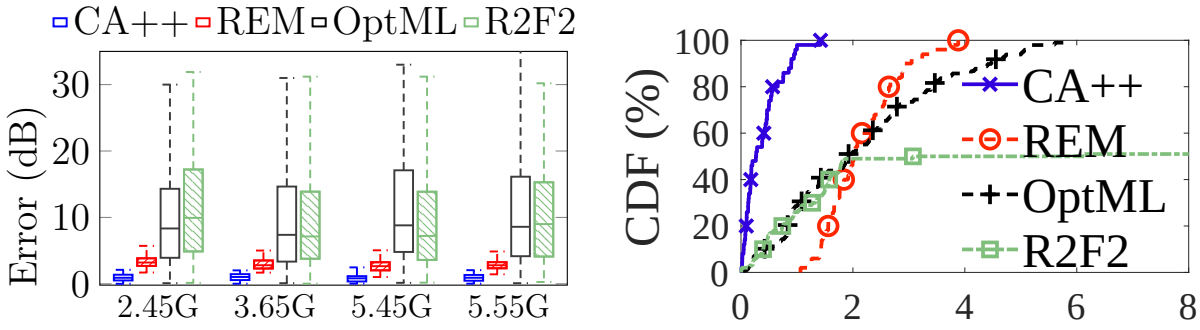
Testbed and experiment settings. Figures 4.9 and 3.5 show our testbed setup. The testbed supports both sub-6G and mmWave bands. The cell tower uses USRP X300 for sub-6G, and operates with the 60GHz HMC6350 TX frontend [44] for mmWave. USRP N210 acts as the sub-6G client device; USRP X300, together with HMC6350 RX frontend and horn antennas, serves as the mmWave client. The USRPs run baseband processing and feed signals to sub-6G and mmWave frontends. They are connected to the server running Ubuntu 18.04 (Intel i7-9700k CPU with 32G RAM) for baseband signal generation. We use the testbed to conduct indoor and outdoor experiments, in both static settings and low-mobility (about 10km/h) scenarios.

Trace-driven emulation. Since `CCE` is not deployed by mobile operators, we use large-

(a) Indoor experiments (b) Outdoor experiments

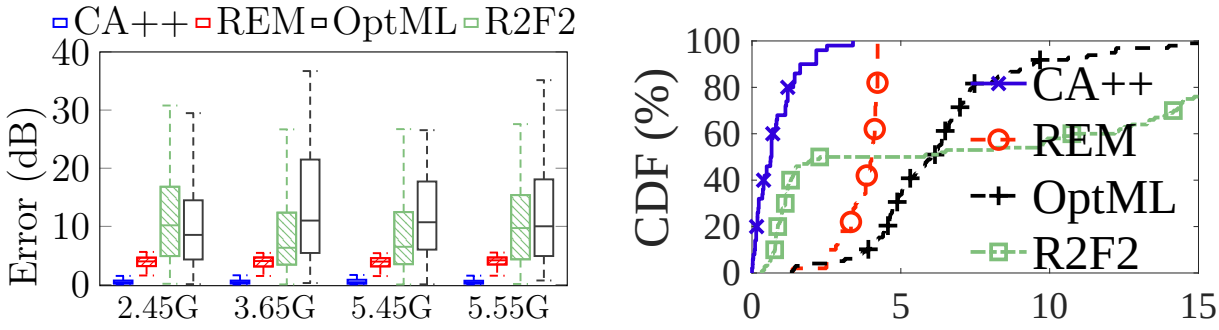
Figure 3.5: Indoor and outdoor setup.

scale datasets collected from commercial networks to perform trace-driven emulations. Effectively, it is a "what-if" study with the following steps. First, we run walking/driving tests with file downloading tasks on phones. Cell deployment, throughput and signal strengths are all recorded. Meanwhile, we extract logs on CA operations using MobileInsight [78]. Second, for each CA instance, we assess *all* available cell combinations and remake the choice on a group basis by exploiting many runs at the given location. As we cannot know the actual signal strength and throughput of missed options at runtime, we feed the emulator with statistical data in history to approximate those conditions. Table 3.4 shows the five datasets used for emulations: (1-4) A-C1, V-C1, T-C1 and A-C2: We run tests with three major operators (A, V and T short for AT&T, Verizon, and T-Mobile, respectively) in one US Midwest city (C1). We also run experiments with A in one large city (C2). Both A and V support 5G over both sub-6G and mmWave bands in the test areas, but T supports only sub-6G. Since we need sufficient data on cell deployment and CA usage, we run extensive experiments to scan the test regions (over 5,700 km and 400 hrs in total). We keep downloading files (500MB each from Google Cloud) to measure the downlink throughput. (5) 4G-HST: We use a public dataset [122] collected on High-Speed-Train (HST) commuting between Shanghai and Beijing, China. 5G was not deployed on HST yet.



(a) From 2.45G (b) From 58G to 62G

Figure 3.6: SNR inference under low mobility.



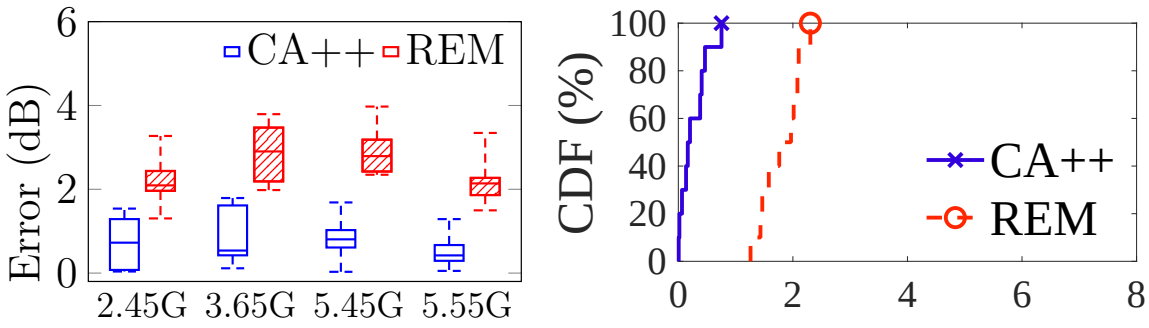
(a) From 2.45G (b) From 58G to 62G

Figure 3.7: SNR inference under high mobility.

3.5.2 Concurrent Channel Inference Accuracy

We assess concurrent channel inference under a variety of settings on frequency bands, mobility, locations and diverse propagation paths, in both indoor scenarios (in a lab space of Figure 3.5a) and outdoor scenarios (at the top level of a parking structure of Figure 3.5b). We compare `CCE` with prior schemes in the delay-Doppler domain (REM [77]) and in the time-frequency domain (OptML [29] and R2F2 [120]). Both OptML and R2F2 require to configure the number of paths for high accuracy. We use their optimal configurations in our tests.

Channel inference accuracy. We first assess the inference accuracy under both low and high mobility. We run low mobility experiments using a mobile cart that carries the client devices and moves at the speed of about 10 km/h. We emulate high mobility



(a) From 5.55G

(b) From 62G to 58G

Figure 3.8: Impact of frequency-to-measure.

Method	**Min**	**Max**	**Mean**
CA++ (**ms**)	0.6	44.1	8.2
REM (**ms**)	2.0	131.1	51.7
OptML (**ms**)	213	288	239
R2F2 (**s**)	2.5	59.3	3.5

Table 3.5: Runtime.

in outdoor scenarios. We extract the propagation model from the low-mobility traces and replay these traces with the mobility settings collected from the HST dataset (i.e., 150 km/h – 300 km/h). The propagation paths are extracted by fitting the channel responses with the optimal Doppler shifts $\{(\upsilon_i)\}_{i=1}^{P}$. To replay the propagation paths under high mobility, we scale the Doppler shifts according to the moving speed v, i.e., $\upsilon_i = \frac{v}{c} f$. We consider two frequency ranges: (1) sub-6GHz (i.e., measuring over 2.45 GHz to infer four unlicensed sub-6G channels at 2.55 GHz/3.65 GHz/5.45 GHz/5.55 GHz) and (2) mmWave (i.e., measuring over 58 GHz to infer 62 GHz). We assess the inference accuracy in terms of SNR/RSRP/RSRQ errors. The results are similar (see next); we only present SNR results in Figures 3.6 and 3.7, due to space limitation.

`CCE` outperforms over mmWave as well. Its error even goes down to 0.37 dB, much lower

than 2.03 dB, 1.92 dB, and 4.60 dB by REM, OptML and R2F2, respectively. All schemes perform better over mmWave, because the measurement-inference frequency gap becomes relatively smaller (4GHz over 58-62 GHz versus >1 GHz over 2.45-5.55GHz except in the 2.55 GHz case). `CCE` performs much better, because its delay granularity increases with the center frequency. We use the median error unless specified.

`CCE` outperforms other approaches even more under high mobility. Compared to low mobility, the error is smaller over sub-6G (1dB $\rightarrow$ 0.4 dB) and slightly larger over mmWave (0.37 dB $\rightarrow$ 0.64 dB). REM, OptML and R2F2 all perform worse over mmWave under high mobility. For sub-6G bands, the inference errors grow to 3.94 dB–4.16 dB (REM), 6.3 dB–10.2 dB (OptML) and 8.53 dB–11.0 dB (R2F2). `CCE` performs better under fast mobility, because the Doppler shifts for diverse paths become more significant. All three prior schemes cannot provide accurate estimation, as they fail to capture or precisely model time-varying Doppler.

We note that, high accuracy gain tends to diminish over mmWave (ultra-high frequencies). It confirms the fact that radio signal over ultra-high frequencies fades much faster than over low frequencies.

Effectiveness over various frequencies and wide spectrum. `CCE` remains effective regardless of measured frequencies. Figure 3.8 shows the results when measuring 5.55 GHz (sub-6G) and 62 GHz (mmWave) under low mobility (outdoor). In fact, we run tests with all five sub-6GHz channels, and omit the results using three other sub-6GHz channels because they are consistent. `CCE` outperforms REM by about 2 dB, reducing the SNR errors from 2.16 dB/2.89 dB (min/max) to 0.5 dB/0.86 dB (min/max). This also applies to mmWave (from 62 GHz to 58 GHz). `CCE` also does well for channel inference across sub-6GHz and mmWave (the plot omitted). It outperforms REM, and reduces the SNR inference error by a factor of 9.0, from 3.16 dB to 0.35 dB.

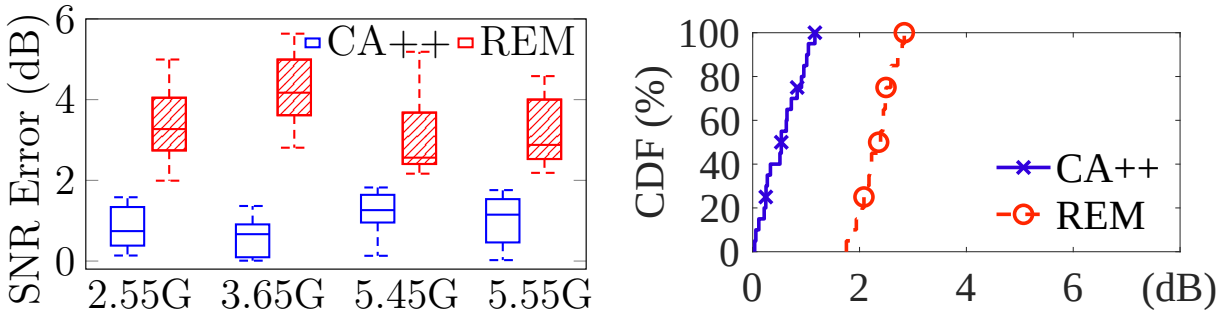
Robustness to measurement settings. CA++ is also invariant of the frequency to

measure. Figure 3.8 shows that CCE (0.50 - 0.86 dB) still outperforms REM (2.16 - 2.89 dB) by around 2 dBs. We also validated the performance under longer distance transmission, where SNR drops from 30dB to 14dB. CCE causes an average error of 0.94 dB to 1.04 dB, while REM causes an average error of 3.11 dB to 3.32 dB. CA++'s error increases by <0.5dB while REM's increases by >1dB. This is because CA++ is more robust in inferring path attenuation under noises.

Robustness to various settings. CA++ is robust to various numerology settings. 5G supports 4 numerology settings, with its subcarrier spacing being 15kHz, 30kHz, 60kHz, 120kHz(denoted with numerology index 0, 1, 2, 3). We test with various subcarrier spacing Δ_f : 15kHz - 30kHz, 15kHz - 60kHz, 15kHz - 120kHz, and 60kHz - 120kHz under high mobility (the plot is omitted). CCE contains the median error within 1dB for cross-numerology channel inference.

We also gauge the inference accuracy under weaker signals and stronger noises, where SNR degrades from 30dB to 14dB, as we move the client away from the cell tower (from 3 m to about 6 m). In case of measuring 2.45 GHz to infer other sub-6GHz channels, CCE increases its inference error by <0.5 dB, whereas REM increases by > 1dB (Figures are omitted). This is because CCE is more robust in inferring path attenuation under noises. We also test CCE in indoor experiments under low-mobility (walking) or static settings (where the cells are placed at one location and the clients are placed at the rest locations). CCE's inference accuracy is comparable in these indoor experiments (the plots are omitted).

Our algorithm provides high-fidelity inference under a widespan of radio spectrum (from 2.45GHz to 5.55GHz for sub-6G bands, and from 58GHz to 62GHz for mmWave bands). We also tested under both outdoor and indoor environments under low mobility. We note that the results under both settings are similar. In the outdoor setting, CCE yields a median error of about 1 dB ([0.91 dB, 1.03 dB]) when measuring 2.45 GHz (over sub-6G). In contrast, REM incurs a median error of about 3 dB ([2.66 dB, 3.21 dB]). In the indoor setting, Figure 3.9 shows that CCE and REM perform similarly. CCE induces an error around 1dB ([0.27 dB,



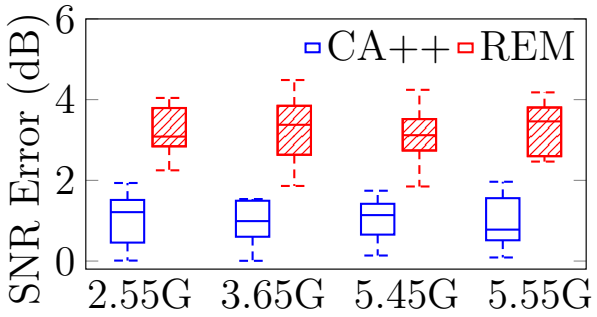
(a) From 2.45 GHz to sub-6GHz

(b) From 62GHz to 58GHz

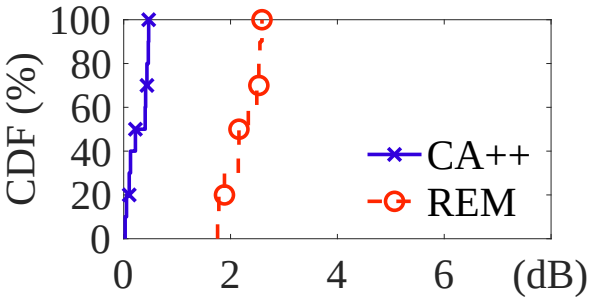
Figure 3.9: SNR inference errors of indoor.

1.26 dB]), while REM induces a median error of [2.56, 4.17 dB] for sub-6G scenarios. For the mmWave bands, CCE induces a median error of 0.54 dB, while REM induces a median error of 2.38 dB. This is because the indoor propagation conditions are more complex in layout with more obstacles. We also vary the position of the clients, and find that CCE consistently outperforms REM. As shown in Figure 3.10a and Figure 3.10b, we move the client away from the cell tower (from 3m to about 6 m). In the case of measuring 2.45 GHz to infer other sub-6GHz channels, CCE increases its inference error by <0.5 dB, whereas REM increases by > 1dB. For indoor experiments with 14 testing locations (Figure 3.10c), CCE's inference accuracy is comparable in these outdoor experiments.

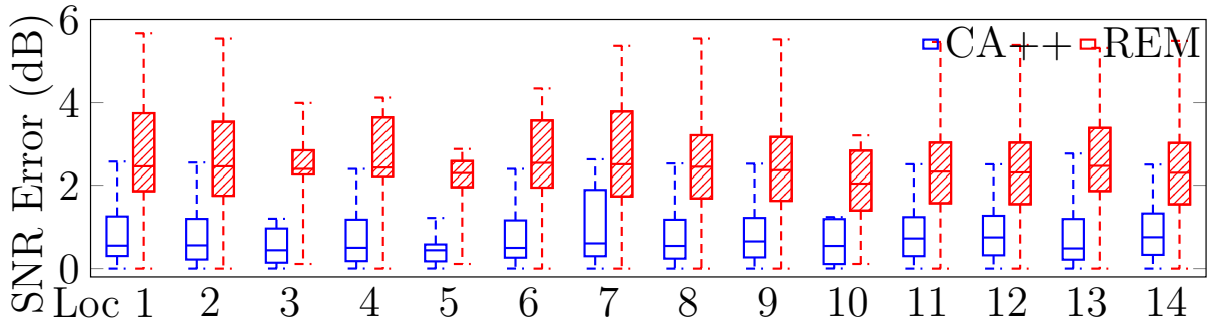
Evaluation on RSRP and RSRQ. We compare RSRP, RSRQ, and SINR estimation via CCE and legacy measurements. Figure 3.13 plots the inference errors under low mobility (outdoor). Results under high mobility are similar and omitted. In Figure 3.13a, we combine all samples when inferring each of four sub-6GHz channels using the measurements over 2.45 GHz. RSRP measurements are within [-140dBm, -45 dBm], and RSRQs are mostly in [-20 dB, -4.5 dB]. The reported RSRP (RSRQ) values are quantized with a resolution of 1 dBm (0.5dB) [16, 11]. CCE consistently outperforms REM, regardless of the radio quality metric. Compared with REM, CCE reduces the median error from 3 dB/dBm to below 1 dB/dBm over sub-6G. Its RSRP/RSRQ errors even go to zero (partly due to quantization)



(a) From 2.45 GHz to sub-6GHz at outdoor



(b) From 62GHz to 58GHz at outdoor



(c) From 5.45GHz to 5.55GHz at indoor

Figure 3.10: SNR inference errors at dif. locations.

over mmWave.

Achievable data rate. To quantify how the channel inference error affects achievable data rate, we gauge the estimation error of data rate based on the SNR inference accuracy. We use the standard SNR to spectral efficiency mapping to estimate achievable data rate in 5G [15]. The base stations decide the spectral efficiency by adapting modulation to received radio quality reports. Figure 3.12 compares `CCE` with REM by data rate estimation error for mmWave scenario. If the estimation is wrong, the base station will aggregate cells with overestimated or underestimated quality, which causes under-utilization. We use 100MHz setting to assess how the data rate deviates from the ground truth with `CCE` or REM under different mobility and SNR. Under low mobility, `CCE` outperforms REM with 87% - 91%

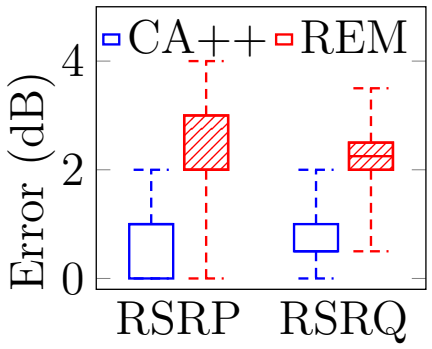
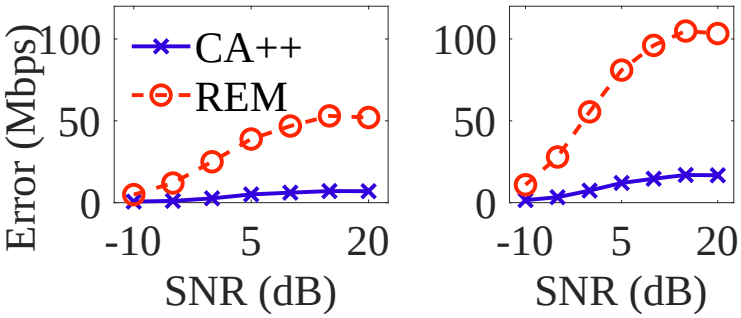


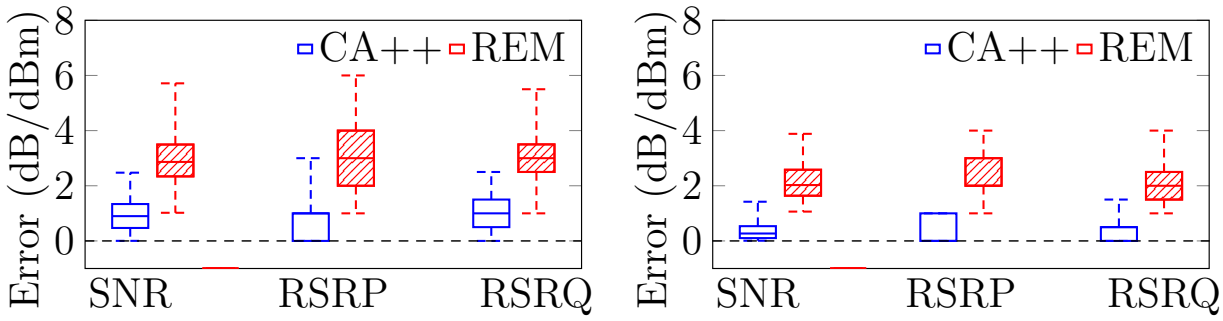
Figure 3.11: RSRP/RSRQ.



(a) Low mob.

(b) High mob.

Figure 3.12: Data rate.



(a) From 2.45 GHz to sub-6GHz

(b) From 58 GHz to 62 GHz

Figure 3.13: SNR/RSRP/RSRQ errors under low mobility.

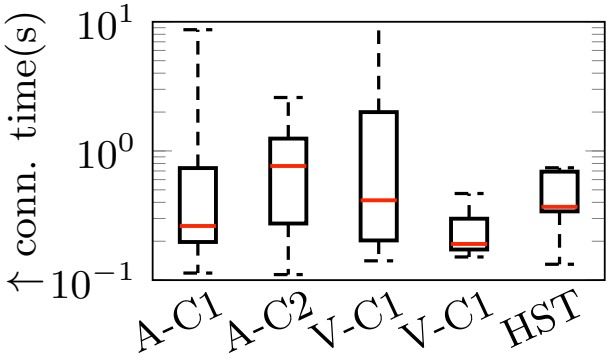


Figure 3.14: Increased connection time.

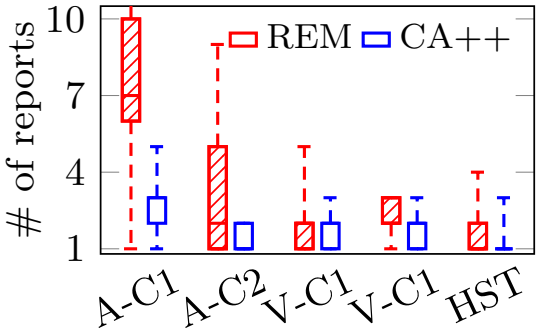


Figure 3.15: Signaling efficiency.

reduction of error. Under high mobility, CCE outperforms REM with 84% - 88% reduction of error.

Efficiency. We compare the efficiency of all four algorithms by measuring the time needed to get channel inference using the same set of data. Table 3.5 shows that CCE takes 8 ms on average, compared with 52 ms by REM and 238 ms by OptML and 3.5 s by R2F2. CCE and REM outperform OFDM-based algorithms by more than an order of magnitude as they do not rely on optimization with many iterations. Compared to REM, CCE further reduces the execution time by decoupling the sparse propagation paths and performing inference separately. Such efficient inference makes CCE promising to accelerate the measurement; The processing is faster than measuring one frequency in 5G, i.e., 40 or 80 ms.

3.5.3 Overall Improvements

We check the overall improvement by CCE. We use trace-driven emulations to assess the overall improvement by CCE. Since the device never knows the runtime performance of those un-selected cell sets, we use historical data to profile the performance of the cell sets available. We perform a "what-if" study to compare the CA options enabled by CCE and selected by the legacy (current) practice. Note that this assessment is only feasible for our 5G datasets; The public 4G-HST dataset lacks sufficient speed tests.

Measurement acceleration. We first estimate how much the group-based operation could speed up measurement. Accelerated measurement leads to an earlier selection of the next serving cell set, and thus prolonged time on a stable connection. As Figure 3.14 presents, the increased connection time (median) is 190.5 ms – 765.0 ms for 5G operators, and 369.3 ms on the high-speed train. It will benefit all scenarios in our experiments. In static/low-mobility cases, mobile users get more time to enjoy the complete aggregation of all serving cells and thus higher data speed. Under extreme mobility like the high-speed train, the network can provide smooth service with a lower risk of connectivity loss or acute performance drop.

Comparing different datasets, we note that `CCE` further accelerates measurements as more frequency channels are deployed. It yields faster measurements for AT&T 5G than Verizon 5G, T-Mobile 5G, and 4G on HST in China. This is because AT&T 5G has deployed more frequencies than others (Table 3.4). A single measurement covers more colocated cells on different channels. Our design thus yields more benefits with expanding spectrum resources in 5G and beyond.

Signaling efficiency. We examine the efficiency of group-based feedback in terms of signaling overhead. We compare the number of reports needed to include all eligible cells under legacy CA and `CCE`. Figure 3.15 proves that `CCE` could save the signaling overhead. Under low/moderate mobility and AT&T 5G in C1(C2), the legacy CA incurs 7(2) reports as the median value, while `CCE` reduces the number to 3(1). Compared to AT&T 5G, the benefit is slighter for other operators and scenarios. For Verizon 5G, T-Mobile 5G and China Unicom 4G (on HST), `CCE` reduces signaling overhead in 13.2%, 66.2%, and 31.0% of cases, respectively; And the median factor of reduction is 2 for all operators. This is because fewer frequency channels are deployed by each cell tower. Therefore, the number of cells to concurrently measure and include in one report is smaller and restricts the efficiency.

3.6 Discussion

Infra side cooperation. Infrastructure side: Reduce the signaling overhead by predicting the user movements, grouping their signaling operations (e.g., tunneling), and proactively performing these operations without delaying user traffic. With our proposed wireless link predictions in §4, the infrastructure can obtain more predictive channel measurements under high-speed mobility and 5G mmWave. By replacing existing wireless estimation with OTFS-powered ones, the wide-area roaming policies will obtain near-constant, predictable inputs for efficient and robust handover decisions.

Client side prediction. Client-side prediction with small-data, latency-friendly distributed learning: To predict the upcoming handover, the client should first infer the operator's handover decision logic. Inferring the network-side handover logic is challenging for two factors. First, the connected-state handover decision logic can be operator specific. The 3GPP standards leave the freedom for operators to customize their decision logic. Second, the device does not have full access to all network-side operations. It has to rely on its observations.

Colocated cells on a tower. In case cells from the same tower use different propagation paths to the device (say, deployed at different tower heights), we may categorize these cells into a few classes based on their path differences to the device. Moreover, beamforming in 5G [6] allows to adjust the antenna direction at runtime. In this case, CA would consider beam management [51] by steering antennas. `CCE` also helps by treating the antenna orientation as a new dimension on measurements and selection in CA. Intuitively, the device still performs once-for-all measurements in each direction per tower. In the extreme case, `CCE` degenerates to the current CA in both measurements and selection, but will never perform worse than the legacy CA.

3.7 Related Work

Channel state inference is becoming more challenging as the mobile network evolves with new mmWave technologies. Recent studies focus on two directions: single-channel state inference and cross-channel state inference.

Single channel state inference. Channel state inference is critical for link rate adaptation and channel selection. Existing works predict the channel in the time-frequency domain (e.g., OFDM). They adopt different approaches to capture the time-frequency domain patterns, e.g. spatial-temporal correlation [81, 130, 90, 110]. The major approach for inference is based on channel modeling and optimization [32, 87], machine learning [80], compressive sensing [28, 33], and subspace algorithm [22]. However, such schemes are susceptible to Doppler effects. Recent works [105, 108] focus on OTFS-based channel state inference with embedded OTFS signals with a pure delay-Doppler domain approach. Nevertheless, these approaches cannot be translated into time-frequency domain channel estimation.

Cross-channel state inference. With the emerging MIMO technology, cellular networks require an immediate downlink channel state without waiting for clients' feedback. R2F2 [120] is the first work that leverages reciprocity to transform uplink channel measurements into downlink channel states, which is inspired by [68]. Recently, deep learning-based approaches [60, 29, 83] are drawing more and more attention recently. Among them, OptML [29] first recovers the underlying paths and then reconstructs the channel states but Fire [83] performs estimation in an end-to-end approach. However, these approaches are all limited by the machine learning models that are hard to generalize. Another related domain is sensing algorithms and systems that leverage motion information for channel estimation with WiFi [23, 67, 111, 101, 69], mmWave [138, 125], RFID [123, 85], and acoustic signals [133, 134, 91, 124]. However, these approaches usually ignore multipath profile and focus on the dominant path only.

Dataset	**A-C1**	**V-C1**	**T-C1**	**A-C2**	**HST** [122]
Date	Apr 2021 - March 2022				Nov 2018
Region	1.65×1.85 km^2			1.2×1.0	1,300-km
Speed (km/h)	driving: 10-40 (mostly); walking: <5				300 - 350
RAT	5G + 4G (T supports 5G over sub-6G only)				4G
Max# CA CH	6	5	5	7	3
# CA groups	5,681	2,037	492	3,031	534
Max CA CW	430 MHz	140 MHz	135 MHz	445 MHz	50 MHz
Summary of 5G cells					
# sub-6G cells	62	21	55	38	N/A
# mmWave cells	372	100	0	39	
# sub-6G CH	3	1	3	2	N/A
# mmWave CH	16	4	0	9	
sub-6G freq.	826–2116	885	626–2608	826–2116	N/A
mmWave freq.	38.6–39.5	27.9–28.3	0	38.6–39.5	
sub-6G CW	5	10	15–100	5	N/A
mmWave CW	100	100	N/A	100	
Summary of 4G cells					
# cells	1,719	1,228	878	1,490	1,910
# Channel	20	40	12	18	8
freq. (Mhz)	709–5824	701–5825	701–2539	709–5824	1740–2155
CW (Mhz)	5/15/10/20 MHz				

Table 3.4: Datasets. CH: Channel. CW: Channel-Width. Sub-6G freq. are in MHz. MmWave freq. are in GHz.

CHAPTER 4

Embracing Interference with X-Domain Cancellation

This chapter is organized as follows. §4.1 introduces the requirements for interference cancellation. We then detail the problem statements and the challenges in §4.2. We devised two essential components for interference management, interference estimation (§4.3) and interference cancellation (§4.4). §4.5 presents our implementation and evaluation. We discuss the open issues in §4.6 and set our work apart from related works in §4.7.

4.1 Motivation

Drones are increasingly valuable in modern industry. We have witnessed the emergence of drone-based surveillance, delivery, and many more new applications [24]. The mobility of drones is qualified to enable applications in more challenging environments and more flexible use cases. For example, drones can enable continuous and all-area 3D data collection at large industrial facilities dueling automated processes and decisions. The remote control of drones increases efficiency and practicality by reducing manual efforts. The drone applications need seamless network support to deliver frequent and real-time data.

Nowadays, the only large-scale mobile network is cellular networks. Thus we ask the question, *Can cellular networks support drones?* As required by always-on surveillance or automated delivery, the network should always be available on demand and ready to deliver high-throughput drone traffic. When the drone moves with three dimensional flexibility, it needs seamless and reliable mobility support. Besides, emerging high-throughput applica-

tions also requires real-time feedback from the other end. It thus demands high bandwidth and ultra-low end-to-end latency.

Current cellular network is a promising approach to enabling drone-based demanding applications with the advances of 5G high-capacity radio and cloud-based core network. In 5G and beyond, the operators are opening up hundreds of GHz bandwidth for demanding applications [74] and offering low-latency services like URLLC [98]. However, there is no study on whether the current cellular network can support drone-based applications. We thus studied the state-of-the-art 5G networks. Unfortunately, we discovered two main problems, mobility, and data reliability.

First, the performance on the move is restricted by poor handovers. Drones rely on handover to gain seamless connection from one serving area to another. We find that handover is more frequent and more error-prone for the cellular network in the air. Handover frequency is two times higher for drones than ground devices along the same route. This is because Drones can measure more cells due to lack of blockage compared with ground devices. More measurements and inappropriate decision metrics trigger more handover. Besides, the handover failure ratio is higher for drones. Drones rely on inaccurate cell quality indicators for handover decisions.

The root cause lies in the design of the OFDM-based channel. OFDM-based handover decision metrics are unreliable. These RSRP/RSRQ/SINR metrics are erroneous due to interference. Thus measured the strength of resource elements is not accurate under Doppler and inter-cell interference. We validated that the correlation between measured channel quality indicators (RSRP/RSRQ/SINR) and ground truth SINR is weak. Besides, these metrics are for the downlink only. However, the DL/UL have different channel quality, and the interference can be very different. Finally, the interference due to drone communication is difficult to eliminate. We discovered that cases where the OFDM SINR is strong, but the BLER is high, which means cells with bad BLER are selected. This leads to handover failures and connection failures.

Interference also reduces the data rate, even in static cases. Our empirical study of commercial networks shows that operators only use a limited number of frequency channels[1] for a frequency band. For example, AT&T only uses two EARFCNs 850 (1.955GHz) / 976 (1.967GHz) in west LA. This overlapping deployment scheme facilitates intra-frequency measurements. Operators deploy intra-frequency overlapping cells ubiquitously to guarantee coverage even when the device only supports one frequency band. For drones, the problem is more severe due to less blockage and side-lobe based propagation. The interference gets high upon less attenuation of interference. Besides, ground-based tests are ineffective in preventing interference in the air. High interference causes a high packet error rate and thus a low data rate [75, 103].

We design CrOss-Domain Interference Management (`CODIM`) for next-generation cellular networks. We find that the synergy between delay-Doppler domain and the time-frequency domain eliminates interference and profiles interference with a new transformation. We exploit the transformation for accurate channel estimation and interference cancellation.

4.2 Problem Statement and Challenges

4.2.1 Physical Layer Basics

Here we introduce the background of physical layer design in 4G/5G. We first introduce the frame structure. We focus on the channel quality measurements and data demodulation. Both rely on the characterization of the propagation channel.

Frame structure. 3GPP standards [12, 13, 14] adopts OFDM as the modulation scheme. OFDM multiplex symbols at time-frequency domain. The multiplexing relies on quantized resources by time slots and subcarriers. Each resource element is defined at a coordinate of slot and subcarrier index. There are two types of symbols: signaling and data. Both signaling

[1]Frequency channel is identified by EARFCN in LTE, NR ARFCN in 5G

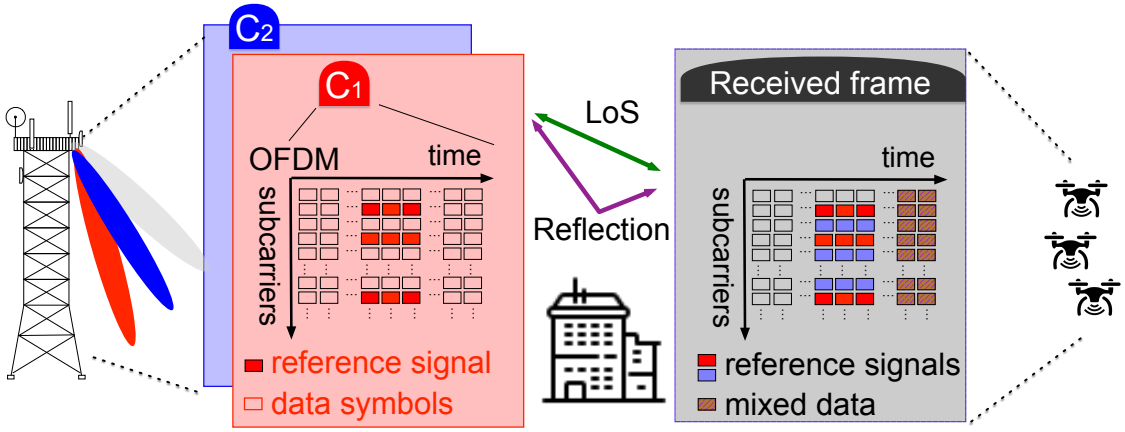


Figure 4.1: Illustration of OFDM frame structure.

and data are scheduled into a set of resource elements. The physical layer finally sends the multiplexed signal at the allocated time and frequency. The key signaling symbols are reference signals. In 5G NR, there are three types of signaling at downlink, SSB blocks, CSI-RS and data demodulation reference signal. These signaling blocks are scheduled with either a subgrid or a set of subgrids in the time-frequency domain. Besides signaling, data symbols are also scheduled based on user traffic. As shown in figure 4.1, a cell tower typically schedules reference signals from different cells with different resources. However, data symbols can collide, resulting in mixed data symbols.

Channel measurement. A key step for the wireless-based mobility management is to measure each cell's channel quality. This requires channel measurements at the physical layer. As shown in figure 4.1, the cell tower schedules each cell's reference signal at designated grids. The device is configured with the specific grid locations by the cell and then starts periodic or on-demand measurements. The device is usually able to measure all the cells on one frequency channel. For example, the device can tune to band 2 and measure all cells with frequency 1930 - 1990MHz. To measure the cells outside of this band, the device has to disconnect with the current serving cell and finish measurement within a period (usually 40ms-80ms [16]). Thus, the common practice is to reduce the measurement on inter-frequency cells located outside the current serving frequency.

In 4G/5G, such measurements are performed by the client at the physical layer in the time-frequency domain, based on the input-output relationship: $\mathbf{Y} = \mathbf{HX} + \mathbf{n}$, where $\mathbf{X}$ is the transmitted signal vector, $\mathbf{Y}$ is the received signal vector, $\mathbf{H}$ is the channel impulse matrix and $\mathbf{n}$ is the noises. By placing the reference signal with pre-defined values X, the receiver can estimate H and n based on received value Y and pre-defined X, and then estimate received signal power (RSRP), quality (RSRQ) and signal-noise-ratio (SNR) for the handover decisions and cell selection under carrier aggregation[87, 9]. These physical measurements are used for control plane procedures.

Data demodulation. Data demodulation relies on the transmission of demodulation reference signals and data symbols. Data symbols can be control messages or user messages. The device is configured with the resource allocated for each cell. As the client moves, the channel will also vary over time. How fast a channel evolves with mobility depends on how it is represented. In OFDM, the channel response $H(t, f)$ remains approximately invariant in a very short duration $T_c \propto \frac{1}{\nu_{max}} \propto \frac{c}{vf}$ due to multipath fading [128], where T_c is the channel coherence time and $\nu_{max} \propto \frac{vf}{c}$ is the maximum Doppler frequency shift, v is client movement speed and c is light speed. So a client under faster movements or higher carrier frequency (e.g., mmWave) will experience more dramatic channel dynamics in OFDM (<1ms under 500km/h and 3.6GHz band).

4.2.2 Problem Statement

We formulate the problem of interference estimation in the context of the cellular network for drones. To illustrate the problem, we first show how the interference happens. The client is receiving the reference signal from cell 1. An interfering drone is transmitting data to cell 2. Since there is no blockage, the data transmission in the air causes interference to the reception of the reference signal. Besides, the uplink-intensive applications for drones intensify the interference. Downlink interference can be generated in a similar setting as the client is receiving the reference signal from cell 1. An interfering cell 2 is transmitting data

to another drone. The interference will cause issues with reference signal measurements and data decoding.

The problem here is how to estimate the channel response for a reference signal for channel quality metrics. We categorize interference into intra-cell interference and inter-cell ones based on where the interference comes from. Based on current practice, even in the case of carrier aggregation [84, 99], a single cell will not multiplex users at the same resource grid, thus orthogonality between transmission inside a single cell is guaranteed. Nevertheless, inter-cell interference can be devastating. The interference might come from the reference signals or data signals from other cells. Here we use data signals to denote all signals, including both control plane and user plane traffic.

System settings Interference cancellation generally requires cooperation between cells with the backhaul network. We assume that the cells can interchange configurations and decode data via the backhaul network. Note that this requirement is satisfied by 3GPP standards. The backhaul network is implemented as X2/Xn interface [18]. Besides, we require the cells remain synchronized at the slot level. All 3GPP standard-compliant cells shall satisfy the 3 μs synchronization requirement [11].

4.2.3 Challenges

4.2.3.1 Challenges of Interference Estimation

Interference from other clients can have an arbitrary channel response matrix and signal structure. Thus it is inherently impossible to infer both matrices perfectly. From Channel theory [119], we know that the possibility of decoding the signal depends on the SINR. However, even when SINR is high, the decoded signal does not reveal the channel response matrix. For example, we have $y_recv = H_1 * x_1 + H_2 * x_2$, supposing x_1 is the reference signal, which is already known from broadcasting configurations. We can only decode the value of H_2 with accurate estimation of H_1 and x_2.

The current interference handling is deficient in differentiating the reference signal and data signal. In current cellular networks, reference signals are usually scheduled apart for different cells. Such a scheme is effective in preventing interference between reference signals. However, this is not sufficient to prevent interference between reference signals and data signals. Although cells can schedule data signals separately from reference signals to prevent interference, it either requires the cell to exchange real-time scheduling information or schedules persistent allocated resources for reference signals. It is thus difficult to strive for a balance between the trade-off communication overhead and spectrum efficiency.

Therefore, `CODIM` shifts the intra-domain interference to inter-domain interference. Intuitively, by scheduling signals at different domains, interference can be minimized. However, the signals will take the same resource elements, where orthogonality does not hold. There are three main challenges to solve this problem.

◦ *C1: Characterizing the frequency-selective channel.* Channel in the time-frequency domain exhibits frequency-selective fading. Such a feature improved OFDM's efficiency since data transmission with OFDM leverages frequency diversity to adaptively modulate data symbols [127]. It's a common practice for the cell to schedule a higher coding rate for devices with lower fading. However, such frequency-selective scheduling requires the channel estimation performed for every single sub-carrier. The frequency-selective fading implies that the OFDM reference signal must be spread over the entire frequency band to capture the fading coefficient at each sub-carrier.

◦ *C2: Differentiating reference signals.* OFDM symbols only contain phase and amplitude, which reveals no information on the transmit signal power. Thus, reference signals can only differentiate themselves based on the transmission location. In such a setting, each cell occupies all sub-carriers of a symbol duration to transmit the reference signal. All other cells cannot transmit data signals at that symbol duration. This results in two limitations: scalability of overlapping cells and waste of resources.

◦ *C3: Reciprocity does not hold for interference.* Downlink interference can be measured.

However, uplink interference cannot. And the common approach of leveraging reciprocity to transform uplink and downlink channel characteristics does not hold. The uplink interference must be measured to be estimated and analyzed. But the cellular network does not support the device to send any signal for the cell to measure uplink. This is a negative outcome due to hard handover. What's worse, the device cannot send signals without disrupting data transmission. Since each cell operates at different channels, the device has to switch to a different frequency to send signals, which disrupts data. Besides, it might cause potential security loopholes for the cell to accept and analyze the signal from a non-connected device.

4.2.3.2 Challenges of Interference Cancellation

Interference cancellation is an attractive technique to boost performance even under high interference. The intuition is that by canceling strong interference sources, the client can decode the remaining meaningful signal with better accuracy. To put it simply, we have $y_{recv} = H_i * x_i + H_0 * x_0$, where x_i is the interference signal and x_0 is the intended signal. After decoding x_i, the client can cancel $H_i * x_i$ and decode x_0. There has been extensive work in this domain [57, 107]. However, there are two main challenges as discovered in previous works.

First, the success of cancellation relies on the assumption of perfect channel estimation. If the estimated channel response is erroneous, the client is not able to cancel interference accurately even with the decoded signal. For example, with a simple two-client case, we have the following equations to denote the remaining signal.

$$\hat{y_0} = y_{recv} - \hat{H}_i * x_i = H_i * x_i - \hat{H}_i * x_i + H_0 * x_0$$

As shown in the equation, the remaining interference depends on the accuracy of estimated channel response H_i. In the worse case, even with accurate decoding of interference signal, the interference can even be higher after cancellation. Although the worse case is not common

to happen, for realistic cases, the remaining interference can reduce the gain greatly.

Second, the gain of interference cancellation depends on the discrepancy between the power of clients. If two clients are transmitting with similar power, the transmitting rate is largely restricted by the capacity of decoding the interfering signal. We can derive the gain of successive interference as the following in the two-client model.

$$R_{legacy} = log(\frac{H_i * x_i}{H_0 * x_0}), R_{sic} = log(\frac{H_i * x_i}{H_0 * x_0}) + log(\frac{H_0 * x_0}{H_i * x_i - \hat{H}_i * x_i})$$

By comparing the legacy capacity rate and SIC capacity rate, the gain will increase as the stronger interferer gets stronger. However, the real-time coordination of power between clients is challenging.

Finally, interference cancellation does not fit in the drone cases where there can be multiple interferers. Due to lack of blockage, the drones can receive the interfering signals of even remote cells. With interference from different cells, the legacy cancellation approach fails to perform accurate cancellation successively for each interferer. For the multiple interferer case, the problem of channel estimation error and power differentiation remains. Besides, the problem is even more challenging upon the interference between OTFS reference signals. The intra-domain interference is harder to cancel than the cross-domain interference.

4.3 Cross-Domain Interference Estimation

4.3.1 Insights from Delay-Doppler Domain

The delay-Doppler domain unveils client movement and multi-path propagation. Compared with the time-frequency domain, it offers a more direct representation of propagation paths. Besides, the delay-Doppler domain channel is resistant to frequency-selective fading. However, the direct application of the delay-Doppler domain affects the data transmission efficiency since the beamforming and MIMO technology with OTFS is rather premature.

We propose to multiplex reference signals and data at both domains to break the trade-

off of choosing the single domain. Our key insight is that, with reference signals at the delay-Doppler domain and data transmission at the time-frequency domain, we can achieve much more than a single domain. We highlight the following insights.

Insight 1. Cross-domain power differentiation. The co-existence of OTFS and OFDM provides natural power control. When the OTFS signal is transmitted via the OFDM modulation, the power of the signal decreases since each symbol is spread among the entire OFDM plane. If we are transmitting a signal of the same strength at the OTFS domain and OFDM domain, the OTFS-generated signal in the time domain can be 10dB lower than the OFDM-generated signal.

Insight 2. Cross-domain non-orthogonal multiplexing. We also find that non-orthogonal multiplexing of OTFS and OFDM convert interference into beneficial signals. Although the OTFS and OFDM generate overlapping waveforms, the processing at a different domain separates the signals apart. The client can decode both the reference signal and the interfering data. The decoded data can be further canceled to improve transmission efficiency.

Insight 3. Exploiting the synergy of both domains. With signals separated by two domains, we achieve inherent differentiation of reference signal. At the delay-Doppler domain, the reference signal goes through a frequency-inselective channel, thus the frequency-selective fading is no longer a problem. However, the transformation between OTFS and OFDM offers us the channel quality in both domains.

Besides, we observe the following features of propagation channels for drones.

Besides, we observe the following features of propagation channels for drones.

◦ *Sparsity of paths.* One unique feature of communication in the air is the sparsity of paths. The sparsity of propagation paths is shown in many previous studies [66, 126, 42].

The sparsity of paths leads to a sparse channel response matrix H in the delay-Doppler domain. This simplifies the problem of estimating H given fewer unknown variables. Since a known element $H(k,l)$ implies a known $Y(k,l)$, our target is simplified to infer a number of unknown elements. With delay-Doppler domain representation, the complexity of frequency selective fading is avoided. In the time-frequency domain, the frequency-selective fading can only be captured with a full-rank channel matrix.

◦ *Limited Doppler.* The channel response matrix is decided by the scale of Doppler, which is dictated by the relatively moving speed between the sender and the receiver. Due to the restriction of drone moving speed[2], Doppler is limited. We further find that the value of H reaches the peak at the index corresponding to the Doppler scale. Since Doppler is limited, the value of H is primarily decided by the first Doppler index given the small time span of the reference signal. In a realistic setting of using 2 slots to transmit reference signal, the Doppler tap is $\frac{1}{N*T} = 7.5$ kHz. With max speed as 100km/h, the Doppler is vf/c.

◦ *Linear Transformation between OTFS and OFDM.* While cross-domain separation seems to be practical, which domain to perform channel estimation remains a question. No matter from which domain, we will observe the mixed signals generated from both domains. To differentiate these two, we derive a frequency-inselective approach leveraging the sparsity of delay Doppler signals. Note that Time-frequency domain symbols are spread over the delay-Doppler domain. We can recover some of Y with known y. Note that this is impossible if both signals are sent in the time-frequency domain since each element is independent.

4.3.2 X-Domain Estimation for Single Path

In this section, we begin with the formal formulation of the interference estimation problem, then use an illustrative example to show how our design works at a high level. Finally, we

[2]According to FAA regulation [47], the maximum allowed speed for commercial drones is 100km/h.

will elaborate on our algorithms step by step.

Problem formulation. We first formulate our problem formally as an optimization problem. `CODIM` enables the co-existence of OTFS domain reference signal Y and OFDM domain data signals y. The OTFS reference signal $Y[k,l]$ is located at k th delay tap and l th Doppler tap. The OFDM data signal $y[m,n]$ is located at m th subcarrier and l th symbol. The receiver receives $Y_{recv} = Y_{RS} + Y_{interf}$, where Y_{RS} is the reference signal and Y_{interf} is the data signals from other cells. Our target is to estimate Y_{RS}. We can further derive H_{RS} since $Y_{RS} = H_{RS} * X_{RS}$.

Formulation of the transformation. A key step of our algorithm relies on the transformation between the delay-Doppler domain and the time-frequency domain. We use matrix $\mathbf{S}$ to capture the transformation since it is linear. The transformation from $\vec{y}$ to $\vec{Y}$ is linear. We denote $\vec{y}$ and $\vec{Y}$ as the vectorization of $y[m,n]$ and $Y[k,l]$. We have: $\vec{Y} = \mathbf{S}\vec{y}$, where $\mathbf{S}$ is a $MN * MN$ matrix. Based on ISFFT, $\mathbf{S}[idx(k,l), idx(m,n)] = e^{-j2\pi(\frac{mk}{M} - \frac{nl}{N})}$, where $idx(k,l)$ is the function to get the index after vectorization with $idx(k,l) = lM + k$. For example, we have $Y[k,l] = \begin{bmatrix} 1 & ... & e^{-j2\pi(\frac{mk}{M} - \frac{nl}{N})} & ... & e^{-j2\pi(\frac{(M-1)k}{M} - \frac{(N-1)l}{N})} \end{bmatrix} \vec{y}$. Note that $\mathbf{S}$ is full-rank. For simplicity, we use $\mathbf{S}_{\mathbb{R}} = \{\mathbf{S}[idx(k,l),:]\}_{(k,l)\in\mathbb{R}, 0\leq:<MN}$ to represent the row vectors to transform $\vec{y}$ to $Y_{\mathbb{R}}$, where $\mathbb{R}$ denotes the indexes of selected $\vec{Y}$. In general, We have $\vec{Y}_{\mathbb{R}} = \mathbf{S}_{\mathbb{R}}\vec{y}$.

An illustrative example. We first show an illustrative example on how we solve the unknown Y_{RS} and Y_{interf}. In this example, we assume that there is only a single propagation path and a single interferer. In such a case, we have $Y_{RS} = H_{RS} * X_{RS}$, where H_{RS} is a matrix with a single non-zero element at $[0,0]$. For simplicity, we assume X_{RS} is a matrix with a single one at $[0,0]$. Our target is to infer the value of $H_{RS}[0,0]$. Equivalent, we can derive our target from either $Y_{RS}[0,0]$ or $Y_{interf}[0,0]$. We will next show how $Y_{interf}[0,0]$ is derived from cross-domain transformation.

Different from legacy channel estimation where we only infer the channel from the features

Algorithm 2 Estimating interference for a single path

Require: Received signal Y_{recv}, Reference signal X_{RS}, delay-Doppler grid setting $N, M, \Delta\upsilon, \Delta\tau$

Ensure: Channel response $\mathbf{H}$

1: Initialize transformation matrix $\mathbf{S}$ with Equation 3.3;

2: Update transformation matrix $\mathbf{S}_{\mathbb{R}^*\setminus(0,0),\mathbb{C}^*\setminus(m,n)}$ by removing the column corresponding to element $[m, n]$;

3: Get $Y_{interf} \leftarrow \mathbf{S}_{[0,0]} inv(\mathbf{S}_{\mathbb{R}^*\setminus(0,0),\mathbb{C}^*\setminus(m,n)}) \vec{Y}^i_{\mathbb{R}^*\setminus(0,0)}$ according to Equation 4.1;

4: $Y_{RS} \leftarrow Y_{recv} - Y_{interf}$;

5: Calculate $\mathbf{H}$ based on X_{RS} and Y_{RS} with $\mathbf{H} = Y_{RS}/X_{RS}$;

6: Derive time-frequency domain signal y_{RS}

7: Calculate SNR/RSRP/RSRQ from signal y_{RS} and interference y_{interf};

of the channel, we actually infer the interference since the interference is more structured. We now need to solve unknown $Y_{interf}[0, 0]$. We will first show how it is feasible with the knowledge of $y_{interf}[M-1, N-1] = 0$. Later, we will remove this assumption and prove it works in general cases.

We use y or Y instead of y_{interf} and Y_{interf} since the following proof works for general OFDM signals. Note that when $k = l = 0$, we have $Y[0, 0] = \begin{bmatrix} 1 & ... & 1 & ... & 1 \end{bmatrix} \vec{y}$. $S_{\mathbb{R}=\{(0,0)\}}$ as a all-ones vector with length MN. We use $\mathbb{R}^*$ to represent all row vectors. We have $S_{\mathbb{R}^*\setminus(0,0)}$ to denote the $MN - 1 * MN$ transformation matrix to transform $\vec{y}$ to $Y_{\mathbb{R}^*\setminus(0,0)}$. Since we know $Y_{\mathbb{R}^*\setminus(0,0)}$, if $Y[0, 0]$ is a linear combination of $Y_{\mathbb{R}^*\setminus(0,0)}$, we can get $Y[0, 0]$. Note that $S_{\mathbb{R}^*\setminus(0,0)}$ is a full rank matrix if we delete any column (proven in Lemma 4.3.4). If the removed columns constitute a full-rank submatrix.

Estimating a single path. We devise Algorithm 2 to estimate interference in a single path scenario. This is to represent the case where the transmitter and receiver have LoS (Line-of-Sight) propagation path. The single path makes sense for drones since the propa-

gation is LoS, especially between drones. The limitation of Doppler also holds due to the drone's speed limitation.

The algorithm requires the received signals at the time-frequency domain. This input can be guaranteed with legacy 4G/5G support. With received signals, the client performs transformations between the time-frequency domain and the delay-Doppler domain. By imposing the restriction of sending the signal, the client is able to recover the signals at both domains. Essentially, it decouples the signals from two sources by enforcing both sides transmit with certain rules.

The algorithm works in three steps. The client first transforms time-frequency domain signals y to delay-Doppler domain signals Y. Note that the signals at both domains are mixed. After getting signals at two domains, the client removes noisy Y and gets a clean estimation of y according to the cross-domain transformation. Finally, the client infers missing Y based on estimated y. In the final step, the estimated y is complemented with the prior information on the transmitted signals.

We next show how the transformation works in rigorous proofs. Note that in step 2, the client gets the noisy Y. However, the Y is only noisy at the elements with the reference signal. Due to the sparsity of reference signals, we can eliminate these noise elements. The key idea here is that even with incomplete y, we can recover complete Y with Equation 4.1.

$$Y^i[0,0] = \mathbf{S}_{[0,0]} inv(\mathbf{S}_{\mathbb{R}^*\setminus(0,0),\mathbb{C}^*\setminus(m,n)}) \vec{Y}^i_{\mathbb{R}^*\setminus(0,0)} \tag{4.1}$$

We rely on Theorem 4.3.1 to recover Y. We show the detailed proofs in Appendix B.1.

Theorem 4.3.1 (Solvability under single path model)**.** *Under the single path propagation model,* $Y[0,0]$ *can be derived with the knowledge of any* $y[m,n]$*.*

Theorem 4.3.1 requires the client to know at least one element of any $y[m,n]$. There are two realistic conditions that make these possible. First, the sender can keep $y[m,n] = 0$ as a guard symbol. This can easily hold since the first several slots are reserved for control

signaling. In 4G/5G, it already holds that different cells use specific slots. This is true in current practice. We are only required to spare one slot. Even for a 600 * 2 grid, the condition still holds. Especially for TDD mode, the drone only sends data in the uplink. The downlink slots are empty most of the time. The second condition is that $y[m,n] = y[i,j]$, which is practical when the channel matrix H does not change during coherence time, and the transmitted signal X is the same. We find that this condition is easy to satisfy for uplink due to the existence of padding data. When the client is assigned more resource elements than the buffer size, it will fill up the remaining elements with zeros or ones. As long as the placeholders are the same, we can use them as prior knowledge for interference estimation.

Error due to noises One possible concern is how the noise affects estimation accuracy. If the received signal is not zero or the equivalence is not guaranteed, the recovery can still work with controlled errors. Theorem 4.3.2 shows that the error is constrained by the error of the assumed-to-know element in the time-frequency domain. Given that the common noise level is tens of dB lower than the signal, the error is controlled. The detailed proof is shown in Appendix B.2.

Theorem 4.3.2 (Approximation under single path model)**.** *Under the single path propagation model,* $Y[0,0]$ *can be derived from* Y_{recv} *with an error* $M * N * y[m,n]$*. The resulting SINR error is linear to* $y[m,n]$*.*

4.3.3 Multiple Paths and Multiple Interferers

For drones, the propagation paths can be more than one when there are reflecting paths. Besides, there can be multiple sources of interference since the drone is able to receive signals from many cells. We next show how our solution deals with multiple paths and multiple interferers.

Challenges and insights. There are two challenges. First, given there are more paths, there are more unknown path variables. For example, the channel response might have

several peaks. Second, the path variables are in unknown delay indexes. Although different paths can be decoupled by the delay indexes, the specific index of each path is unknown. This implies the number of unknown variables is decided by the maximum delay.

We make two observations to simplify the problem. First, the sparsity of the path guarantees a limited number of unknown path variables. The number of paths is equivalent to the number of unknowns. With more unknowns, we need more prior knowledge regarding time-frequency domain interference. The next observation is the limited delay in the air. The major in-the-air indirect path would be caused by ground or building reflection. Measurements of drone networks [55] profile the max delay for drones as 500 ns delay, which is equivalent to a distance of 150m. With the setting of M=600, N=2, Δf=15kHz, Delay tap is 100ns. Thus the unknown delay indexes will be 5 at maximum.

Estimating multiple paths. Multi-pathing increases the number of unknown channel response variables. To reveal them, we devise the recovery based on the nice features of the transformation matrix $\mathbf{S}$. We prove that P paths are recoverable if we know P of vector $\vec{y}$. The idea is that P paths corresponds to the first P elements for $\vec{Y}[p,0]$, $p = 0, ..., P-1$. Theorem 4.3.3 backups our idea. The detailed proof is shown in Appendix B.3.

Theorem 4.3.3. *$\vec{Y}_{\mathbb{P}}$ can be inferred with the knowledge of any P unique elements from $\vec{y}[p,i]$ for slot i.*

Linearity of Multiple interferers. We discover that the interference from multiple sources is linearly aggregated at the receiver. Since we recover the unknown channel response variables based on linear transformation, the linearity of interference guarantees the applicability even under more than one interferer. For a single path, the derivation is a known linear transformation, which is decided by the M/N grid setting only. We want to infer $Y[0,0]$. However, there are I interferers. We need to deduct the sum of interferer at $Y[0,0]$. Given F interferer, we have $Y[0,0] = Y_{recv}[0,0] - \sum_f^F Y^f[0,0]$. Thus we have $\sum_f^F Y^f[0,0] = \sum_f^F \mathbf{S}_{[0,0]} inv(\mathbf{S}_{known}) \vec{Y}^f_{known}$. The equation holds when $\mathbf{S}$ for all the interfering

cells are the same. Since S are the same for all interferers, we have

$$\sum_{i}^{I} Y^{i}[0,0] = \mathbf{S}_{[0,0]} inv(\mathbf{S}_{\mathbb{R}^*\setminus(0,0),\mathbb{C}^*\setminus(m,n)}) \sum_{i}^{I} \vec{Y}^{i}_{\mathbb{R}^*\setminus(0,0)}$$

For the case of multi-pathing, we want to infer $Y[p,0]$, However, there are I interferer. We need to deduct the sum of interferer's energy $Y[p,0]$. . We have

$$\sum_{i}^{I} Y^{i}_{\mathbb{P}} = \mathbf{S}_{\mathbb{P}} inv(\mathbf{S}_{\mathbb{R}^*\setminus\mathbb{P},\mathbb{C}^*\setminus\mathbb{K}}) \sum_{i}^{I} \vec{Y}^{i}_{\mathbb{R}^*\setminus\mathbb{P}} \tag{4.2}$$

, where $\mathbb{P}$ denotes unknown delay-Doppler domain indexes and $\mathbb{K}$ denotes known time-frequency domain indexes. Note that we also require that all senders have the same M/N grid setting. This requirement is easy to satisfy in the context of cellular networks. For cells interfering with each other, they are operating on the same frequency since the operators only deploy on certain frequencies. Thus the M/N grid setting is the same. The second requirement is that all the cells operating at the same area are synchronized and all follow the prior knowledge of y. Given all interferers at the same frequency channel are slot level synchronized [11], all the cells operating at the same area will sync. The operator just needs to send configurations to each cell to control the transmission. The detailed proof is shown in Appendix B.4.

Theorem 4.3.4. *$\vec{Y}_{\mathbb{P}}$ can be inferred with the knowledge of any P unique elements from $\vec{y}[p,i]$ of all interferers for slot i if all senders holds the same grid setting.*

Step-by-step algorithm. Algorithm 3 shows how the client estimate interference with the received signals at the time-frequency domain. With received signals, the client performs transformations between the time-frequency domain y and delay-Doppler domain Y. The main difference is how the client removes noisy Y and gets a clean estimation of y according to the cross-domain transformation. Different from the single path algorithm, the transformation matrix S is derived by removing all the columns corresponding to the known time-frequency domain signal. With estimated missing y, the client estimates the interference at the time-frequency domain and reference signal at the delay-Doppler domain.

Algorithm 3 Estimating interference for multipathing

Require: Received signal Y_{recv}, Reference signal X_{RS}, delay-Doppler grid setting $N, M, \Delta v, \Delta \tau$

Ensure: Channel response $\mathbf{H}$

1: Initialize transformation matrix $\mathbf{S}$ with Equation 3.3;
2: **for** each known $[m, n]$ from $\mathbb{K}$ and unknown $[k, l]$ from $\mathbb{P}$ **do**
3: Update transformation matrix $\mathbf{S}_{\mathbb{R}^*\setminus\mathbb{P},\mathbb{C}^*\setminus\mathbb{K}}$ by removing the rows and columns;
4: **end for**
5: Get $Y_{interf} \leftarrow \mathbf{S}_{\mathbb{P}} inv(\mathbf{S}_{\mathbb{R}^*\setminus\mathbb{P},\mathbb{C}^*\setminus\mathbb{K}}) \sum_i^I \vec{Y}^i_{\mathbb{R}^*\setminus\mathbb{P}}$ according to Equation 4.2;
6: $Y_{RS} \leftarrow Y_{recv} - Y_{interf}$ for unknown indexes $\mathbb{P}$;
7: Calculate $\mathbf{H}$, y_{RS} and SNR/RSRP/RSRQ;

Transformation back to OFDM. The paradigm of cross-domain transmission is promising. But one headache is how to transform the channel response derived from the delay-Doppler domain to the time-frequency domain. This is the final step to derive the channel quality metrics like RSRP, RSRQ, or SINR. Current 5G networks support all three metrics [16]. The accuracy of the metrics is paramount to select the best cells during handover and even initial connection setup. How to transform the derived channel matrix to get SINR? This is trivial since we already separated the received signals and interferences. Then the derivation of RSRP and RSRQ is obvious. We calculate RSRP by the power of the received signal. For RSRQ, we divide the RSRP by the average amplitude of interference and noise.

Overhead analysis. We require all the senders follow the rules to send guard symbols for channel estimation. The overhead of resources depends on

$$P = max\ \{number\ of\ path,\ number\ of\ delay\ indexes\}.$$

Note that the number of paths is restricted. It can be 14 paths as specified by the maximum number of paths by 3GPP [7], The challenge is the number of delay indexes. For mmWave bands, the delay index would be more fine-grained. For example, with $\Delta f = 240$kHz, the

delay tap is around 10 ns, which is equivalent to 3 m. For paths with 150 m distance, the max delay index could be up to 50. When P is 50, the overhead is $50/1200 = 4.2\%$ (assuming M=600, N=2). We can further reduce the overhead for mmWave scenarios at the loss of accuracy. For mmWave, path propagation loss increases as the distance increases. Energy will be concentrated on paths with smaller delays. The 100 ns delay spread covers most path energy. Recent measurements [36] show 40-60 ns delay spread dominates.

4.4 Cross-Domain Interference Cancellation

4.4.1 Insights

The main challenge for interference cancellation is how to cancel interference from meaningful data. With cross-domain interference estimation, the client can derive the interference for the slots of the reference signal. However, canceling interference is non-trivial. It requires the client to not only estimate the interference but also derive the interference for other time slots. The following insights enable the precise derivation of OTFS reference signals and interference over the shared grid.

Insight 1. Cross-domain scheduling. The co-scheduling of the time-frequency and delay-Doppler domain signals facilitates interference cancellation. A cell needs to differentiate the interference from other cells to cancel it. Facilitated by the accurate channel estimation, the cell can not only infer the channel response of the reference signal, but also recover the data from the interference. Note that the interference is derived from data channel response and data. After data is decoded, the cell can infer the data channel response. Co-scheduling also enables intelligent power control. Even when two clients adjust OFDM sending power independently, the received OTFS sending power is different. With such natural power differentiation, the sender can better leverage channel diversity for a higher data rate.

Insight 2. Cooperative decoding with backhaul. Base stations utilize backhaul networks to exchange UE traffic and control information. Our solution utilizes an inter-BS backhaul network for cooperative decoding. The cooperative decoding facilitates two cells to decode mixed signals with cancellation. Note that cancellation relies on the accurate estimation of channel response and decoded interference (data signal). The estimated channel response H for reference signals still applies for the entire slot. The decoded interference can be sent to all overlapping cells. For a single cell, it can try to decode the signal from itself or other cells. After decoding the strongest signal, it cancels it with the channel response and then proceeds to decode other data signals.

Insight 3. Slot-level synchronization with cloud infrastructure. Synchronization enables slot-level coherence. Per required by 3GPP standards [10], cells operating at the same frequency band are synchronized. The synchronization is realized with the high-speed backhaul network and precise time synchronization protocols. Benefiting from synchronization, the coherence time is guaranteed to cover a slot (14 symbols). Therefore, the channel response matrix H for data and reference signal are the same within coherence time (2 slots). For example, when F = 3GHz and v = 160 km/h, the Doppler shift is 450Hz. We can derive the coherence time as T $1/\nu$ (2.2 ms), while slot time is 1 ms. This implies, for a single slot, the cell is able to infer the channel response matrix for all other cells sending at the shared slot.

4.4.2 Design

We first use an illustrative example to show the intuition of our algorithm. We will then detail the design, including the requirements and each design component.

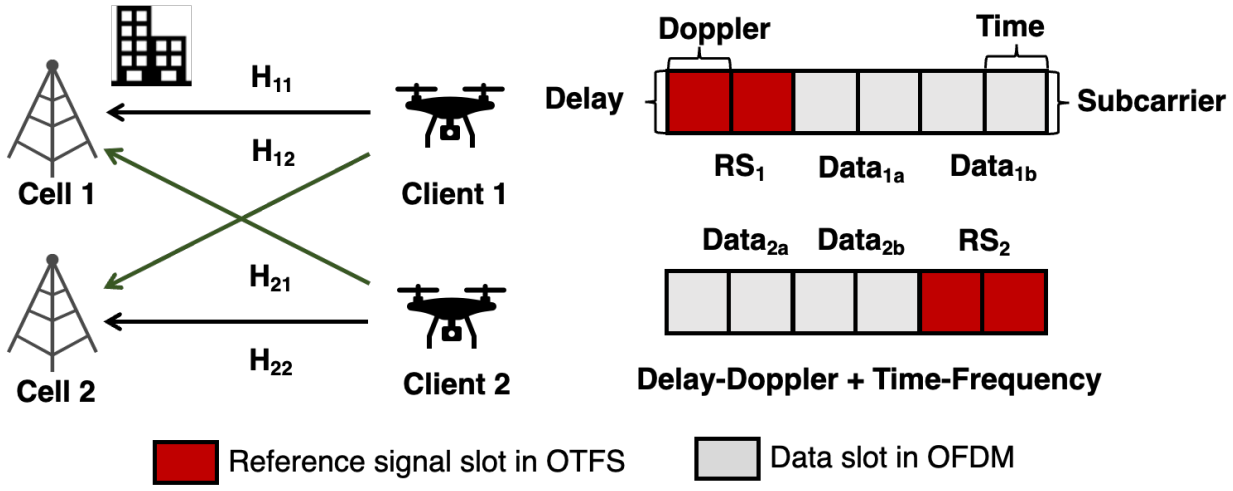


Figure 4.2: Illustration of interference between two clients.

4.4.2.1 An Illustrative Example

Figure 4.2 shows an example of two cells operating in the same area. The transmission from client 2 of cell 2 will interfere with the communication between client 1 and cell 1. Here we use H_{ij} to denote the channel response matrix between client i and cell j. If we look at the received signal at cell 1, we can formulate the following equations.

$$Y_I = H_{11} * X_1^{RS}(OTFS) + \mathbb{S} * h_{21} * x_{2a}^{Data}(OFDM) \tag{4.3}$$

$$Y_{II} = \mathbb{S} * h_{11} * x_{1a}^{Data}(OFDM) + \mathbb{S} * h_{21} * x_{2b}^{Data}(OFDM) \tag{4.4}$$

$$Y_{III} = \mathbb{S} * h_{11} * x_{1b}^{Data}(OFDM) + H_{21} * X_2^{RS}(OTFS) \tag{4.5}$$

We categorize two types of interference: cross-domain interference as formulated by Equation 4.3 and Equation 4.5, and intra-OFDM domain interference as shown in Equation 4.4. For cross-domain interference, `CODIM` guarantees that there is only one cell transmitting the reference signal at one time. Thus the receiver is able to decode the channel response from all senders. For intra-OFDM domain interference, `CODIM` utilizes the accurate channel estimation to decode data. The accuracy benefits from the cooperation between cells.

At a high level, the receiver performs three steps to cancel interference: 1) estimate

interference; 2) decode the strongest data and cancel it; 3) iterate to decode the remaining signal. In this example, the receiver first estimated H_{11} and H_{21} at OTFS domain given X_1^{RS} and X_2^{RS}. This is achieved through Algorithm 3. Then the receiver cancels the reference signal by $H_{21} * X_2^{RS}$ and decodes x_{1b}^{Data} with h_{11}. Note that h_{11} is derived from H_{11}. As the final step, the receiver decodes x_{1a}^{Data} with H_{11} and H_{21} by two options depending on which cell first decodes the signal. The first option is to let the Cell 2 send decoded $Data2b$ to Cell 1 and Cell 1 then eliminates $\mathbb{S} * h_{21} * x_{2b}^{Data}$ and decodes x_{1a}^{Data}. The second option is to let Cell 1 decode x_{1a}^{Data} and send it to Cell 2 for cancellation. Which option is adopted depends on the SINR at both cells. The cell with higher SINR will decode the data faster and initiate cooperative decoding. Note that the decoded data from two cells will not conflict due to the ID-bonded CRC checking.

Compared with legacy OFDM-based interference cancellation, `CODIM` outperforms mainly due to accurate interference estimation. With OFDM, we might adopt similar steps, but the effective decoding is lower given inaccurate channel estimation and power control. Our solution infers accurate H_{11} and H_{21} while the OFDM-based approach cannot reach the ideal SINR with inaccurate channel estimation. Besides, our solution perfectly cancels the interference of other cells' reference signals. Thus the effective decoding rate for Data $1b$ is $\frac{H11*Data1b}{H21*RS2+Noise}$ with OFDM, while with OTFS, the rate is $\frac{H11*Data1b}{Noise}$. Finally, our solution can recover with accurate channel estimation. The cancellation will remove the signal without leaving any remaining interference.

4.4.2.2 Requirements and Design

The key problem is to accurately cancel decoded data signal interference. There are three major goals for interference cancellation. First, the intra-cell interference shall not affect the multiplexing between clients. Second, the intra-OTFS interference is minimized to mitigate reference signal estimation error. Thirds, inter-domain interference should be canceled with high accuracy. With all three goals, `CODIM` is composed of three components to mitigate and

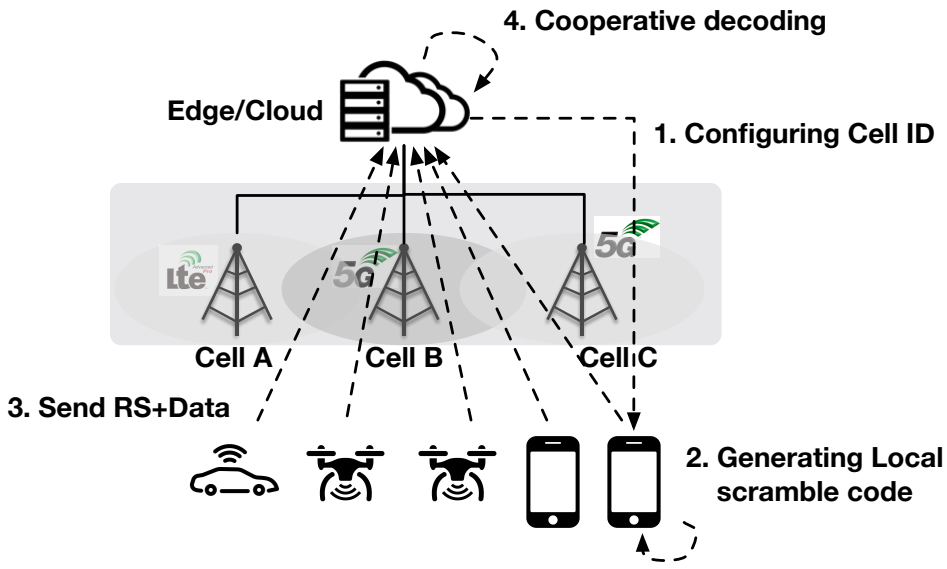


Figure 4.3: The overall design.

embrace interference. Figure 4.3 shows the overall design of CODIM.

Persistent and runtime scheduling. CODIM utilizes persistent scheduling and runtime scheduling to flexibly allocate resources while guaranteeing channel estimation accuracy. To avoid intra-OTFS-domain interference, we require that each cell reserves slots for OTFS reference signal. Such persistent scheduling places OTFS reference signals and OFDM data signals over the same grid. One requirement is to prevent interference between OTFS reference signals. Note that intra-delay-Doppler domain interference diminishes if the signals are transmitted at separate time-frequency grids. To do so, CODIM schedules reference signals apart at the time-frequency domain. Our design guarantees the flexibility of the cells to design its over reference signal inside the allocated subgrid.

One feature provided by runtime scheduling is informative modulation. For the decoded signal, it embeds the cell ID. The cell ID indicates the location of a reference signal. Thus the receiving cell can map the decoded data and the corresponding reference signal together. CODIM proposes to combine current CRC checking with ID binding. During CRC checking, the receivers get the UE-specific scrambling code. CODIM will embed cell ID into the scram-

Algorithm 4 Interference cancellation at the receiver

Require: Received signal Y_j^{recv} at cell j, Reference signal X_i^{RS} for cell i, delay-Doppler grid setting $N, M, \Delta\upsilon, \Delta\tau$

Ensure: Decoded data x_{ij}^{Data} from client i to cell j

1: Initialize transformation matrix $\mathbf{S}$ with Equation 3.3;
2: **for** each cell i with X_i^{RS} at grid G_i **do**
3: Estimate channel response $\mathbf{H}_{ij}$ via Algorithm 3;
4: Derive $Y_i^{data} \leftarrow Y_i^{recv} - \mathbf{H}_{ij} * X_i^{RS}$ for each grid;
5: Perform successive interference cancellation for $inv(\mathbf{S})Y_i^{data}$;
6: **end for**
7: Transform delay-Doppler channel response $\mathbf{H}_{ij}$ to time-frequency domain $\mathbf{h}_{ij}$
8: **for** grids G_k without other cells' reference signals **do**
9: Check each channel response $\mathbf{h}_{ij}$ for equalization to y_k^{recv};
10: Broadcast decoded data x_{ij}^{Data} to neighboring cells;
11: **end for**

bling code together with the UE ID to guarantee that decoded data can be differentiated.

Cross-domain cancellation. The key component of `CODIM` is Algorithm 4. The algorithm relies on the prior configurations of each cell's reference signals and OTFS grid setting. During the configuration, each cell is assigned with specific reference signal slots. When it sends out its data signal, the data signal embeds the cell ID. For a cell to decode data, it will check the ID after decoding it. If the ID is itself, it stops the process. Otherwise, it will keep decoding. Nonetheless, the receiver will send the decoded data to other cells for cancellation no matter who generates the data.

Note that the algorithm deals with inter-domain interference and intra-OFDM interference differently. The inter-domain interference is resolved by deriving the channel response matrix at the delay-Doppler domain. Then the receiver transforms the channel response to

the time-frequency domain. This is because the interference cancellation is performed at the time-frequency domain. For intra-OFDM interference cancellation, the receiver adopts a simplified equalization procedure that benefits from the accurate channel response. The receiver will try with each channel response to identify the strongest signal. If decoding succeeds, the receiver cancels the interference and proceeds on the remaining signal.

Gain analysis. We compare the performance with our design to legacy OFDM-based interference cancellation. The gain is different for cross-domain interference cancellation and intra-OFDM interference cancellation. For the first type, we have Equation 4.3 and Equation 4.5. The gain with OTFS is $\frac{\mathbb{S} * h_{21} * x_{2a}^{Data}}{n}$ while the legacy OFDM can only provide $\frac{\mathbb{S} * h_{21} * x_{2a}^{Data}}{H_{11} * X_1^{RS} + n}$. `CODIM` provides better gain because the receiver can differentiate data and reference signal with cross-domain transformation. Considering a realistic case with 20 dB SNR, 5 dB SIR and nominal Gaussian noise, the gain with `CODIM` is 20 dB while the gain is 5 dB with the legacy approach assuming perfect interference estimation. Even under a common 12% error, `CODIM` still provides 10 dB gain. As SIR gets lower, `CODIM` will provide higher gain compared with the legacy approach.

For intra-OFDM interference, `CODIM` provides a similar gain for decoding the strongest signal. However, for the later signals, `CODIM` provides better gain since the accurate channel estimation cancels the strongest signal. Specifically, `CODIM` provides $\frac{\mathbb{S} * h_{21} * x_{2b}^{Data}}{n}$ gain for second strongest signal (assuming 1a is stronger than 2b). We have accurate cancellation with correct h_{11}. However, the gain is only $\frac{\mathbb{S} * h_{21} * x_{2b}^{Data}}{h_{11}^{err} * x_{1a}^{Data} + n}$ with legacy OFDM. Note that even small drifting of h_{11}^{err} deteriorates the gain since the amplitude for Data 1a is much higher. For the previous setting assuming the legacy approach fails to eliminate half of the interference, `CODIM` offers 14dB gain while the legacy offers 3dB.

4.4.3 Practical Problems

Extension to multiple interferers. It is very likely that decoded time-freq signal is from multiple interferers. For example, we need to decode the mix of data from three or more cells. `CODIM` could decode the strongest signal first and then cancel it. The benefits of managing cross-domain interference canceling provide accurate channel estimation for all interferers. For cross-domain interference cancellation, the performance is the same regardless of the number of interferers. For intra-OFDM interference, performance will drop upon more interferer. However, `CODIM` still outperforms the legacy approach due to more accurate channel response estimation.

Specifically, `CODIM` provides $\frac{h_i * x_i^{Data}}{\sum_{j=i+1}^{I} h_j * x_j^{Data} + n}$ gain for i strongest signal. We have accurate cancellation with correct h_k for $\forall k < i$. Therefore, the capacity is only inverse-proportional to the remaining signal from client j after cancellation. However, the gain is only $\frac{h_i * x_i^{Data}}{\sum_{k=1}^{i-1} h_k^{err} * x_k^{Data} + \sum_{j=i+1}^{I} h_j * x_j^{Data} + n}$ with legacy OFDM. The error in channel estimation aggregate and affect the decoding of all successive packets.

Downlink interference cancellation. Downlink is more challenging than uplink since clients cannot exchange decoded data. With such constraints, each client can only perform successive interference cancellations independently. We propose that the cell configures the clients with new configurations to support cell-specific ID. Besides, the cell estimates downlink channel response and aligns interference through coordination.

Our design is backward-compatible with legacy clients without the capacity to decode advanced configurations. Although our design of uplink cancellation leverages the backhaul network between base stations, the backhaul traffic is not necessary for interference cancellation. For downlink interference cancellation, we can still keep the primary steps to decode overlapping reference signals and data signals. We expect a smaller gain but still benefit from accurate channel estimation.

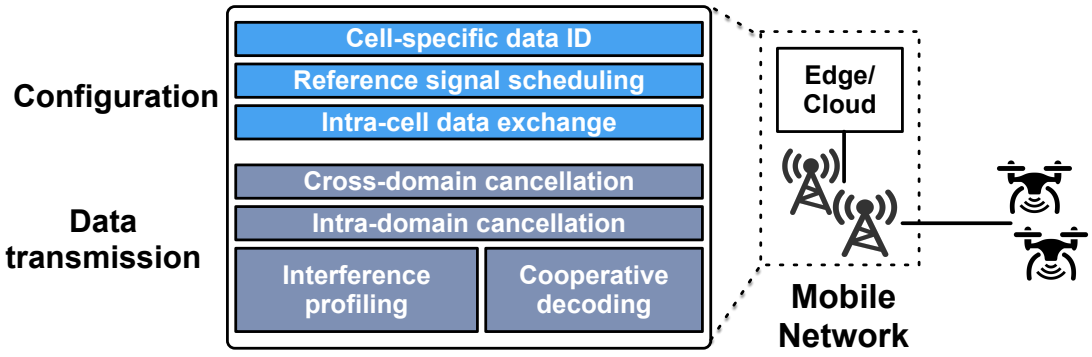


Figure 4.4: Architecture at the network side.

4.5 Evaluation

In this section, we evaluate the performance of our design with both real-world experiments and large-scale trace-driven simulations.

4.5.1 Implementation and Setup

We detail the implementation of pre-processing before our estimation algorithm and cancellation. The overall architecture is shown in Figure 4.4.

Configuration engine. Cells rely on the configuration engine to coordinate for interference management. There are three components in the configuration engine, ID manager, reference signal scheduler, and data exchange policy manager.

Data transmission. After configuration, each cell is prepared to transmit and decode data with the existence of interference. There are four key components at this stage.

- *Interference profiling.* Each cell profiles the interference with the cross-domain interference estimation algorithm 3. After each client is configured with a sounding reference signal, it sends a periodic reference signal and data demodulation reference signal. These reference signals assist with interference profiling.

- *Cooperative decoding.* The decoding Algorithm 4 is implemented at both the cell and

client side. For the cells, they exchange the decoded data via backhaul network through X2/Xn interface [17]. For the clients, they do not exchange decoded data.

◦ *Noise estimation.* We use the legacy least square scheme to estimate noise. Given Gaussian-distributed noises, the least square scheme works well. Typically, the noise level compared with signals and interference is small.

◦ *Carrier frequency offset.* We estimate frequency offset based on the synchronization signal. After detecting the synchronization signal, the client calculates the frequency offset and demodulates the received waveform. Our implementation assumes static carrier frequency offset for simplicity.

Experimental setup. To approximate real extreme mobility, we run trace-driven emulations over a USRP-based testbed. Our testbed consists of servers running OAI [1] cellular protocol stack and the USRPs as clients and base stations. The servers run OAI [1] cellular protocol stacks. We have USRP B210/N210 to test with real channels, which are connected to servers with Intel Xeon CPU E5-2420 v2 and 16GB memory. To emulate operational settings, we configure the testbed based on the above datasets. Specifically, we extract protocol configurations and mobility policies for each cell from the dataset and test with various settings. To compare `CODIM` with legacy 4G/5G, we replay the mobility traces from our datasets and evaluate if `CODIM` can prevent failures under the same settings. Note we run the USRPs under the unlicensed 2.45G band and 58G band instead of licensed ones to comply with FCC regulations.

4.5.2 Overall Performance

We evaluate our algorithm's benefits for mobility and data rate. Our algorithm provides better channel quality indicators with accurate interference profiling. We first compare the accuracy of predicted channel quality indicators and the impact on mobility support. We then evaluate the data rate and packet loss ratio with interference cancellation, comparing

CODIM and legacy OFDM.

Reduction of packet loss. We show how interference cancellation reduces packet loss at different SINR and even mobile cases. We compare our solution with legacy OFDM-based as a baseline, where we estimate the channel with the legacy approach and cancel it. For current indoor 2.45GHz setting, CODIM reduces the loss from >24% to 0.5% compared with the legacy 5G (Figure 4.5). From static to low mobility case, packet loss ratio with legacy scheme increase from 24.5% → 28.4% while CODIM is not affected by mobility (0.5% → 0.4%). Under mmWave setting (Figure 4.6), CODIM reduces the loss from 19.1% to 3.5%. Under low mobility case, CODIM reduces the loss from 36.8% to 9.4%.

We also show the results from the optimal approach (Figure 4.7). We assume the channel estimation is perfect and cancel it. The OTFS-based solution is where we adopt the designed algorithm and we estimate reference signal response and cancel it. On average, our design reduces the BER from 23.9% to 8.8%. Our solution is close to the optimal interference cancellation.

Data rate gain. We evaluate the data rate gain with CODIM compared with the legacy approach. We estimate data rata based on the SINR before and after cancellation. We compare CODIM with the legacy. We combine all the samples shown before that are collected in both indoor and outdoors. As shown in Figure 4.8, CODIM improves the data rate from 23-38Mbps to 148-244Mbps. The improvements are significant regardless of the moving speed of the clients.

4.5.3 Benefits for Applications

How can CODIM benefit real applications with enhanced mobility and efficiency? In this section, we evaluate the performance improvement for emerging applications with stringent latency requirements like AR/VR.

Experimental setup. To evaluate the performance of mobile VR/AR applications, we

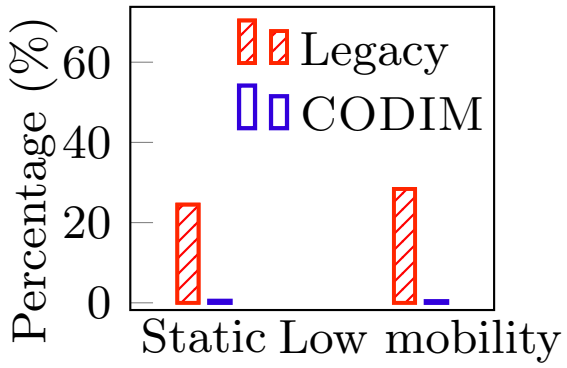


Figure 4.5: Packet loss ratio for Sub-6GHz.

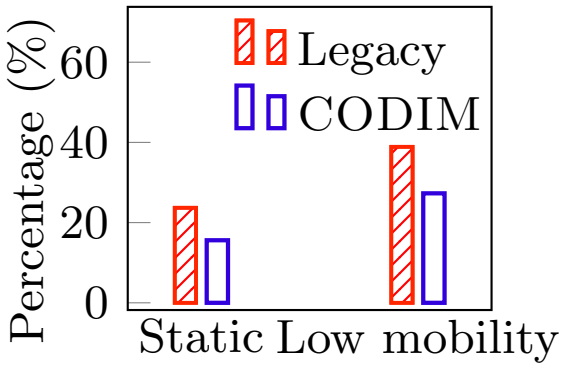


Figure 4.6: Packet loss ratio for mmWave.

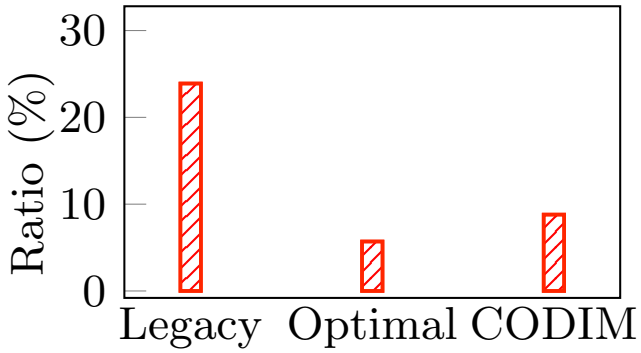


Figure 4.7: Comparison to optimal.

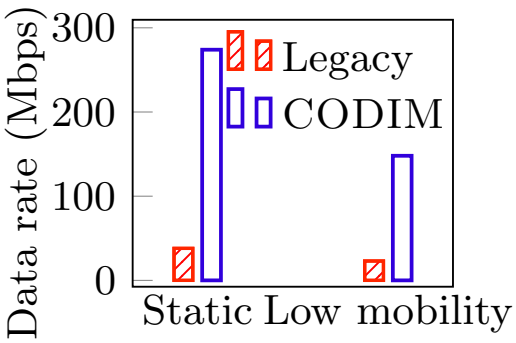
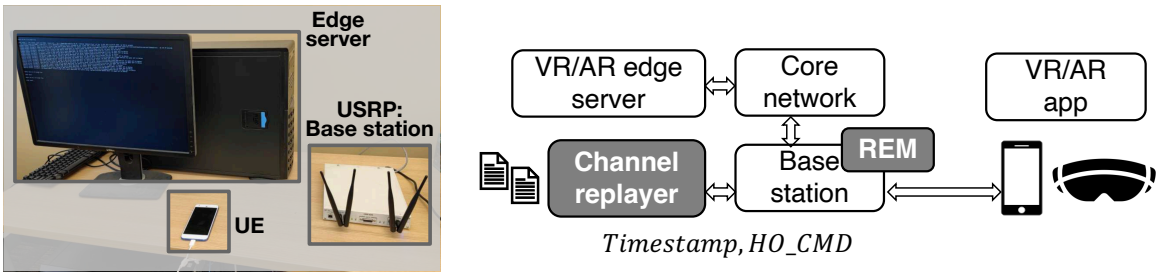


Figure 4.8: Data rate.

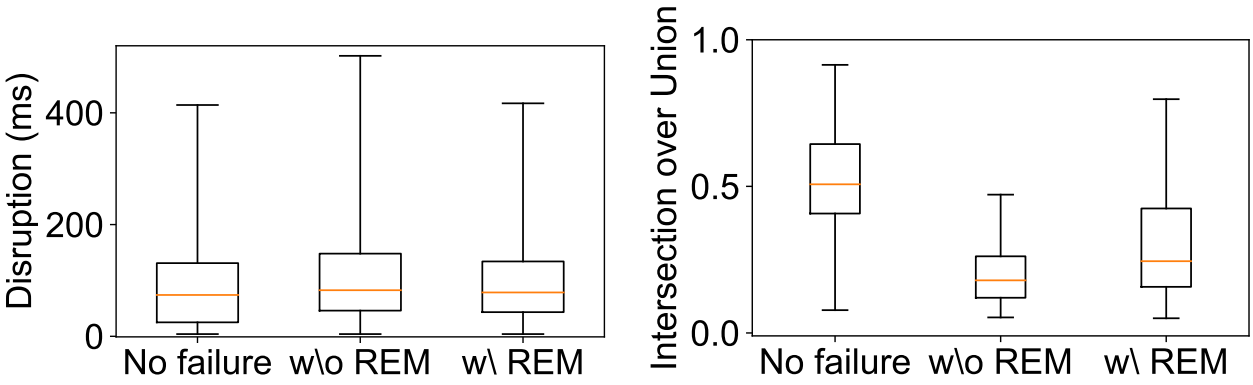
use Pixel 2XL as the client and deploy the edge server with the testbed. The overall setup is shown in Figure 4.9. The client is connected with the base station within the coverage of the USRP, so handover will not be triggered without channel dynamics in the lab environment. Therefore, we replay HSR traces to emulate the scenario with and without `CODIM`. Specifically, we inject handover to the testbed and control the delay based on replayed traces.

We launch AR/VR demo applications to test performance. The following settings are consistent with either legacy mobility management or `CODIM`. We deploy the VR and AR engine at the edge server co-located with the core network. For the mobile VR application, we consider VR streaming of medium quality with the same setting as [114]. The client sends motion updates to the server, expecting to receive a streamed VR frame and render the view



(a) Testbed (b) Emulator architecture

Figure 4.9: Testbed setup for edge-based VR/AR.



(a) VR frame disruption (b) IoU for AR object recognition

Figure 4.10: Performance of mobile VR/AR.

with the updated frame. We evaluate performance by checking whether the request frame is missing after the client renders an updated view. In our experiments, we let the client send periodic motion updates. For the mobile AR application, the client streams real-time video to the edge server for object detection. After receiving the identified location of recognized objects, the client will render the bounding box of the object on the current frame. If the object recognition result is delayed due to network failure, the rendered bounding box might not overlap with the ground truth bounding box. We use the same streaming content for Mobile VR and AR to guarantee the results are not affected by streaming content. We quantify the timeliness of the recognition result for evaluation.

Disruption reduction for mobile VR. We evaluate the disruption that a requested frame is missing when the user updates its view. Figure 4.10a shows CODIM reduces the median (95th-percentile) disruption from 82.5 (508.7) ms to 78.5 (418.4) ms for affected frames. We also evaluate the disruption under the static case. The median (95th-percentile) disruption is 74.0 (415.8) ms, which proves CODIM reduces the median disruption added by handover failure by 47.1%. Note not all frames experience disruption; we find that the percentage of affected frames is similar for the case with and without CODIM. CODIM outperforms legacy 4G/5G since it mitigates the disruption by reducing failure-caused disruption to normal handover latency.

Recognition performance for mobile AR. We evaluate the performance based on Intersection over Union (IOU), which is a common metric to evaluate whether the identified object bounding box matches with the ground truth in object detection and tracking [71]. Figure 4.10a shows shows CODIM improves the median (95th-percentile) IoU from 0.18 (0.43) to 0.24 (0.59) for affected frames. To quantify the overall performance, we take the IoU threshold as 0.25 as proposed in [71]. The ratio exceeding the threshold is 49.7% (88.3% improvements) with CODIM compared with 26.4% in the case without CODIM. We notice that CODIM's benefit is more significant under low-IoU samples. This is because CODIM reduces the probability of handover failure where IoU is low due to failure-caused disruption.

4.6 Open Issues

We discuss the limitations of interference cancellation.

Deployability of CODIM. The interference estimation needs to be deployed at both the device side and the base station side. At the first stage, incremental deployment is feasible with dynamic scheduling of OFDM-based reference signal and OTFS-based reference signal. Thus both legacy receivers and OTFS-enabled receivers can estimate channels from the corresponding reference signal. For interference cancellation, the deployment solely depends

on the base station side. The client only needs to send a sounding reference signal per configuration. The deployment of the backhaul exchange network is Open-RAN [117] friendly since the RAN is implemented in the cloud in Open-RAN.

Synchronization at the network side. Our design requires that the network side coordinates reference signals when sharing the spectrum. With the recent advances in time-synchronization techniques like PTP (Precision Time Protocol) [65], the under-ms synchronization is no longer challenging. When the backhaul network exchanges packet information for interference cancellation, there is no synchronization requirement. However, the processing and transmission need to be completed before the deadline for sending packet acknowledgment. Looking forward, we expect that strong synchronization will be feasible in the near future. At that time, strong-consistent interference cancellation will provide better accuracy for cancellation.

Sensing with OTFS. We note that sending with OTFS can enable more applications, e.g., differentiating UE type. We note that drones and ground devices require different handover policies. However, drones have arbitrary trajectories and flexible moving patterns. Operators can not tune handover parameters for both drones and ground devices at the same time. In 4G/5G, operators adjust handover triggering time and condition to reduce failure rate for devices with different speed [41]. We thus identify a key application: how to differentiate drones and ground devices? The key insight is that we can leverage a single UE & single BS to estimate the device height or speed pattern. The UE could estimate the distance to all antennas of the serving BS. We leave this as future work.

4.7 Related Work

Interference cancellation has been studied for decades under the context of WLANs and cellular networks [26]. There are two approaches in general, pre-transmission-based and post-transmission-based.

Pre-transmission-based. Before transmission, access points can coordinate the antennas to mitigate and even cancel interference. This approach is commonly used for 802.11-based WLANs since there is no synchronization between access points. Recent studies [132, 136] investigated how to apply cancellation under the context of MIMO, achieving a high throughput boost compared to legacy approaches. On the receiver side, there is another stream of work called Interference alignment [53], which requires the receivers to tune antennas for interference alignment. These approaches are generally restricted by their practicality in obtaining accurate channel state estimation.

Post-transmission-based. A representative post-transmission based approach is successive interference cancellation (SIC) [57, 107]. However, naive SIC is known to be less effective when the power gap between senders is not coordinated [107]. Another practical approach to resolving collision without canceling the interference is another stream [72]. These earlier approaches do not consider the cooperation between access points through the backhaul networks. The exploration of backhaul is also analyzed in [59, 137, 31] as the inter-cell networks have higher bandwidth. However, existing approaches are still restricted to negotiating data rates before transmission.

Our approach of cross-domain interference cancellation is inspired by the cross-technology interference cancellation [52]. These works view interference as informative features instead of random noises. Our work builds on the same idea but adopts a fundamentally different approach; it explores the diversity between domains without the requirement of negotiated orthogonality in improving the throughput of wireless networks.

CHAPTER 5

Integrating X-Domain Designs in 6G and Beyond

This chapter is organized as follows. We first introduce the goals (§5.1) and the challenges of integration (§5.2). We then detail the design of a novel overlay to enable standard-compliant integration in §5.3. We elaborate on the details of implementation in §5.4. §5.5 presents the evaluation. We discuss the extension of this work in §5.6 and related works in §5.7.

5.1 Introduction

We first enable group-based measurements within 5G PHY. As shown in Algorithm 1, a single cell measurement is used to infer the entire group of cells residing on the same tower. We next address two issues: (1) How to multiplex the cells' reference signals in the standardized OFDM PHY? (2) How to cope with heterogeneity among co-located cells inside a group? Our solution uses two ideas to address both issues: (a) embedding the delay-Doppler reference signals into the current 5G OFDM PHY as an overlay; (b) transforming measurements for cells with heterogeneous PHY configurations.

In this chapter, we describe `NEMONET` (NExt-gen MObile NEtwork Testbed), which enables researchers to perform trace-driven emulation with real-world traces collected by off-the-shelf mobile devices in the commercial network. It includes two parts, the open radio platform and the trace-driven emulator.

5.2 Challenges of Integration

Although attractive, OTFS cannot be directly applied to signaling traffic only. It suffers from the coexistence with OFDM data. We are neutral to whether data should also use OTFS. OTFS for data would also reduce Doppler shifts for faster data speed [122, 56], but at the cost of more processing delays and system complexity. Instead of mandating if OTFS should be used for data or not, we leave this decision to operators and future designs, and offer a universal and seamless solution for *both* choices. To function correctly, OTFS requires a continuous $M \times N$ OFDM grid. But in 4G/5G, both the mobility signaling traffic and data are multiplexed over the OFDM grid[1]. In case data still uses OFDM, the signaling traffic may span on *discontinuous* OFDM resource elements, and cannot apply OTFS directly.

One solution is to decouple the data and signaling channels, which requires redesigning the 4G/5G physical channel and thus not backward compatible. Our solution should naturally support coexistence and backward compatibility with OFDM data, *without* changing the 4G/5G physical channel designs.

5.3 Integration with Existing Standards

5.3.1 Embedding OTFS Signal Over OFDM

The algorithms rely on transformations in the delay-Doppler domain at the client and the cell tower. We thus embed the signaling in the DD domain to the current 5G signals. To this end, we adapt Orthogonal Time Frequency Space (OTFS), a delay-Doppler modulation scheme, onto the current physical layer (PHY) measurement blocks. Therefore, our design still fits in the current 5G OFDM-based PHY.

Our design requires standard-compliant changes on measurement blocks of the PHY

[1]RRC signaling traffic is also transmitted over PDSCH/PUSCH as data, not PDCCH/PUCCH.

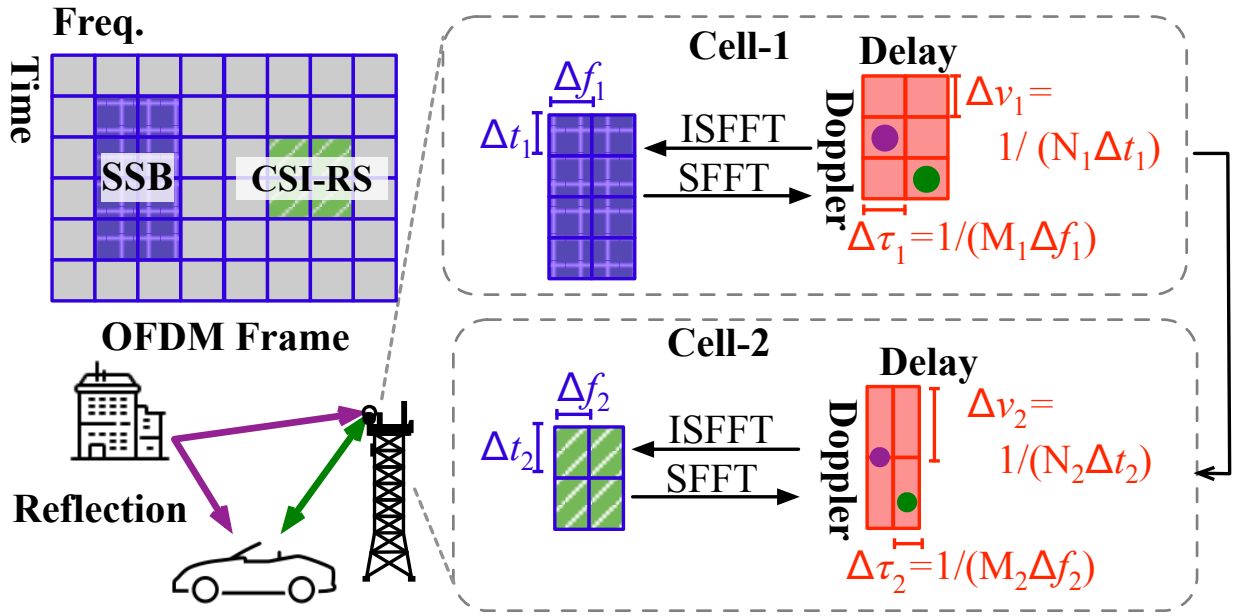


Figure 5.1: Concurrent channel inference in delay-Doppler domain (OTFS embedded into OFDM).

layer. 5G NR utilizes two types of resource blocks for cell measurement, SSB blocks and CSI-RS blocks [12]. Each cell schedules SSB blocks with reserved resources (i.e., subcarriers and symbols) for measurements. Multiple SSBs are transmitted continuously, which form an SSB burst (spanning 1∼5ms and 240 frequency subcarriers). Different from SSB, CSI-RS blocks are dynamically scheduled for each device with more flexibility. One CSI-RS block can take up to 2 symbols and 4 sub-carriers. The cell might prompt the device to measure several CSI-RS blocks to cover wider bandwidth during measurement.

We construct an OTFS overlap upon the current 3GPP standardized OFDM layer. We place OTFS grids over the SSB burst and dynamic CSI-RS blocks for measurements. Since an SSB burst is a whole chunk on the time-frequency domain, `NEMONET` can place an OTFS grid over it and measure cells on the delay-Doppler domain. A burst may last 1 - 5 ms and appear with periodicity of 5 - 160 ms[2]. This way, the mobile network can avoid the prohibitive cost of completely changing the modulation scheme.

[2]The options include 5, 10, 20, 40, 80 or 160 ms.

We place new OTFS grids over the existing OFDM ones (illustrated in Figure 5.1), and largely reuse the current scheme to multiplex reference signals from different cells. Specifically, OTFS symbols are placed at the same locations as SSB burst and dynamic CSI-RS grids for cell measurements. In Figure 5.1, a 4x2 grid for Cell 1 is scheduled for SSB. `NEMONET` transforms OTFS into a 4x2 OFDM grid and embeds it into the OFDM frame. We thus obtain OTFS measurements via the OTFS overlay on top of existing OFDM grids. We then run Algorithm for the entire cell group. With measurements from the OTFS signals, we further transform the measurements into standard-compliant SNR/RSRP/RSRQ metrics on radio signal level and quality [12]. RSRP and RSRQ are used for cell selection in the current practice [43]. The transformation from OTFS to OFDM is realized with ISFFT. We thus transform the OTFS signal into the OFDM domain, compute the adopted metrics, and follow the current 5G practice.

Note that, the OTFS scheme is used for reference signals only, whereas OFDM is still used for other data types (e.g., packet delivery). We thus reduce the prohibitive cost of completely revamping the OFDM based modulation at PHY.

Handling block heterogeneity. In the above framework, we make further changes to address heterogeneity in numerology and subgrid size. First, 5G NR adapts the grid to handle heterogeneity over a much wider range of frequency spectrum. Cells with frequencies far apart probably have different numerology, i.e., symbol duration Δt and sub-carrier Δf[3]. Therefore, the transformation to the DD domain brings diverse numerology $(\Delta\tau, \Delta\upsilon)$. `NEMONET` tackles the diversity by mapping multi-path profiles from one grid to a different one. Figure 5.1 illustrates the details with an example of inferring cell-2 based on cell-1; Two cells use different numerologies. First, we estimate cell-2's multi-path profile based on parameters retrieved from cell-1. Next, we map those parameters to coordinates on the OTFS grid with cell-2's numerology.

[3]5G currently supports 4 numerologies, with 15kHz, 30kHz, 60kHz, 120kHz as subcarrier spacing Δf (extended to 960kHz in recent releases) [16]

Second, the subgrid size can be different for different cells with flexible scheduling of SSB and CSI-RS blocks. Because the inference algorithm can handle fractional Doppler, the selection on grid size (i.e., N, M) becomes much more flexible. The state-of-the-art [77] relies on long duration of symbols on TF grid (i.e. large N equivalent to 40ms) to achieve acceptable accuracy of integer Doppler (i.e. fine-grained Δv). As `NEMONET` takes fractional parts into account, it could relax the above restriction and adopt a grid lasting as short as an SSB burst. As shown by Figure 5.1, the reflection path can be represented with fractional Doppler coordinates on cell-2, as the Doppler step runs larger and is less precise. Choosing a smaller N would save time spent on OTFS signaling and thus improve wireless efficiency by allocating more blocks for OFDM data. In the example of Figure 5.1, cell-1 utilizes only a (4x2) subgrid and cell-2 utilizes a (2x2) subgrid.

5.3.2 Dynamic Scheduling for Signaling

We design a dynamic overlay via scheduler adaptation at the base station. As shown in Figure 5.1, the base station's scheduler can allocate all resource elements for signaling messages (e.g. measurement reports and handover command) to an $M' \times N'$ subgrid of the 4G/5G resource grid ($M' \leq M, N' \leq N$). In this way, All signaling messages can directly apply OTFS in $M' \times N'$ sub-grid.

Our solution supports the following features. 1) Backward compatible with 4G/5G. All data traffic can still use OFDM today; No interference between signaling and data; No change for the physical channel design, backward compatible with 4G/5G: Only refine base station's traffic scheduler, which is simply an implementation and operational issue. 2) Always achievable with marginal impact on scheduler today, because signaling traffic (SRB) has the highest priority by design [5, 16]. Given signaling and data, signaling traffic will always be transferred first due to its importance; Signaling traffic is always scheduled first due to its importance for functional correctness. So the base station can always find a sub-grid for signaling traffic first, and then allocate the remaining for data transfer; 3) Marginal impact

on scheduler today, which has already prioritized signaling traffic (though not necessarily allocate them into a sub-grid); In 4G/5G design, the base station schedules the traffic based on a per-logical-channel (rather than per-user) basis. So different users' signaling traffic can be bundled together to occupy the $M' \times N'$ sub-grid; 4) Efficient spectral utilization: (M', N') will dynamically adapt with runtime signaling traffic volume, so no resource waste or additional signaling overhead.

Overhead for signaling. Additional $O(M'N'log(M'N'))$ SFFT/ISFFT pre/post-processing cost for signaling only (similar to LTE/5G uplink SC-FDMA's overhead compared to downlink OFDMA). There is no additional overhead on data.

Limitation. M', N' can be small (depending on signaling traffic volume) to sample real Doppler shift and path delay. Nevertheless, the cost is unavoidable if data does not use OTFS; To improve it, the scheduler may allocate more guard resource elements to enlarge M', N', at the cost of resource waste. Such a tradeoff is similar to 5G's additional DMRS for high mobility today [12], but is better than DMRS since it helps stabilize the signaling channel. If data uses OTFS, this problem will disappear since $M' = M, N' = N$.

5.3.3 Design of Channel Quality Metrics

Channel profiling before handover. One missing piece of channel estimation is that uplink channel quality is not measured for drones. Current 4G/5G adopts hard handover, with no uplink traffic before handover succeeds. Our finding on the uplink interference features enables `NEMONET` to derive the uplink channel quality of the target cell without uplink measurement. The main challenge is that reciprocity does not apply to interference. We leverage the separation of channel quality into path-dependent channel response and receiver-location-dependent interference. The interference depends on the location of the receiver only. For example, all UEs' uplink at a base station share the same uplink interference. Given that 4G/5G enables uplink interference measurement with uplink sounding signal, which is

Table 5.1: Standardized QCI characteristics [8]

QCI	Type	Priority	Delay budget (ms)	Loss rate
3	GBR	3	50ms	10^{-3}
6	Non-GBR	6	300ms	10^{-6}
9	Non-GBR	9	300ms	10^{-3}

constantly sent by the UE for uplink channel estimation, `NEMONET` shares the measurement of other UEs without adding over-the-air transmission.

The naïve way to measure uplink interference is to configure the client to send the over-the-air signal. However, this approach is infeasible in the cellular network since it requires every potential client to send the signal to the candidate cell. The signaling storm will paralyze the candidate cell, let alone cause high interference with other cells. Our approach is efficient without adding over-the-air signaling. The client A sends the measured downlink channel response to the serving cell. The candidate cell sends the interference measure by sounding the reference signal of active clients to the serving cell of client A. The serving cell combines the channel response and interference to make a handover decision. It's worth noting that communication between cells can be resource-wasting. We propose an event-based scheme to reduce the signaling overhead between BSes. Only when UL interference drops more than a threshold, the cells sends updated interference level to other cells.

5.4 Implementing the Next-Gen Open Radio Platform

5.4.1 Scalability Under MmWaves

In 2018, we extended NEMONET for the below-IP user-plane analytics. Different from the control plane, the user plane faces an explosive growth of over-the-air messages. This poses challenges to energy-efficient, real-time analytics in commodity devices.

Characteristics of user-plane analytics. Compared to control-plane messages, user-plane messages are *simpler* with fewer fields, but *more intensive* with a massive amount of packets to deliver. Figure 5.2 shows the user plane's messages are 2∼3 orders of magnitude more frequent than the control plane's.

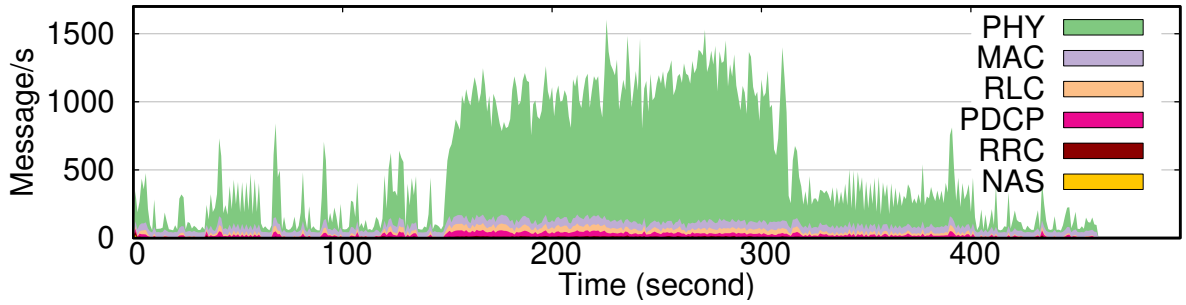
We next quantify how well the initial NEMONET tackles user-plane analytics inside devices. We enable all messages in Table 5.2 to evaluate the initial NEMONET's runtime responsiveness, energy, and CPU usage. We repeat this test with data collection only, collection + message parsing, and collection + parsing + analysis. For energy usage, we also compare NEMONET with the worst-case background scenario when the screen is always on. Figure 5.4 shows the results. We make three observations:

• **Real-time responsiveness.** Surprisingly, even with intensive user-plane messages, NEMONET can still timely process them *before* the next message arrives. For each message, we define its accumulative lag as the elapsed time that its processing is after the next message's arrival. Figure 5.3 shows NEMONET can analyze ≥95% user-plane messages with ≤1ms lag. The maximal lag is ≤8ms. This is because most user-plane messages are simple to process.

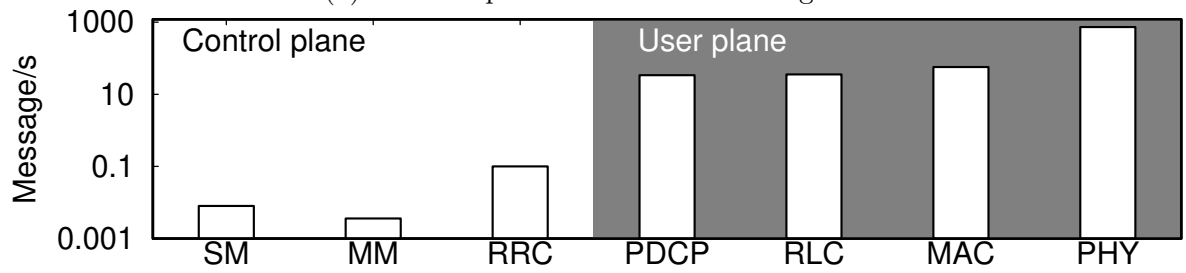
• **Energy deficiency.** The initial NEMONET's responsive real-time user-plane analytics is at the cost of huge energy and CPU usage. Figure 5.4c shows that, with all messages enabled, NEMONET consumes 21% battery in 1 hour, which is 1.5× compared to the scenario with the always-on screen. The battery is mostly used by software, since the data collection from the chipset consumes comparable energy to the scenario without NEMONET. Its energy consumption is proportional to message volumes.

• **Heavy CPU usage.** With all messages, Figure 5.4b shows the initial NEMONET occupies one core and uses 12–23% CPU in total. The CPU usage is proportional to message volumes, and dominated by message parsing (≥99%) in software space.

Vanilla solution: Domain-specific independent sampling. To save the battery and CPU, NEMONET should reduce the cost of processing intensive link/physical-layer messages



(a) An example of the runtime messages



(b) Average message arrival rates

Figure 5.2: Runtime messages at the control and user planes.

and retain high analytics accuracy. For non-real-time tasks, `NEMONET` can collect raw messages in device and analyze them offline. For real-time analytics, `NEMONET` can sample the messages to analyze. The initial version of `NEMONET`'s user-plane analytics uniformly sampled each physical/link-layer's messages independently. Figure 5.4b implies sampling can be approximated by parsing only a subset of messages in software. This approach still retains all the raw data in the collection phase to facilitate full-fledged offline analysis. Figure 5.5a confirms it effectively reduces the CPU usage.

Moreover, the sampling can be optimized with domain-specific knowledge. We find 4 types of messages (MAC buffer status, block error rate, serving cell measurements, and uplink transmission power) contribute $\geq$71.8% of total messages. We customize their analytics to be sampling-aware, treat un-sampled messages as missing data, and optimize their analytics accuracy with the missing data inference. Figure 5.5b exemplifies the customization for the uplink MAC-layer queuing delay for each segment. To track a segment's queuing delay,

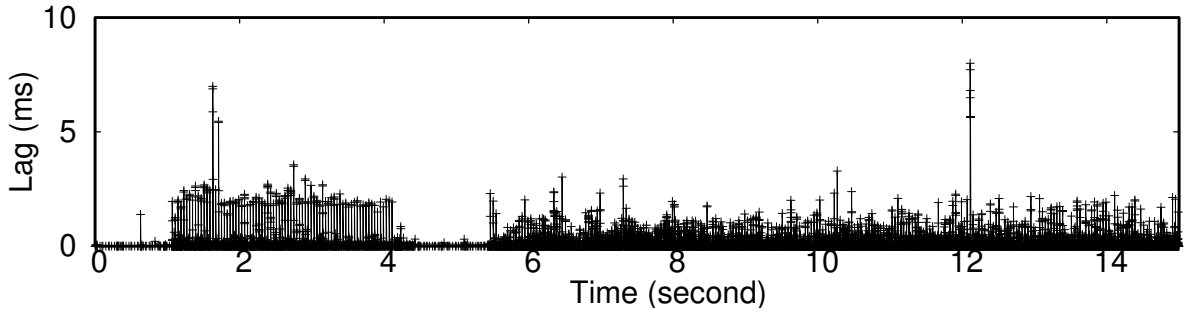
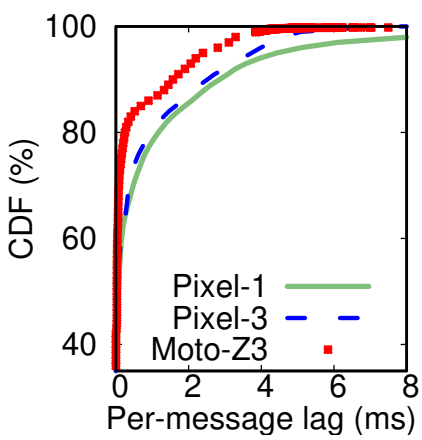


Figure 5.3: Runtime per-message processing lags.

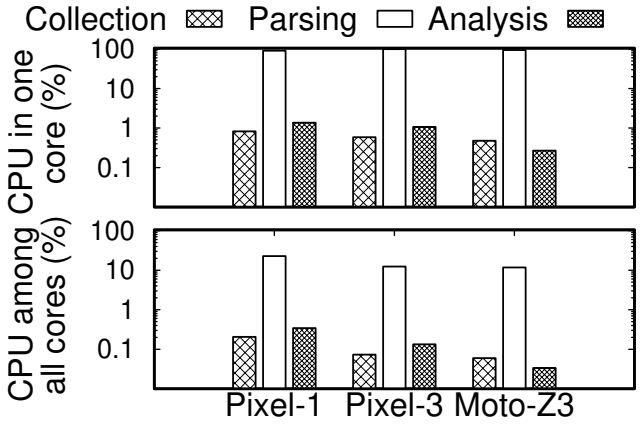
NEMONET needs two timestamps for it when entering and leaving the buffer. The naive sampling is unaware of this timestamp dependency and simply samples the MAC logs uniformly. This is prone to miss one of the timestamps and thus failure of tracking the latency. Instead, by tracking logs continuously to cover both timestamps for each segment, the optimized sampling is more accurate under the same sampling ratio.

Our solution: Domain-specific cross-layer sampling. However, independent sampling turns inaccurate in *cross-layer* analytics. An IP packet will traverse across link/physical layers for delivery, during which it can be divided into multiple RLC segments and MAC blocks based on available physical radio resources. If messages from different layers are sampled independently, the cross-layer dependency between the IP/PDCP packets, RLC segments, MAC blocks, and physical resources can be lost. This causes not only inaccurate analytics, but also inefficient sampling since sampled messages across layers are mismatched and wasted. Figure 5.5c exemplifies this deficiency when tracking the uplink packet latency. With 10%, 20% and 50% sampling ratio, independent sampling only ensures 0.04%, 0.52% and 11.38% of IP packets can be fully tracked across layers. This leads to 34.2%, 33.6% and 8.6% estimation errors for uplink packet latency, respectively.

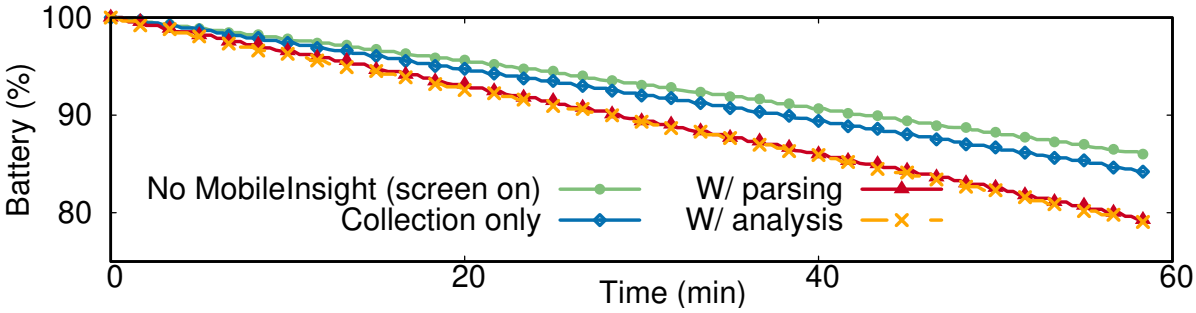
To this end, we devise *cross-layer sampling*. Rather than independent sampling among layers, NEMONET first uniformly samples the IP packets at PDCP based on the target sampling ratio. Then for each sampled IP packet, NEMONET runs the cross-layer dependency tracking to locate the corresponding messages related to its RLC segments, MAC blocks, and PHY radio



(a) Real-time responsiveness



(b) CPU utilization

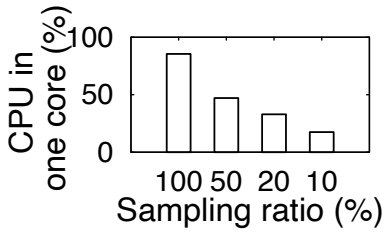


(c) Battery consumption at Motorola Z3.

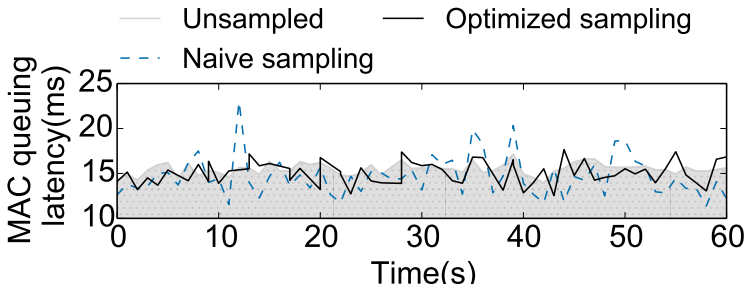
Figure 5.4: The initial NEMONET's user-plane analytics.

resource allocation. NEMONET only parses (samples) these messages and drops the remaining for efficiency. In this way, the cross-layer dependencies for these sample IP packets are all retained for high analytics accuracy. This approach is applicable to both uplink and downlink data transmission, because the dependency across the PHY, MAC, RLC, and PDCP layers exists for both uplink and downlink.

Figure 5.5 evaluates NEMONET's cross-layer sampling. In terms of its accuracy, Figure 5.5c shows NEMONET reduces the data latency estimation error by 4.1×, 4.5×, and 3.4× with 10%, 20%, and 50% sampling ratio, respectively. It retains comparable actual sampling ratios (i.e., the percentage of IP packets that can be correctly tracked across layers) to the



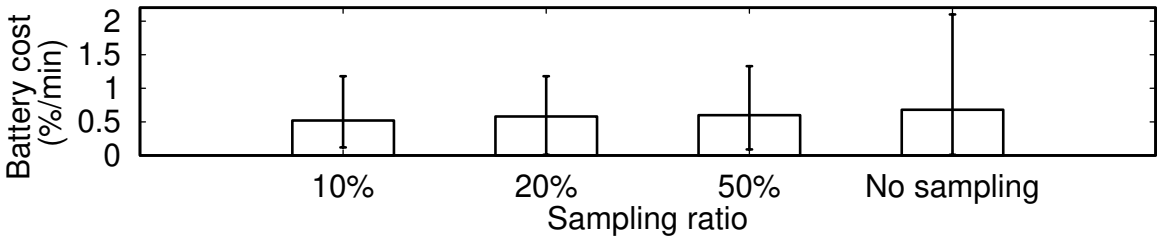
(a) CPU reduction



(b) Single-layer sampling at MAC layer

	Actual sampling ratio			Per-packet latency analysis error		
Target sampling ratio	10%	20%	50%	10%	20%	50%
Independent sampling	0.04 %	0.52%	11.38%	6.76ms (34.2%)	6.64ms (33.6%)	1.69ms (8.6%)
Cross-layer sampling	9.37%	18.1%	49.45%	1.63ms (8.3%)	1.48ms (7.5%)	0.50ms (2.5%)

(c) Effectiveness of domain-specific cross-layer sampling



(d) Energy efficiency of cross-layer sampling (Motorola Z3)

Figure 5.5: NEMONET's energy-efficient real-time user-plane analytics (uplink data latency analysis as an example).

target. Figure 5.5d confirms viable energy saving (by up to 47.6%) with NEMONET's cross-layer sampling. Due to the variance of traffic and battery drain, the energy saving is not strictly proportional to the actual sampling ratio. This suggests more energy savings are possible with further refined solutions.

```
# Initialize an online monitor
src = OnlineMonitor()
# Initialize the KPI manager
kpi = KPIManager()
# Enable KPI analyzer on handover latency
kpi.enable_kpi("Mobility.HANDOVER_LATENCY")
# Bind the analyzers to the monitor
kpi.set_source(src)
# Start analysis
src.run()
```

KPI analyzers

```
# check whether handover is triggered
# by tracking the RRC state machine
if ho_istriggered(event):
    HO_triggered = True
    ts = event.timestamp
# check whether handover is completed
# by tracking the RRC state machine
if HO_triggered and ho_iscomplete(event):
    latency = event.timestamp - ts
    HO_triggered = False
```

Simplified underlying implementation

Figure 5.6: APIs of KPI analyzers in NEMONET (handover disruption latency as an example).

5.4.2 Standard-Compliant Analytics

The new features empower NEMONET with deep mobile network analytics. But they were not widely used as expected, since they are unfriendly to new users because of their low-level natures. So in 2019, we started to streamline NEMONET's analytics to balance its comprehensiveness and user-friendliness.

We extend NEMONET with user-friendly, extensible performance indicator (KPI) analyzers. Table 5.3 summarizes the latest NEMONET's available KPIs from 3GPP standards [20, 19] and user requests. They have covered the control and user planes, and various aspects such as reliability, performance, and energy efficiency. They are not available from the legacy mobile OS APIs. Compared to the protocol/packet state machines in basic NEMONET analyzers, KPIs are more intuitive for users to understand runtime mobile network operations. The KPI analyzers can simplify many usage scenarios by the community.

Figure 5.6 exemplifies NEMONET's APIs for the runtime KPIs (control-plane handover latency as an example). NEMONET defines user-friendly hierarchical names (category.kpi_name) for the KPIs in Table 5.3. For each KPI, NEMONET defines a KPI analyzer, and calls the corresponding state machine-based analyzers to track it. Instead of calling the complex low-level analyzers, a new user can easily track KPIs by simply declaring them via names,

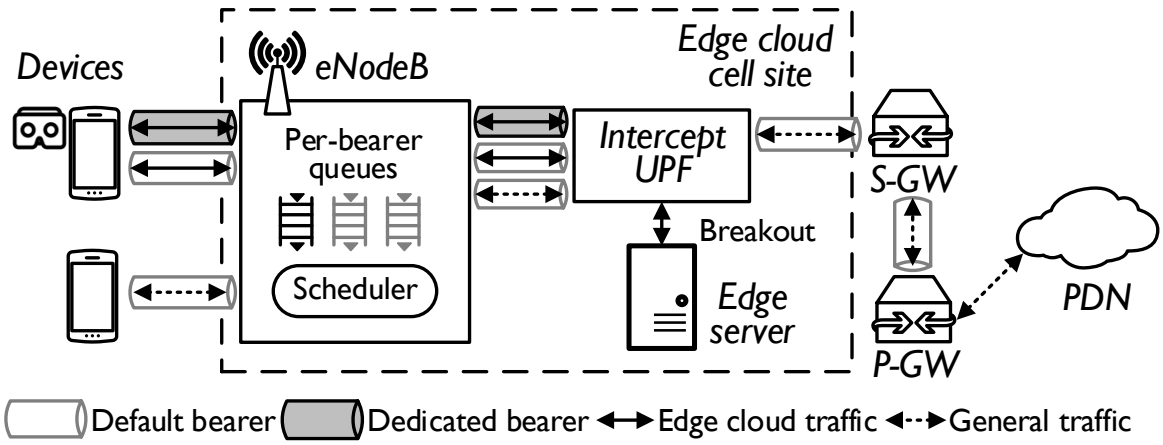


Figure 5.7: Edge based remote rending.

without worrying about the low-level details of mobile network protocols. Meanwhile, an experienced user can dive into the underlying behaviors related to KPIs. `NEMONET` supports both *periodic* and *event-driven* runtime KPI reporting. By locally storing the historical data, it allows users to query the aggregated KPI statistics by time, location, network node (e.g., cells) and client.

5.5 Evaluation

5.5.1 Efficiency

Testbed setup. We evaluate end-to-end latency with an edge-based remote rendering video streaming system on a private LTE testbed as shown in Figure 5.7. The LTE testbed consists of an eNodeB with LTE release 12, a data interceptor and a co-located edge server. The eNodeB manages two cells running on band 30 with 10MHz channel bandwidth. The packet interceptor user plane function (UPF) implements local breakout and routes edge cloud traffic to the edge server. Other general traffic is routed via the regular path onwards

to the SGW and PGW. The PGW is connected to the Packet Data network (PDN). The testbed establishes dedicated bearers for different QCIs and routes traffic through dedicated bearers based on IP/Port/Protocol filters. The edge server is deployed as a VM with eight CPU cores and 16GB memory. The MEC platform complies with ETSI specifications [46]. We tested with Samsung Galaxy S7 and Samsung Galaxy Note 9.

Remote Rendering Application. Since our main focus is on measuring network side latency, we develop a simplified application to simulate the behavior of the thin-client system in [82]. On the server side, we read a video file (H.264, 1080p, using I and P frame only) and store every video frame in memory. Each frame size ranges from 41 KB to 368 KB with an average of 58.3 KB. The client sequentially transmits a fetch request for one frame and the server replies with the frame data upon receiving the request. After receiving the full frame, the client uses MediaCodec that calls hardware codec APIs to decode the video and display it on the screen. We use TCP in our experiments to guarantee reliable delivery. In an accompanying paper, we have also observed that TCP outperforms UDP for VR delivery using MEC [109]. The client only runs the streaming application unless specified. This simplified application allows us to accurately measure the end-to-end system latency for every frame and easily separate the latency caused by network and mobile devices. We consider rendering and transcoding latencies on the server to be negligible. Note that cloud platforms typically take 5ms to process one frame [93].

Data collection and analysis. We collect three types of traces to quantify network latency: (i) the application logs on server and client; (ii) the LTE network signaling traces with MobileInsight [78]; and (iii) the TCP/IP traces on clients and servers with tcpdump. We measure end-to-end latency on the client side. Since the server processing latency is negligible, the overall latency is broken down into network latency and client latency. To compare client latency and network latency, we further extract network latency by timestamping when the frame request was sent and when the frame was received. In total, we analyzed 35.6 million frames traces in both the testbed and the simulation.

We begin by looking at overall latency performance and dissect network latency from the whole. The end-to-end latency composes of network latency, server-side latency and client-side latency. Here we define network latency as the time elapsed from the packet sent to the time response received at the client-side application layer. Since the server side only transmits pre-render frames in our experiments, the server-side latency is negligible. As shown in Figure 5.8, device latency is stable compared with network latency. Also, network latency constitutes 77.5% of total latency (96.1 ms of 124.0 ms). The medians are 81 ms and 112 ms for network latency and total latency. Though device side latency depends on device capability, we believe Samsung S7 and Note 9 are mid to high-performance phone models.

We then investigate the root cause of abnormal network latency. As compared with the ping latency, the median of ping latency is only 61 ms and the minimum is 31 ms. Since ping latency is collected using ICMP probing packets, the latency shall be longer than the data packet. The reason why the ping latency is less than the network latency for the streaming application can be one of the followings: processing latency, transmission latency and queuing latency caused by congestion. Since there is only one device connected with the cell during testing, there should be minimum congestion. As shown in Figure 5.8, the minimum network latency is only 61 ms, which is far beyond standardized 4ms processing delay limits [4]. So we can safely assume the prolonged latency is transmission latency.

We further investigate the root cause of prolonged transmission latency. We prove that transmission latency is not prolonged by limited bandwidth but the retransmission. With 256QAM enabled, peak throughput is around 50.7Mbps during frame transmission. So the normal transmission latency for a single 58KB frame only takes 9 ms. In our experiments, 0.4% packets are retransmitted by TCP, which is higher than previously measured 0.06% in [61]. The retransmission delay is 154 ms on average. Packet loss can happen at the wireless channel, interceptor, or the end hosts. We validate that the wireless channel packet corruption is all recovered by RLC retransmission. Then we can conclude that packet loss

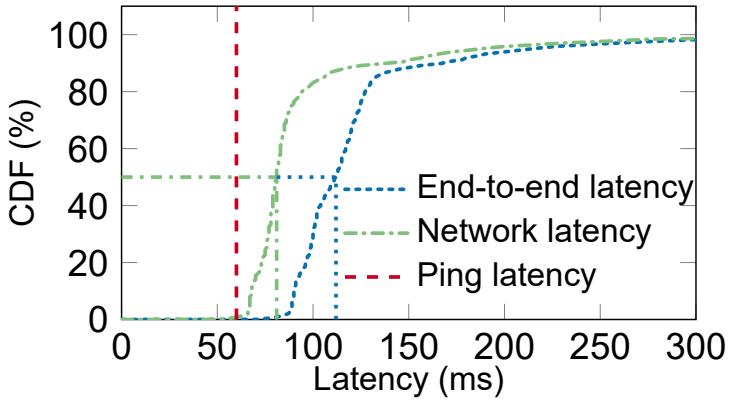


Figure 5.8: End-to-end latency breakdown.

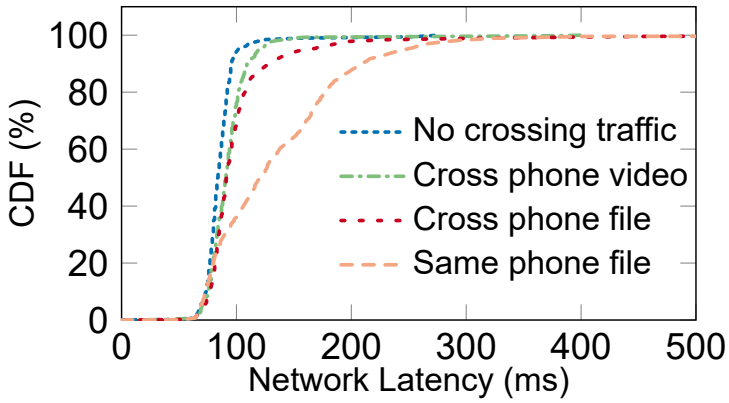


Figure 5.9: Latency with conflicting traffic.

happens at either the interceptor or the end hosts. We further check the interceptor logs and validate that no packets are discarded. So we conjecture that TCP retransmits packets because packets are lost over the core network.

Takeaway: Network latency is the key driver of overall latency, rather than device processing latency. Excessive network latency mostly comes from TCP retransmission rather than frame transmission.

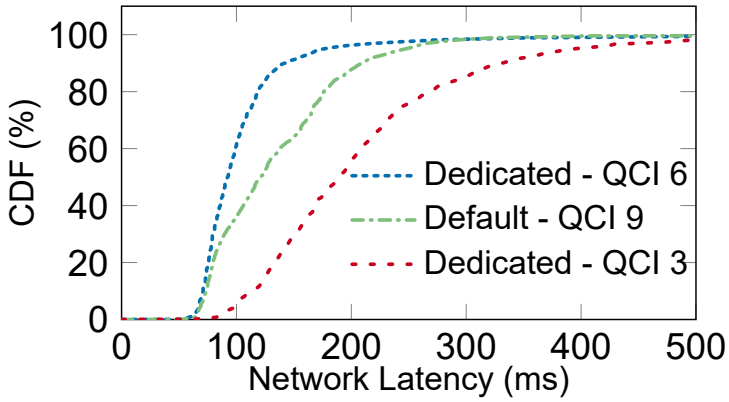


Figure 5.10: Latency with different QCI.

5.5.2 Effectiveness for Latency Measurements

As discussed in the previous section, TCP retransmission prolongs network latency. Another portion of network latency is queuing latency especially when the base station needs to schedule limited radio resources between multiple users. According to a recent study [30], most base station tends to serve only one user at a time, which add queuing latency to other waiting users. In this section, we study how congestion prolongs queuing latency.

We conducted experiments with two congestion scenarios: single phone and cross phone scenario. In the single phone scenario, the VR application and the traffic generator run on the same phone. The single phone scenario mimics the case of simultaneously downloading large files (at a data rate of around 10Mbps) or streaming background video frames (at a data rate of around 6.8Mbps) [70]. In the cross-phone scenario, two applications run on different phones. The cross-phone scenario emulates a congested cell with crossing traffic.

Experiment with testbed shows that under crossing traffic the overall latency will increase significantly. As shown in Figure 5.9, latency increases in both the single-phone scenario and the cross-phone scenario. In the single-phone scenario, per-frame network latency with crossing traffic increases by 56.9% (86.3 ms to 135.4 ms). Latency 75 percentile increase from 91 ms to 168 ms. In the cross-phone scenario, per-frame network latency with crossing traffic

increases by 8.2% (86.3 ms to 93.4 ms) for video traffic and 19.6% (86.3 ms to 103.2 ms) for file traffic. Latency 75 percentile increases from 91 ms to 100 ms and 103 ms, respectively.

The intuitive solution to avoid competition between latency-sensitive traffic and latency-insensitive traffic is to guarantee latency requirements by differentiating these two traffic and priority latency-sensitive ones. In current streaming frameworks, pre-fetched based scheme is commonly adopted since it masks long latency to users [102, 70]. However, when pre-fetching fails, the user needs to wait for an entire round trip for an emergent frame request. The user tolerance difference motivates us to differentiate emergent frame requests from pre-fetching frame requests.

Crossing traffic significantly prolongs network latency, especially for crossing traffic from the same device. The network side should differentiate emergent streaming traffic and serve with higher priority.

Varying testbed configurations. Each operator can customize radio configurations to guarantee service quality at different network entities. For example, different operators might assign different profiles for the same VoLTE traffic. It is often neglected by researchers how radio configuration can make impacts beyond the radio connection part. In this section, we will show how inappropriate radio configuration deteriorates latency performance. Since in 5G, the layering design does not change significantly and studied radio configurations are still effective [5], the following findings still hold in 5G.

We first examine the generality of testbed radio configurations by comparing them with existing configurations in operational LTE networks. We analyze traces from LTE-VR database [113], which includes 3.2 million cellular messages collected over 8 months on four major operators in the U.S. Our analysis of operational LTE networks aligns with our testbed setting. We checked configurations from all four major operators by examining RRC signaling messages [5]. All of them use RLC AM mode for default bearer, which is the same in our testbed. In addition, the RLC re-transmission timer range from 35 ms to 60 ms in all four

operators with 38 ms on average, which corresponds with 40 ms setting in the testbed. Thus we believe testbed configurations align with real-world scenarios and measurement should be consistent with experiments in the wild.

To investigate the impact of radio configurations, we consider three QoS with significantly different QoS parameters listed in Table 5.1. Note that QCI 3 and QCI 6 are for the dedicated bearer and QCI 9 is for the default bearer. We first show a comparison of two non-GBR bearers, QCI 6 and QCI 9, the former with higher priority than another. We then show a comparison of QCI 3 (GBR bearer) and QCI 9 (non-GBR bearer). Since the single phone experiments show more latency, we adopt the experiment setting of a single phone generating the crossing traffic.

We first compare two non-GBR bearers setting of QCI 6 and QCI 9. We observe latency reduces by sending traffic through a dedicated bearer with QCI 6 as shown in Figure 5.10. Per frame network latency with the dedicated bearer (QCI 9) is smaller than the default bearer (QCI 6). Average latency reduces by 26.7% (135.4 ms to 106.9 ms). The 75 percentile reduces from 177 ms to 115 ms. From the above analysis, we can draw the conclusion that network-side service differentiation can reduce network latency significantly.

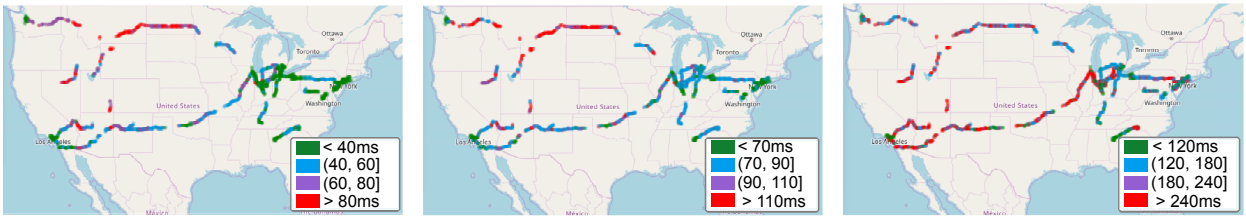
We then compare the non-GBR bearer of QCI 9 with the GBR bearer of QCI 3. The network should allocate resources to QCI 3 bearers prior to QCI 9 bearers. However, We observe that sending traffic with QCI 3 does not reduce latency. Average latency increases by 55.6% (135.4 ms to 210.7 ms). The root cause is improper radio configurations under GBR bearer that prolong packet loss recovery latency. RLC layer configuration for QCI 3 bearer is unacknowledged (UM) mode in the testbed. RLC mode defines whether RLC layer recovery is enabled. Since RLC is configured as UM mode, packet loss recovery solely relies on TCP re-transmission, which could potentially prolong latency. We also get testbed side limited logs to verify our findings. Based on eNodeB side records, the queuing latency for the dedicated bearer is 8.2 ms, while for the default bearer is 16.0 ms. The default dearer IP packet discard rate is 0, while the dedicated bearer IP packet discard rate is 1.2 Kbps.

Prioritizing streaming traffic can help to reduce latency only when radio configurations are properly configured. Configuration sacrificing reliability help to reduce queuing latency but increase transport layer re-transmission, which prolongs overall latency even under good channel quality.

Efficiency of analytics. This section uses network latency to exemplify how NEMONET empowers in-device mobile network analytics. We run a 5-year longitudinal study to unveil and understand end-to-end (E2E) latency over operational mobile networks. More results have been available from NEMONET with the collected dataset being released. In this work, we showcase how to use NEMONET's new modules, and how these modules unveil some new insights that were not visible. Note these lessons become visible until NEMONET's new cross-layer analytics, user-plane analytics, and user-friendly KPIs become available.

Since 2016, we have sporadically collected 4G/4.5G LTE over-the-air messages with the evolving NEMONET and accumulated a five-year dataset. We ran NEMONET over the test phones when using ping (primary), iperf, web, video streaming, or virtual reality applications in static, walking and driving scenarios. summarizes our dataset as of August 15, 2020. It has been collected from 50+ phone models and 58 global operators over 20+ countries and regions, including the USA, China, India, South Korea, Singapore, France, Spain, Germany, Norway, Hungary, Egypt, Australia, New Zealand, etc. Most data (82%) is collected in the US (AT&T, Verizon, T-Mobile, Sprint, Google Fi), covering 39 states and 260,000+ miles. The early tests in 2016–2018 only enabled the control-plane messages and partial user-plane messages. With NEMONET's user-plane features in §5.4.1, we further enabled full link/physical-layers messages for cross-layer analytics in late 2018 and afterwards. We ran a large-scale across-the-US driving test in May - July 2019, resulting in a surge of the dataset size in 2019 Q2.

Analytics methodology. We apply NEMONET's new features to analyze the mobile network latency. We group our dataset by quarter/month and device contexts (phone models, mobility patterns, and locations) and replay them to extract the mobile network latencies



(a) AT&T (1%) (b) AT&T (50%) (c) AT&T (95%)

Figure 5.11: Ping RTT latency over the US (AT&T).

using KPI analyzers in §5.4.2. As summarized in Table 5.3, these KPI analyzers provide simple interfaces to unveil the user-perceived network latencies from the control plane, user plane, and across planes. They depend on the cross-layer analytics and sampling (§5.4.1) for efficient and accurate KPI extractions. To analyze the root causes of the latency variations, we further perform the cross-layer analytics over the dataset to track the signaling procedures at the control plane, IP packet delivery across link/physical layers at the user plane, and the interplays between the control and user plane.

5.5.2.1 Latency analysis with NEMONET

Figure 5.11 plots ping latency (round-trip time) measured in AT&T in the past year (Q2 2019 - Q2 2020). Results for other carriers in the U.S. (Verizon, T-Mobile, Sprint and Google Fi) are similar and omitted. Unsurprisingly, E2E latency varies over the location (much worse in freeways, mountain and rural areas than in big cities). In particular, the 1th (a robust estimate of minimal latency), 50th (median) and 95th percentiles of RTT are below 40ms, 70ms and 120ms in almost all the cities; But the 1th percentile is $>$ 80ms in mountain areas and $>$ 40 ms over the freeways across the US. These results are consistent with other measurement studies by FCC [49, 48], Ookla speedtest [94, 3] and OpenSignal [95, 96, 97], but offer finer-grained spatial analysis.

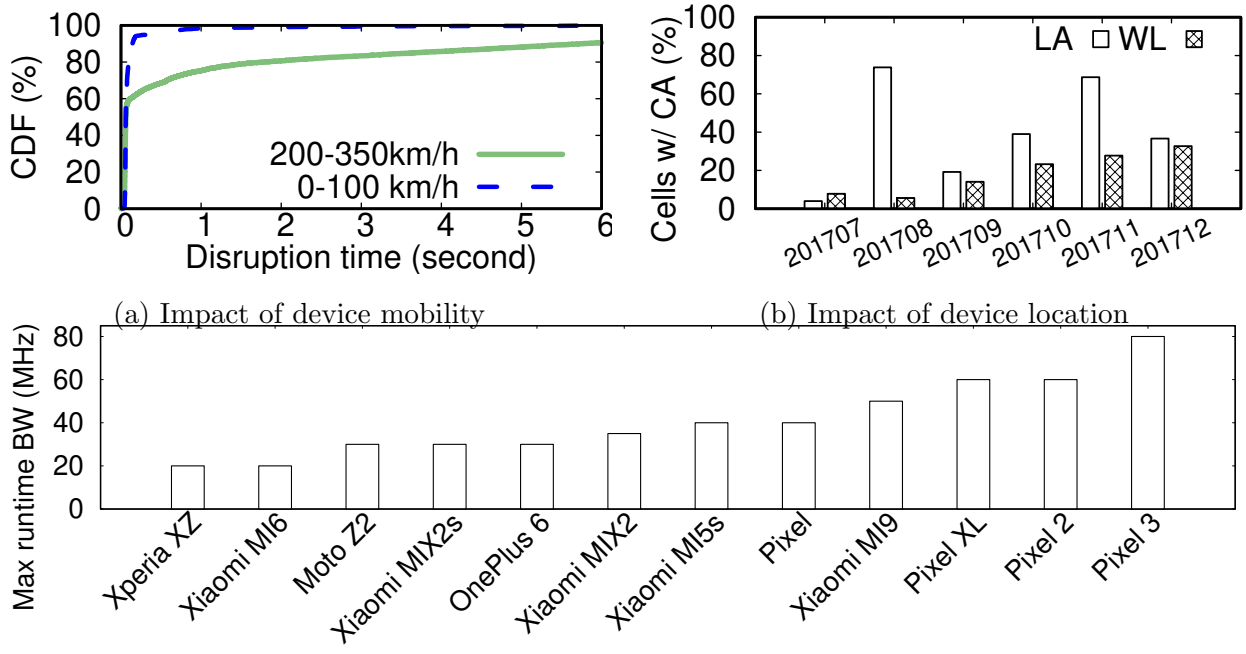
NEMONET's KPI analyzers report that, in the past 5 years, LTE network latency has decreased at both the control *and* user planes overall. At the control plane, the signaling

protocols take time to establish the radio connectivity and migrate it in device mobility, before which all data delivery will be blocked. For a fair comparison, we examine the latency evolution in one eastern U.S. city using AT&T. Similar trends hold for other cities and operators in the same dataset. Results show a 38.2% average latency reduction in radio connectivity setup, and 48.1% reduction in handover. Note the missing data will incur the incomplete signaling procedure (e.g., lost connectivity setup), which is not counted in this figure. The sampling is not applied because the control-plane signaling messages are marginal (Figure 5.2). At the user plane, we observe a 50.0% average reduction of the first uplink packet's delivery latency after radio connectivity setup, and an increment of downlink bandwidth (thus faster transmission).

Different mobile devices have heterogenous hardware modems and capabilities that affect their latencies and bandwidth. Based on `NEMONET`'s runtime physical-layer logs, Figure 5.12c shows the maximum radio bandwidth allocated for each device. In general, a device with earlier or lower-end generations of hardware modem has less bandwidth and thus a slower data speed. With advanced radio technologies in newer chipsets, and the device can enjoy more bandwidth and less latency at runtime.

5.6 Discussion

Beamforming and MIMO. Beamforming is one of the selling points for 5G. With the increasing need for high bandwidth, beamforming and MIMO have become the norm of broadband communication. Our testbed is extensible to beamforming and MIMO with hardware upgrade of the RF frontends and software upgrade of beamforming/MIMO management. For example, to support 2x2 MIMO, the testbed should have 2 TX and 2 RX. Our testbed with USRP X300 supports simple 2x2 MIMO. In the next-gen cellular network, MIMO is expected to reach the capability of up to 512 TX/RX. Our design of channel state inference will still benefit the applications of beamforming/MIMO.



(a) Impact of device mobility

(b) Impact of device location

(c) Impact of device types

Figure 5.12: The device context's impact on LTE latency.

Variants of PHY. The testbed only supports OFDM and OTFS-based PHY for now. To embrace future variants of PHY, we devise the PHY-MAC API to be open and standard-compliant. In order to support another variant of PHY, e.g., visible light communication, the PHY entity only needs to provide standard compliant API by encapsulating the variants of the coding scheme, channel state inference, etc. Our design of trace-driven emulation of the above-PHY still applies to the variants of the PHY layer.

Beyond RAN. Our testbed only discusses the design and implementation of the last hop, RAN and device side. Beyond RAN, the core network is another battlefield of next-gen cellular network design. The cloudification of the core network indicates future core will be closely tied with the commercial core. There are mnay options of core networks, e.g. magma[100], OpenEPC[2]. Our prior work on Flora[54] builds on open-source core Magma [100] and supports customized over-the-air communication between the core and the customized SIM. We leave the integration of this testbed with Flora as future work.

5.7 Related Work

Onward to 6G and beyond, there are generally two types of innovations, the wireless side, and the system side. On the wireless side, there are different approaches to boost bandwidth and enable more radio technologies, e.g., full-duplex radio to better utilize spectrum [38, 62], backscattering to enable ubiquitous access [37]. These approaches are still at the prototype stage. There are active efforts to study commercial networks and re-design the protocols on the system side. The major stream of these works focuses on improving the throughput using different techniques. Recent works [73, 43] re-design cell selection with either network or device-centric solutions. Another stream of studies investigates resource scheduling in the context of carrier aggregation or multiple connectivity [64, 50, 79]. Compared with these efforts, our work is readily deployable while fundamentally changing the lower radio layer to enable more usage scenarios and applications.

Table 5.2: Available messages from NEMONET.

Protocol	Message Types	# Msg	# Element
4G-PHY	PDSCH signal; Cell Measurement	2^4	N/A[5]
4G-MAC	Uplink/downlink transport blocks; MAC config; Buffer status report	5^a	54
4G-RLC	Control/data packet data unit	2^a	N/A^b
4G-PDCP	Control packet data unit	1^a	N/A^b
4G-RRC	System info blocks; Connection setup/release/re-establish/reconfig; Handover command; Measurement control/report; Radio capability equerry; Paging; Security model command	45	185
3G-RRC	Same as 4G-RRC	32	108
3G-MM/GMM	Attach/detach; Authentication request/response; Location update; Security mode control; Identification request/response; Service request; Paging	41	63
4G-EMM	Same as 3G-MM/GMM	32	108
3G-CM/SM	Session (EPS bearer/PDP context) setup/modify/release; PDN connect/release/modify	58	54
4G-ESM	Same as 3G-CM/SM	22	71
CDMA/EvDo	Paging information; connectivity establishment/release; radio link protocol status	5^a	N/A^b

Table 5.3: Available runtime KPIs from NEMONET.

Category	Key performance indicator (KPI)	Protocol
Radio	Signal strength (RSRP, RSRQ, RSSI, EvDo, RSCP)	PHY
	Channel estimation (CQI, PMI, RI)	PHY
	Modulation and coding scheme	PHY
	Radio resource block allocation	PHY
	Block error rate (BLER)	PHY
	Downlink path loss	PHY
Accessibility	Radio connection setup success rate	RRC
	Attach success rate	MM
	Session setup success rate	SM
	Service request success rate	RRC+MM
	Session QoS/billing class	MM+SM
Mobility	Tracking area update success rate	MM
	Tracking area update latency	RRC+MM
	Handover success rate	RRC
	Handover disruption latency	RRC
	Handover HOL blocking latency	RLC+RRC
Retainability	Abnormal RRC connection drop rate	RRC
Integrity	Data throughput	PDCP
	Packet loss ratio	PDCP
	Link-layer data loss ratio	RLC+MAC
Energy	Uplink transmission power	PHY
	Downlink reception power (DRX)	PHY+RRC

CHAPTER 6

Conclusion

This chapter briefly summarizes our work, highlighting the novel designs with a new domain to overcome wireless dynamics. We then present insights and lessons learned during the investigation. Finally, we discuss the future work in embracing new radio design in aerial networks.

6.1 Summary

Our approach is inspired by the delay-Doppler domain, which reveals frequency-decoupled physical path features. With explicit physical path modeling, we design practical systems in the delay-Doppler domain to profile channel states and interferences. We further exploit the shared features between cells to enable decentralized access. Based on this approach, we make three main contributions.

Concurrent channel state inference. The legacy approach of channel state inference is performed by each cell independently, ignoring the common path features shared by cells. We exploit the similarity benefitted from shared physical paths. We present our design on multipath decoupling that enables high-fidelity channel state inference for drones. The delay-Doppler path features separate path delays from frequency domain coupling. By characterizing path delays, we can decouple multiple path propagation with adaptive scheduling of delay-Doppler signals. We also show that our design improves the accuracy by orders of magnitude compared with existing approaches.

Interference cancellation with cooperative decoding. The main challenge to high-throughput data transmission is interference. Although interference cancellation is promising to turn interference into assisting data, it usually requires data rate and power coordination, adding restrictions on spectrum utilization. We show a novel cross-domain interference cancellation scheme, which requires no coordination between cells. The insight is that sharing channel state knowledge between cells enables decoding for all modulation. We leverage cross-domain multiplexing of signal and data to cancel interference. Hence, we can transform interferences into meaningful interference profiles to boost data speed.

Interference management for high-fidelity cell selection. The characteristics of air propagation (e.g., lack of blockage, side-lobe-based antenna) cause excessive interference. The receivers cannot accurately profile interference and decode data without redesigning the channel estimation algorithm. Given current OFDM-based cell quality indicators fail to capture interference, we leverage a novel transform and inter-cell collaboration to profile interference. The intuition is that interference only depends on the location of the receiver. Therefore, we design the collaborative interference profiling algorithm to exchange interference knowledge between cells leveraging 5G high-speed backhaul. We quantify the impact of interference and improve the channel estimation accuracy to enhance the performance.

Experimental testbed for XG. We presented an open-radio platform enabling researchers to perform the trace-driven evaluation of new radio designs. The platform provides mmWave support and offers an end-to-end evaluation pipeline. Unlike existing mmWave platforms with signal-level testing, we assisted experiments with abundant traces collected during driving and high-speed train. Therefore, users can replay the traces with desired channel and mobility setting. We show several use cases to demonstrate how researchers can leverage our platform for PHY layer optimization. We hope our work can add something to the developing record of the field.

6.2 Insights and Lessons Learned

During the investigation of how to innovate the legacy cellular network designs to adapt the networks in the sky, We discovered the following insights and lessons.

Physical paths are sparse and of low variability. With the rolling-out of mmWave bands, there turns out to be a gap between the PHY design and the wireless link. The PHY design in current cellular networks assumes that the link is stable in the microscopic view. However, the wireless channel changes much faster than the assumption. Our design thus unravels the variant wireless channel features by modeling physical paths. We solve the major obstacle to profiling the multi-pathing effect with detailed modeling of per-path features. Besides, the sparse path features are shared between co-located cells, which matches the increasingly dense cell deployment trend.

Cross-domain multiplexing compresses interferences. Our designs are all based on overlaying OTFS modulation with OFDM modulation. Our main insight is that the non-orthogonality causes strategical interferences, which can be transformed into benign signals or assisting signals for data decoding. The cross-domain approach is different from the legacy approach in that all signals and data are coupled in a single domain. The intra-domain coupling posts restrictions on how to improve transmission gain under multiple interferers. With a cross-domain approach, the interferences between domains can be easily canceled. Besides, there is no restriction on what signal the cells send at the time-frequency domain. The fundamental reason why the cross-domain approach enables cancellation is that the delay-Doppler domain spreads signals over the entire frequency band, causing linear correlation between frequencies. The correlation can thus be interpreted into channel knowledge.

Adapt to new radios with software designs. Embracing new radios into cellular networks is challenging. While mmWaves bring a wide spectrum and potential of high band-

width, it is difficult to fully release the potential due to the complexity of the wireless environment. Our insight in defeating the complex propagation hassles in mmWave is that we finally rely on software designs to enable a high-performance PHY layer. We also gain wisdom that bringing new innovations in protocol designs can be easy and fast as long as the initial design is flexible and open. Looking forward, we expect the integration of cross-domain signals into 6G or beyond to be easy since the overall mechanism stays the same. This benefitted from the current flexible PHY frame design.

6.3 Future Work

We envision the following future work to continue our exploration in a cross-domain multiplexing design with integration of the delay-Doppler domain.

Incorporating IoT devices. In the foreseen future, personal-use IoT devices will largely increase together with the network with wider coverage [89]. Compared with commodity phones or professional drones, IoT devices are constrained by limited battery power. Fundamentally, the communication based on time-frequency domain scheduling is power-consuming since the power change drastically for all the sub-carriers. Recent works discuss how the OTFS domain solution relieves IoT devices from battery-consuming modulation [112, 131]. We expect the design of cross-domain multiplexing shed light on a future exploration into cross-domain multiplexing-enabled power saving. Besides, as IoT devices become prevailing, the mmWave-enabled short-range network is expected to cover more devices. In the long term, we envision IoT devices enabling learning-based paradigms with coarse-grained data. All diverse devices could share their network knowledge and achieve collaborative mobile network intelligence.

Cross-layer design with new radios. The current deployment of mmWaves is still constrained by the limited beam setting offered by the physical hardware. We expect that the integration of mmWaves with software-defined PHY enables more flexible services with

mmWaves. For such services, the PHY has to provide ultra-low latency to adapt to the short symbol duration at high frequency. In our design of OTFS-based signaling, we add an extra layer to coordinate OFDM and OTFS symbols. To further enable coordination between different beams, we need a slim layer for PHY management for beam, modulation, coding, and so on. By separating management from lower-layer signal transmission, the design enables more innovation with closer integration between the PHY and the upper application layers.

Future of aerial networks with ground networks. With mmWaves in aerial communication, drones can act as a replay between ground networks. Such deployment reveals a large potential of bandwidth wasted due to NLOS transmission. Past work has shown the potential of supporting ground networks with mmWaves. However, the details on how to overcome the interference between drones are not clear. Our design of cross-domain interference cancellation removes the obstacles due to interference. However, the practicality of such a design remains to be validated in a real deployment.

Applicability to emerging satellite networks. Another direction for aerial networks is closely bonded with emerging satellite networks. Starlink [116, 118] will include 4,425 LEO satellites (57 available until 2019) at an altitude of 550–1,150 km. All satellites use high-frequency bands (12– 18GHz Ku band or 26.5–40GHz Ka-band), and traveling at high speeds (around 7.3km/s). Such a setting results in a viable Doppler effect (300KHz-900KHz frequency shift) and dramatic wireless dynamics (thus latency variations). The unique propagation features bring extra challenges [76]. We expect to witness more innovations in the cooperation between low-height drone-based networks and satellite networks.

Appendix A

Supplementary Materials for Chapter 3

A.1 Proof for Theorem 2

Proof. We prove that the SVD decomposition leads to perfect decomposition under single path model. We first prove that the channel matrix $\mathbf{H} = \mathbf{\Gamma P \Phi}$ is a possible compact SVD decomposition. We then prove that the compact decomposition is unique.

The decomposition $\mathbf{H} = \mathbf{\Gamma P \Phi}$ is only a SVD decomposition when $\mathbf{P}$ is a diagonal matrix with non-negative real numbers. Given $P = 1$, we construct $\mathbf{P}$ as a 1×1 matrix with $\mathbf{P}(1,1) = |h_p|$. Thus $\mathbf{P}$ satisfy the condition. We then prove the delay spread matrix $\mathbf{\Gamma}$ is semi-unitary. Note that $\mathbf{\Gamma}$ is a $M \times 1$ vector. We only need to prove that the delay spread matrix satisfies $||\mathbf{\Gamma}|| = 1$. Note we only need to consider the case that τ_p is divisible by $\Delta\tau$. In the delay-Doppler domain, we have

$$\Gamma(k) = \frac{1}{M} \sum_{d=0}^{M-1} e^{-j2\pi(\tau_p - k\Delta\tau)d\Delta f}$$

. Since $\Delta\tau = \frac{1}{\Delta f M}$ based on the grid setting, we have

$$|\Gamma(k)| = |\frac{1}{M} \frac{1 - e^{-j2\pi(\frac{\tau_p}{\Delta\tau} - k)}}{1 - e^{-j2\pi(\frac{\tau_p}{\Delta\tau} - k)/M}}| = |\frac{1}{M} \frac{sin(\pi x - \pi k)}{sin(\pi \frac{(x-k)}{M})}|$$

, where $x = \frac{\tau_p}{\Delta\tau}$. Proving $||\mathbf{\Gamma}|| = 1$ is equivalent to prove that

$$\sum_{d=0}^{M-1} |\frac{sin(\pi x - \pi d)}{sin(\pi \frac{(x-d)}{M})}|^2 = |sin(\pi x)|^2 \sum_{d=0}^{M-1} csc^2(\pi \frac{(x-d)}{M}) = M^2$$

From the series representation of *csc* funtion, we have

$$\csc^2(\pi\frac{(x-d)}{M}) = \sum_{r\in\mathbb{Z}} \frac{1}{(\pi\frac{(x-d)}{M} - r\pi)^2}$$

Through the expansion, we have

$$\sum_{d=0}^{M-1} \csc^2(\pi\frac{(x-d)}{M}) = \sum_{d=0}^{M-1}\sum_{r\in\mathbb{Z}} \frac{1}{(\pi\frac{(x-d)}{M} - r\pi)^2}$$
$$= N^2 \sum_{r\in\mathbb{Z}}\sum_{d=0}^{M-1} \frac{1}{(\pi x - r\pi)^2} = N^2 \frac{1}{sin(\pi x)}$$

Similarly we can prove $\mathbf{\Phi}$ is also semi-unitary when ν_p is not divisible by $\Delta\nu$. Thus we conclude that $\mathbf{\Gamma}$, $\mathbf{P}$ and $\mathbf{\Phi}$ satisfy the SVD decomposition.

According to [63], the SVD decomposition is unique if the values of eigenvalues are distinct. Since there is only one eigenvalue for P, the compact SVD is distinct. The compact SVD decomposition of H results in the delay spread matrix $\mathbf{\Gamma}$, attenuation $\mathbf{P}$, and Doppler spread matrix $\mathbf{\Phi}$. □

A.2 Number of Paths

Proof. The channel matrix representation $\mathbf{H} = \mathbf{\Gamma P \Phi}$ indicates that

$$rank(\mathbf{H}) = rank(\mathbf{\Gamma P \Phi}) \leq min(rank(\mathbf{\Gamma}), rank(\mathbf{P}), rank(\mathbf{\Phi}))$$

By Sylvester's rank inequality [88], we have

$$rank(\mathbf{\Gamma P \Phi}) \geq rank(\mathbf{\Gamma}) + rank(\mathbf{P}) + rank(\mathbf{\Phi}) - 2P$$

Assuming condition (i) and condition (ii), $\mathbf{\Gamma}$ and $\mathbf{\Phi}$ are unitary matrix. We have $rank(\mathbf{\Phi}) = P$ and $rank(\mathbf{\Gamma}) = P$, where P is the number of physical paths. Then $rank(\mathbf{H}) = rank(\mathbf{P}) = P$. □

A.3 Proof of Proposition 1

Proof. We detail (3.6)'s proof; (3.7)'s proof follows similar procedure. Let $x_T[k,l] = x_1[k,l]c[k]$, then $x_2[k,l] = x_T[kT,l] = x_1[kT,l]c[kT] = x_1[kT,l]$. Then

$$\begin{aligned}
X_2[n,m] &= \sum_{k=0}^{M_2-1}\sum_{l=0}^{N-1} x_2[k,l]e^{-j2\pi(\frac{mk}{M_2}-\frac{nl}{N})} \\
&= \sum_{k=0}^{M_2-1}\sum_{l=0}^{N-1} x_T[kT,l]e^{-j2\pi(\frac{mk}{M_2}-\frac{nl}{N})} \\
&\textit{(Let } k' = kT\textit{. Note } M_2T = M_1\textit{, and } x_T[k,l] = 0 \textit{ if } k \neq k'T\textit{)} \\
&= \sum_{k'=0}^{M_1-1}\sum_{l=0}^{N-1} x_T[k',l]e^{-j2\pi(\frac{mk'}{M_2T}-\frac{nl}{N})} \\
&= \sum_{k'=0}^{M_1-1}\sum_{l=0}^{N-1}\frac{1}{T}\sum_{q=0}^{T-1} x_1[k',l]e^{-j\frac{2\pi}{T}qk'}e^{-j2\pi(\frac{mk'}{M_1}-\frac{nl}{N})} \\
&= \frac{1}{T}\sum_{q=0}^{T-1}\sum_{k'=0}^{M_1-1}\sum_{l=0}^{N-1} x_1[k',l]e^{-j2\pi(\frac{(m+\frac{qM_1}{T})k'}{M_1}-\frac{nl}{N})} \\
&= \frac{1}{T}\sum_{q=0}^{T-1} X_1[n,[m+\frac{qM_1}{T}]_{M_1}] = \frac{1}{T}\sum_{q=0}^{T-1} X_1[n,[m+M_2]_{M_1}]
\end{aligned}$$

thus concluding the proof. □

Appendix B

Supplementary Materials for Chapter 4

B.1 Single Path Estimation

Proof. Here we give the proof for Theorem 4.3.1. We show the general case where we have a general y[u, v]. We will prove that the the columns of **S** are linear independent. We first formulate the vector equations

$$\sum_{(m,n)\neq(u,v)} \alpha_{(m,n)} e^{-j2\pi(\frac{1}{M})mk} e^{-j2\pi(-\frac{1}{N})nl} = 0, for\ \forall k,l \neq (0,0)$$

, where $\alpha_{(m,n)}$ is the coefficient of column corresponding to $y[m,n]$. For each row, we use

$$\beta_k^n = \begin{cases} \sum_m \alpha_{(m,n)} e^{-j2\pi(\frac{1}{M})mk}, & \text{if } n \neq v \\ \sum_{m\neq u} \alpha_{(m,n)} e^{-j2\pi(\frac{1}{M})mk}, & \text{otherwise} \end{cases} \tag{B.1}$$

to represent the coefficient for $e^{-j2\pi(-\frac{1}{N})nl}$. For l varying from 0 to N, for any k, we have

$$\sum_{(m,n)\neq(u,v)} \alpha_{(m,n)} e^{-j2\pi(\frac{1}{M})mk} e^{-j2\pi(-\frac{1}{N})n*0} = 0 \rightarrow \sum_n \beta_k^n e^{-j2\pi(-\frac{1}{N})n*0} = 0$$

$$\sum_{(m,n)\neq(u,v)} \alpha_{(m,n)} e^{-j2\pi(\frac{1}{M})mk} e^{-j2\pi(-\frac{1}{N})n*1} = 0 \rightarrow \sum_n \beta_k^n e^{-j2\pi(-\frac{1}{N})n*1} = 0$$

$$\cdots$$

$$\sum_{(m,n)\neq(u,v)} \alpha_{(m,n)} e^{-j2\pi(\frac{1}{M})mk} e^{-j2\pi(-\frac{1}{N})n*(N-1)} \rightarrow \sum_n \beta_k^n e^{-j2\pi(-\frac{1}{N})n*(N-1)} = 0$$

Thus we can formulate a $N * N$ matrix that

$$\begin{bmatrix} e^{-j2\pi(-\frac{1}{N})0*0} & e^{-j2\pi(-\frac{1}{N})1*0} & \dots & e^{-j2\pi(-\frac{1}{N})(N-1)*0} \\ e^{-j2\pi(-\frac{1}{N})0*1} & e^{-j2\pi(-\frac{1}{N})1*1} & \dots & e^{-j2\pi(-\frac{1}{N})(N-1)*1} \\ \dots & & & \\ e^{-j2\pi(-\frac{1}{N})0*(N-1)} & e^{-j2\pi(-\frac{1}{N})1*(N-1)} & \dots & e^{-j2\pi(-\frac{1}{N})(N-1)*(N-1)} \end{bmatrix} * \vec{\beta}_k = 0$$

Note that the left matrix is full rank since it is a Vandermonde matrix with distinct scale factor $e^{-j2\pi(-\frac{1}{N})n}$ for each column $n = 0, ..., N-1$. Thus we have $\beta_k^n = 0 \ \forall k, n$. Note that $\beta_k^n = \sum_m \alpha_{(m,n)} e^{-j2\pi(\frac{1}{M})mk}$. For $n = v$, we have $\beta_k^v = \sum_{m \neq u} \alpha_{(m,v)} e^{-j2\pi(\frac{1}{M})mk} = 0$. Varying k form 1 to $M-1$, we can formulate a $M-1 * M-1$ matrix that

$$\begin{bmatrix} e^{-j2\pi(\frac{1}{M})0*1} & \dots & e^{-j2\pi(\frac{1}{M})(u-1)*1} & e^{-j2\pi(\frac{1}{M})(u+1)*1} & \dots & e^{-j2\pi(\frac{1}{M})(M-1)*1} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ e^{-j2\pi(\frac{1}{M})0*(M-1)} & \dots & e^{-j2\pi(\frac{1}{M})(u-1)*(M-1)} & e^{-j2\pi(\frac{1}{M})(u+1)*(M-1)} & \dots & e^{-j2\pi(\frac{1}{M})(M-1)*(M-1)} \end{bmatrix} * \vec{\alpha}_v = 0$$

Note that the left matrix is full rank since it is a Vandermonde matrix with distinct scale factor $e^{-j2\pi(\frac{1}{M})m}$ for each column $m \neq u$. Thus $\alpha_{m,v} = 0 \ \forall m$. Next we show that $\alpha_{m,n} = 0 \ \forall n \neq v$. We have

$$\sum_{n \neq v} \alpha_{(m,n)} e^{-j2\pi(\frac{1}{M})mk} e^{-j2\pi(-\frac{1}{N})nl} = 0, for \ \forall k, l \neq (0,0)$$

we formulate that $A \bigotimes B$, where $\mathbf{A} = e^{-j2\pi(\frac{mk}{M})}$ and $\mathbf{B} = e^{j2\pi(\frac{nl}{M})}, n \neq v, l \neq 0$.

We then show how $\mathbf{A}$ and $\mathbf{B}$ are both full rank. We only need to prove A, a $M * M$ matrix, is full rank, where

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & ... & 1 \\ 1 & e^{-j2\pi(\frac{1}{M})} & e^{-j2\pi(\frac{2}{M})} & e^{-j2\pi(\frac{3}{M})} & ... & e^{-j2\pi(\frac{M-1}{M})} \\ 1 & e^{-j2\pi(\frac{2}{M})} & e^{-j2\pi(\frac{4}{M})} & e^{-j2\pi(\frac{6}{M})} & ... & e^{-j2\pi(\frac{2(M-1)}{M})} \\ ... & ... & ... & ... & ... & ... \\ 1 & e^{-j2\pi(\frac{k}{M})} & e^{-j2\pi(\frac{2k}{M})} & e^{-j2\pi(\frac{3k}{M})} & ... & e^{-j2\pi(\frac{(k-1)(M-1)}{M})} \\ ... & ... & ... & ... & ... & ... \\ 1 & e^{-j2\pi(\frac{M-1}{M})} & e^{-j2\pi(\frac{2(M-1)}{M})} & e^{-j2\pi(\frac{3(M-1)}{M})} & ... & e^{-j2\pi(\frac{(M-1)(M-1)}{M})} \end{bmatrix}$$

, where k varying from 1 to M-1. Note that $\mathbf{A}$ is a Vandermonde matrix. Thus the determinant of $\mathbf{A}$ is the product of each row's scalar factor.

$$\det(\mathbf{A}) = \prod_{1 \leq i < t \leq M-1} (e^{-j2\pi(\frac{i}{M})} - e^{-j2\pi(\frac{t}{M})})$$

To prove the determinant of $\mathbf{A}$ is non-zero, we model the $e^{-j2\pi(\frac{i}{M})}$ as vectors of length 1 at plane constructed by $1 + 0j$ and $0 + 1j$. Note these vectors are not the same. Thus the determinant of $\mathbf{A}$ is non-zero, which means $\mathbf{A}$ is full rank.

Now we prove $\mathbf{B}$ is full rank. We write it out as

$$\begin{bmatrix} e^{-j2\pi(-\frac{1}{N})0*1} & \dots & e^{-j2\pi(-\frac{1}{N})(v-1)*1} & e^{-j2\pi(-\frac{1}{N})(v+1)*1} & \dots & e^{-j2\pi(-\frac{1}{N})(N-1)*1} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ e^{-j2\pi(-\frac{1}{N})0*(M-1)} & \dots & e^{-j2\pi(-\frac{1}{N})(v-1)*(N-1)} & e^{-j2\pi(-\frac{1}{N})(v+1)*(M-1)} & \dots & e^{-j2\pi(-\frac{1}{N})(N-1)*(N-1)} \end{bmatrix}$$

Each column of this matrix is a geometric progression with scaling factor $e^{-j2\pi(-\frac{1}{N})l}, l = 1, ..., N-1$.. This is also a Vandermonde matrix with distinct scale factor. Thus $rank(\mathbf{B}) = N - 1$.

We have

$$rank(\mathbf{S}) = rank(\mathbf{A}) * rank(\mathbf{B}) = M(N-1)$$

This concludes out proof that $\mathbf{S}$ is full rank. Thus $\alpha_{m,n} = 0 \; \forall n \neq v$. We have $\alpha_{m,n} = 0 \; \forall m, n$. This is equivalent to prove that $\mathbf{S}$ is full rank. □

B.2 Approximation

We provide the proof for Theorem 4.3.2.

Proof. We calculate $Y[0,0]$ with $Y[\hat{0},0] = \mathbf{S}_{[0,0]} inv(\mathbf{S}_{\mathbb{R}^* \setminus (0,0), \mathbb{C}^* \setminus (m,n)}) \vec{Y}_{\mathbb{R}^* \setminus (0,0)}$. The ground truth is $Y[0,0] = \mathbf{S}_{[0,0]} inv(\mathbf{S}_{\mathbb{R}^* \setminus (0,0), \mathbb{C}^* \setminus (m,n)}) \vec{Y}_{\mathbb{R}^* \setminus (0,0)} - \mathbf{S}_{[0,0]} inv(\mathbf{S}_{\mathbb{R}^* \setminus (0,0), \mathbb{C}^* \setminus (m,n)}) \vec{S}_{\mathbb{R}^* \setminus (0,0), \mathbb{C}=(m,n)} * y[m,n] + y[m,n]$. We have

$$Y[\hat{0},0] - Y[0,0] = \mathbf{S}_{[0,0]} inv(\mathbf{S}_{\mathbb{R}^* \setminus (0,0), \mathbb{C}^* \setminus (m,n)}) \vec{S}_{\mathbb{R}^* \setminus (0,0), \mathbb{C}=(m,n)} * y[m,n] - y[m,n]$$

We will show that $inv(\mathbf{S}_{\mathbb{R}^*\setminus(0,0),\mathbb{C}^*\setminus(m,n)})\vec{S}_{\mathbb{R}^*\setminus(0,0),\mathbb{C}=(m,n)}$ is a vector of all-negative-ones $\tilde{-\mathbf{1}}$. It's equivalent to prove that $\mathbf{S}_{\mathbb{R}^*\setminus(0,0),\mathbb{C}^*\setminus(m,n)})\tilde{-\mathbf{1}} = \vec{S}_{\mathbb{R}^*\setminus(0,0),\mathbb{C}=(m,n)}$. This is equivalent to prove each row of $\mathbf{S}_{\mathbb{R}^*\setminus(0,0),\mathbb{C}^*}$ is 0.

To prove that $\mathbf{S}_{\mathbb{R}^*\setminus(0,0),\mathbb{C}^*}$ is 0. We have the first row (M=4 as an example)

$$\begin{bmatrix} 1 & e^{-j2\pi(\frac{1}{M})} & e^{-j2\pi(\frac{2}{M})} & e^{-j2\pi(\frac{3}{M})} & 1 & e^{-j2\pi(\frac{1}{M})} & e^{-j2\pi(\frac{2}{M})} & e^{-j2\pi(\frac{3}{M})} \end{bmatrix}$$

We only consider the first half. Thus the sum is $\sum_m e^{-j2\pi(\frac{m}{M})} = \frac{1-e^{-j2\pi(\frac{M}{M})}}{1-e^{-j2\pi(\frac{1}{M})}}$. This holds for rows. For row k, we have $\sum_m e^{-j2\pi(\frac{km}{M})}$. □

B.3 Multiple Paths

Proof. The (MN-P) * (MN-P) submatrix of transformation matrix $\mathbb{S}$ is full rank. Assume we have N = 2. We can reformat the matrix $\mathbf{H}$. We keep the removed rows and columns within the same block.

$$\begin{bmatrix} 0 & (1-e^{j2\pi\frac{1}{N}}) * A_{\setminus\mathbb{P},\mathbb{K}} \\ A & e^{j2\pi\frac{1}{N}} * A_{\setminus\emptyset,,\mathbb{K}} \end{bmatrix}$$

Then we can show $A_{\setminus\mathbb{P},\mathbb{K}}$ is full rank. Note that the complete A is full rank, we need to prove reduced A is still full rank,

A, a $M * M$ matrix, is full rank, where

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & ... & 1 \\ 1 & e^{-j2\pi(\frac{1}{M})} & e^{-j2\pi(\frac{2}{M})} & e^{-j2\pi(\frac{3}{M})} & ... & e^{-j2\pi(\frac{M-1}{M})} \\ 1 & e^{-j2\pi(\frac{2}{M})} & e^{-j2\pi(\frac{4}{M})} & e^{-j2\pi(\frac{6}{M})} & ... & e^{-j2\pi(\frac{2(M-1)}{M})} \\ 1 & e^{-j2\pi(\frac{3}{M})} & e^{-j2\pi(\frac{6}{M})} & e^{-j2\pi(\frac{9}{M})} & ... & e^{-j2\pi(\frac{3(M-1)}{M})} \\ ... & ... & ... & ... & ... & ... \\ 1 & e^{-j2\pi(\frac{k-1}{M})} & e^{-j2\pi(\frac{2(k-1)}{M})} & e^{-j2\pi(\frac{3(k-1)}{M})} & ... & e^{-j2\pi(\frac{(k-1)(M-1)}{M})} \\ ... & ... & ... & ... & ... & ... \\ 1 & e^{-j2\pi(\frac{M-1}{M})} & e^{-j2\pi(\frac{2(M-1)}{M})} & e^{-j2\pi(\frac{3(M-1)}{M})} & ... & e^{-j2\pi(\frac{(M-1)(M-1)}{M})} \end{bmatrix}$$

, where k varying from 1 to M-1. In the case of multiple path propagation, the unknown paths constitute the first P rows of A. We then need to select P columns to make the remaining $A_{\setminus\mathbb{P},\mathbb{K}}$ invertible.

Please note that for any columns removed, the remaining columns remaining a geometric progression, with the scaling factor $e^{-j2\pi(\frac{m}{M})}$. Since the scaling factors are unique, the Vandermonde matrix $A_{\setminus\mathbb{P},\mathbb{K}}$ is invertible. We have $\vec{Y}_{\mathbb{P}} = \mathbf{S}_{\mathbb{P}} inv(\mathbf{S}_{\mathbb{R}^*\setminus\mathbb{P},\mathbb{C}^*\setminus\mathbb{K}})\vec{Y}_{\mathbb{R}^*\setminus\mathbb{P}}$. □

B.4 Multiple Interferers

Given F interferer, we have $Y[0,0] = Y_{recv}[0,0] - \sum_f^F Y^f[0,0]$. Thus we have $\sum_f^F Y^f[0,0] = \sum_f^F \mathbf{S}_{[0,0]} inv(\mathbf{S}_{known})\vec{Y}^f_{known}$. The equation holds when $\mathbf{S}$ for all the interfering cells are the same. Since S are the same for all interferers

$$\begin{aligned}\sum_i^I Y^i[0,0] &= \sum_i^I \mathbf{S}_{[0,0]} inv(\mathbf{S}_{\mathbb{R}^*\setminus(0,0),\mathbb{C}^*\setminus(m,n)})\vec{Y}^i_{\mathbb{R}^*\setminus(0,0)} \\ &= \mathbf{S}_{[0,0]} inv(\mathbf{S}_{\mathbb{R}^*\setminus(0,0),\mathbb{C}^*\setminus(m,n)}) \sum_i^I \vec{Y}^i_{\mathbb{R}^*\setminus(0,0)}\end{aligned}$$

For the case of multi-pathing, we have

$$\begin{aligned}\sum_i^I Y^i_{\mathbb{P}} &= \sum_i^I \mathbf{S}_{\mathbb{P}} inv(\mathbf{S}_{\mathbb{R}^*\setminus\mathbb{P},\mathbb{C}^*\setminus\mathbb{K}})\vec{Y}^i_{\mathbb{R}^*\setminus\mathbb{P}} \\ &= \mathbf{S}_{\mathbb{P}} inv(\mathbf{S}_{\mathbb{R}^*\setminus\mathbb{P},\mathbb{C}^*\setminus\mathbb{K}}) \sum_i^I \vec{Y}^i_{\mathbb{R}^*\setminus\mathbb{P}}\end{aligned}$$

, where $\mathbb{P}$ denotes unknown DD domain indexes and $\mathbb{K}$ denotes known TF domain indexes.

REFERENCES

[1] OpenAirInterface. https://gitlab.eurecom.fr/oai/openairinterface5g/wikis/home, 2019.

[2] OpenEPC. https://www.corenetdynamics.com/products, 2019.

[3] Speedtest Global Index: United States. https://www.speedtest.net/global-index/united-states#market-analysis, 2020. Accessed on 07/18/2020.

[4] 3GPP. TS36.321: Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) protocol specification, Mar. 2014.

[5] 3GPP. TS36.331: Radio Resource Control (RRC), Mar. 2015.

[6] 3GPP. Study on new radio access technology Physical layer aspects, 2017.

[7] 3GPP. TR38.901: Study on channel model for frequency spectrum above 6 GHz , 2018.

[8] 3GPP. TS23.203: Policy and charging control architecture (Release 15), Sept. 2018.

[9] 3GPP. TS36.141: Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) Conformance Testing, Oct. 2019.

[10] 3GPP. TS38.104: NR; Base Station (BS) Radio Transmission and Reception, Oct. 2019.

[11] 3GPP. TS38.133: 5G NR; Requirements for support of radio resource management, Jun. 2019.

[12] 3GPP. TS38.211: 5G NR; Physical channels and modulation, Jun. 2019.

[13] 3GPP. TS38.212: 5G NR; Multiplexing and channel coding, Jun. 2019.

[14] 3GPP. TS38.213: 5G NR; Physical layer procedures for control, Jun. 2019.

[15] 3GPP. TS38.214: 5G NR; Physical layer procedures for data, Jun. 2019.

[16] 3GPP. TS38.331: 5G NR: Radio Resource Control (RRC), Jun. 2019.

[17] 3GPP. TS38.401: 5G NR; Architecture description, Jun. 2019.

[18] 3GPP. TS38.420: 5G NR; Xn general aspects and principles , Jun. 2019.

[19] 3GPP. TS28.554: 5G end to end Key Performance Indicators (KPI), Jul. 2020.

[20] 3GPP. TS32.450: Key Performance Indicators (KPI) for Evolved Universal Terrestrial Radio Access Network (E-UTRAN), Jul. 2020.

[21] 3GPP. TS38.300: 5G NR: Overall description; Stage-2, Jan. 2020.

[22] Ramoni Adeogun. A method for parameter extraction and channel state prediction in mobile-to-mobile wireless channels. 2014.

[23] Fadel Adib, Zach Kabelac, Dina Katabi, and Rob Miller. WiTrack: motion tracking via radio reflections off the body. In *Proc. of NSDI*, 2014.

[24] Mikhail Afanasov, Alessandro Djordjevic, Feng Lui, and Luca Mottola. Flyzone: A testbed for experimenting with aerial drone applications. In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*, pages 67–78, 2019.

[25] 5G America. Global 5g: Implications of a transformational technology. `https://www.5gamericas.org/wp-content/uploads/2019/09/2019-5G-Americas-Rysavy-Implications-of-a-Transformational-Technology-White-Paper.pdf`, 2019.

[26] J.G. Andrews. Interference cancellation for cellular systems: a contemporary overview. *IEEE Wireless Communications*, 12(2):19–29, 2005.

[27] Phone Arena. 5g bands cheat sheet: Verizon vs at&t vs sprint vs t-mobile vs world.

[28] Waheed U. Bajwa, Jarvis Haupt, Akbar M. Sayeed, and Robert Nowak. Compressed channel sensing: A new approach to estimating sparse multipath channels. In *Proc. of IEEE*, 2010.

[29] Bakshi, Arjun and Mao, Yifan and Srinivasan, Kannan and Parthasarathy, Srinivasan. Fast and Efficient Cross Band Channel Prediction Using Machine Learning. In *The 25th Annual International Conference on Mobile Computing and Networking (MobiCom)*, page 37. ACM, 2019.

[30] Arjun Balasingam, Manu Bansal, Rakesh Misra, Kanthi Nagaraj, Rahul Tandra, Sachin Katti, and Aaron Schulman. Detecting if LTE is the Bottleneck with BurstTracker . In *Proc. ACM MobiSys'19*, 2019.

[31] Tarun Bansal, Bo Chen, Prasun Sinha, and Kannan Srinivasan. Symphony: Cooperative packet recovery over the wired backbone in enterprise wlans. In *Proceedings of the 19th Annual International Conference on Mobile Computing and Networking*, MobiCom '13, page 351–362, New York, NY, USA, 2013. Association for Computing Machinery.

[32] Philip Bello. Characterization of randomly time-variant linear channels. *IEEE transactions on Communications Systems*, 11(4):360–393, 1963.

[33] Christian R. Berger, Zhaohui Wang, Jianzhong Huang, and Shengli Zhou. Application of compressive sensing to sparse channel estimation. *IEEE Magazine*, 2010.

[34] Bloomberg. Vehicle-to-Everything Market Communications Ecosystem 2030 Account for a Market Worth $1.2 Billion by 2022. https://www.bloomberg.com/press-releases/2019-09-23/vehicle-to-everything-v2x-market-communications-ecosystem-2030-account-for-a-market-worth-1-2-billion-by-2022, 2019.

[35] Business Wire. Global Commercial Drones Market Opportunity Analysis and Industry Forecasts, 2021-2022 & 2030. `https://www.businesswire.com/news/home/20220525005498/en/Global-Commercial-Drones-Market-Opportunity-Analysis-and-Industry-Forecasts-2021-2`
`com`, 2022.

[36] Lele Cheng, Qiuming Zhu, Cheng-Xiang Wang, Weizhi Zhong, Boyu Hua, and Shan Jiang. Modeling and simulation for uav air-to-ground mmwave channels. In *2020 14th European Conference on Antennas and Propagation (EuCAP)*, pages 1–5, 2020.

[37] Zicheng Chi, Xin Liu, Wei Wang, Yao Yao, and Ting Zhu. Leveraging ambient lte traffic for ubiquitous passive communication. In *Proceedings of the Annual Conference of the ACM Special Interest Group on Data Communication on the Applications, Technologies, Architectures, and Protocols for Computer Communication*, SIGCOMM '20, page 172–185, New York, NY, USA, 2020. Association for Computing Machinery.

[38] Jung Il Choi, Mayank Jain, Kannan Srinivasan, Philip Levis, and Sachin Katti. Achieving single channel, full duplex wireless communication. In *Proc. of ACM Mobicom*, 2010.

[39] Alcatel-Lucent Corporation. "TDD eNodeB LTE Parameters User Guide". `https://www.academia.edu/30447258/Title_page_TDD_eNodeB_LTE_Parameters_User_Guide_LTE_Parameters_User_Guide_LPUG_LPUG`.

[40] Huawei Corporation. "Intra-RAT Mobility Management in Connected Mode Feature Parameter Description". `https://www.honorcup.ru/upload/iblock/164/6.pdf`.

[41] ZTE Corporation. "ZTE UMTS handover description". `www.slideshare.net/quyetnguyenhong/zte-umtshandoverdescription`.

[42] Nicolai Czink, Markus Herdin, Hüseyin Özcelik, and Ernst Bonek. Number of multipath clusters in indoor mimo propagation environments. *Electronics letters*, 40(23):1498–1499, 2004.

[43] Haotian Deng, Qianru Li, Jingqi Huang, and Chunyi Peng. Icellspeed: Increasing cellular data speed with device-assisted cell selection. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, pages 1–13, 2020.

[44] Anolog Devices. Getting started with the ek1hmc6350 evaluation kit and software. `https://www.analog.com/media/en/technical-documentation/user-guides/EK1HMC6350-UG-1031.pdf`, 2022.

[45] Ericsson. Carrier aggregation in 5g. `https://www.ericsson.com/en/ran/carrier-aggregation`, 2021.

[46] ETSI. Multi-access Edge Computing (MEC); Framework and Reference Architecture , Jan. 2019.

[47] FAA. FAA Part 107 Fact Sheet. `https://www.nifc.gov/drones/blm/FAA%20Part%20107%20Fact%20Sheet.pdf`, 2022.

[48] Federal Communications Commission (FCC). Measuring Broadband America Mobile 2013-2018 Coarsened Data. `https://www.fcc.gov/reports-research/reports/measuring-broadband-america/measuring-broadband-america-mobile-2013-2018`, 2019.

[49] Federal Communications Commission (FCC). Measuring Mobile Broadband. `https://www.fcc.gov/general/measuring-mobile-broadband-performance`, 2020.

[50] Weidong Gao, Lin Ma, and Gang Chuai. Joint optimization of component carrier selection and resource allocation in 5g carrier aggregation system. *Physical Communication*, 25:293–297, 2017.

[51] Marco Giordani, Michele Polese, Arnab Roy, Douglas Castor, and Michele Zorzi. A tutorial on beam management for 3gpp nr at mmwave frequencies. *IEEE Communications Surveys & Tutorials*, 21(1):173–196, 2018.

[52] Shyamnath Gollakota, Fadel Adib, Dina Katabi, and Srinivasan Seshan. Clearing the rf smog: Making 802.11n robust to cross-technology interference. In *Proceedings of the ACM SIGCOMM 2011 Conference*, SIGCOMM '11, page 170–181, New York, NY, USA, 2011. Association for Computing Machinery.

[53] Shyamnath Gollakota, Samuel David Perli, and Dina Katabi. Interference alignment and cancellation. *SIGCOMM Comput. Commun. Rev.*, 39(4):159–170, aug 2009.

[54] UCLA WING Group. "Flora: Flexible Mobile Network Platform". `https://project-flora.github.io/Flora/`.

[55] Richard M. Gutierrez, Hanguang Yu, Yu Rong, and Daniel W. Bliss. Time and frequency dispersion characteristics of the uas wireless channel in residential and mountainous desert terrains. In *2017 14th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, pages 516–521, 2017.

[56] Ronny Hadani, Shlomo Rakib, Michail Tsatsanis, Anton Monk, Andrea J Goldsmith, Andreas F Molisch, and R Calderbank. Orthogonal time frequency space modulation. In *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6. IEEE, 2017.

[57] Daniel Halperin, Thomas Anderson, and David Wetherall. Taking the sting out of carrier sense: Interference cancellation for wireless lans. In *Proceedings of the 14th ACM International Conference on Mobile Computing and Networking*, MobiCom '08, page 339–350, New York, NY, USA, 2008. Association for Computing Machinery.

[58] Songtao He, Favyen Bastani, Sofiane Abbar, Mohammad Alizadeh, Hari Balakrishnan, Sanjay Chawla, and Samuel Madden. RoadRunner: Improving the Precision of Road Network Inference from GPS Trajectories. In *Proc. of SIGSPATIAL*, 2018.

[59] Ye Hu, Mingzhe Chen, and Walid Saad. Joint access and backhaul resource management in satellite-drone networks: A competitive market approach. *IEEE Transactions on Wireless Communications*, 19(6):3908–3923, 2020.

[60] Chongwen Huang, George C Alexandropoulos, Alessio Zappone, Chau Yuen, and Mérouane Debbah. Deep learning for ul/dl channel calibration in generic massive mimo systems. In *ICC 2019-2019 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE, 2019.

[61] Junxian Huang, Feng Qian, Yihua Guo, Yuanyuan Zhou, Qiang Xu, Zhuoqing Mao, Subhabrata Sen, and Oliver Spatscheck. An in-depth study of lte: effect of network protocol and application behavior on performance. pages 363–374, 08 2013.

[62] M. Jain, J. Choi, T. Kim, D. Bharadia, K. Srinivasan, P. Levis, S. Katti, P. Sinha, and S. Seth. Practical real-time full duplex wireless. In *Proc. of ACM MobiCom*, 2011.

[63] Allan Jepson and Fernando Flores-Mangas. Intro to svd. 2011.

[64] Roghayeh Joda, Medhat Elsayed, Hatem Abou-Zeid, Ramy Atawia, Akram Bin Sediq, Gary Boudreau, and Melike Erol-Kantarci. Qos-aware joint component carrier selection and resource allocation for carrier aggregation in 5g. In *ICC 2021-IEEE International Conference on Communications*, pages 1–6. IEEE, 2021.

[65] John Eidson. IEEE-1588 Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems. `https://www.nist.gov/system/files/documents/el/isd/ieee/tutorial-basic.pdf`, 2005.

[66] Kiran Joshi, Dinesh Bharadia, Manikanta Kotaru, and Sachin Katti. WiDeo: Fine-grained Device-free Motion Tracing using RF Backscatter. In *12th USENIX Symposium on Networked Systems Design and Implementation (NSDI'15)*, pages 189–204, 2015.

[67] Kiran Joshi, Dinesh Bharadia, Manikanta Kotaru, and Sachin Katti. Wideo: Fine-grained device-free motion tracing using RF backscatter. In *Proc. of NSDI*, 2015.

[68] Kaltenberger, Florian and Jiang, Haiyong and Guillaud, Maxime and Knopp, Raymond. Relative Channel Reciprocity Calibration in MIMO/TDD Systems. In *2010 Future Network & Mobile Summit*, pages 1–10. IEEE, 2010.

[69] Bryce Kellogg, Vamsi Talla, and Shyamnath Gollakota. Bringing gesture recognition to all devices. In *Proc. of NSDI*, volume 14, pages 303–316, 2014.

[70] Zeqi Lai, Y. Charlie Hu, Yong Cui, Linhui Sun, and Ningwei Dai. Furion: Engineering high-quality immersive virtual reality on today's mobile devices. In *Proc. ACM MobiCom'17*, 2017.

[71] Laura Leal-Taixé, Anton Milan, Ian Reid, Stefan Roth, and Konrad Schindler. Motchallenge 2015: Towards a benchmark for multi-target tracking. *arXiv preprint arXiv:1504.01942*, 2015.

[72] L. Erran Li, K. Tan, Y. Xu, H. Viswanathan, and Y. Richard Yang. Remap decoding: Simple retransmission permutation can resolve overlapping channel collisions. In *Proc. of ACM MOBICOM*, Sept. 2010.

[73] Qianru Li and Chunyi Peng. Reconfiguring cell selection in 4g/5g networks. In *2021 IEEE 29th International Conference on Network Protocols (ICNP)*, pages 1–11. IEEE, 2021.

[74] Tianxiang Li, Mohammad Hossein Mazaheri, and Omid Abari. 5g in the sky: The future of high-speed internet via unmanned aerial vehicles. In *Proceedings of the 23rd Annual International Workshop on Mobile Computing Systems and Applications*, HotMobile '22, page 116–122, New York, NY, USA, 2022. Association for Computing Machinery.

[75] Yi Li, Lili Qiu, Yin Zhang, Ratul Mahajan, Zifei Zhong, Gaurav Deshpande, and Eric Rozner. Effects of interference on throughput of wireless mesh networks: Pathologies and a preliminary solution. In *In Proc. of HotNets-VI*, 2007.

[76] Yuanjie Li, Hewu Li, Wei Liu, Lixin Liu, Yimei Chen, Jianping Wu, Qian Wu, Jun Liu, and Zeqi Lai. A case for stateless mobile core network functions in space. In *Proceedings of the ACM Special Interest Group on Data Communication (SIGCOMM)*. ACM, 2022.

[77] Yuanjie Li, Qianru Li, Zhehui Zhang, Ghufran Baig, Lili Qiu, and Songwu Lu. Beyond 5g: Reliable extreme mobility management. In *Proceedings of the Annual conference of the ACM Special Interest Group on Data Communication on the applications, technologies, architectures, and protocols for computer communication*, pages 344–358, 2020.

[78] Yuanjie Li, Chunyi Peng, Zengwen Yuan, Jiayao Li, Haotian Deng, and Tao Wang. Mobileinsight: Extracting and analyzing cellular network information on smartphones. In *Proc. ACM Mobicom'16*, 2016.

[79] Hong-Sheng Liao, Po-Yu Chen, and Wen-Tsuen Chen. An efficient downlink radio resource allocation with carrier aggregation in lte-advanced networks. *IEEE Transactions on Mobile Computing*, 13(10):2229–2239, 2014.

[80] Run-Fa Liao, Hong Wen, Jinsong Wu, Huanhuan Song, Fei Pan, and Lian Dong. The rayleigh fading channel prediction via deep learning. *Wireless Communications and Mobile Computing*, 2018.

[81] Lihong Liu, Hui Feng, Tao Yang, and Bo Hu. Mimo-ofdm wireless channel prediction by exploiting spatial-temporal correlation. *IEEE transactions on wireless communications*, 13(1):310–319, 2013.

[82] Luyang Liu, Ruiguang Zhong, Wuyang Zhang, Yunxin Liu, Jiansong Zhang, Lintao Zhang, and Marco Gruteser. Cutting the cord: Designing a high-quality untethered vr system with low latency remote rendering. In *Proc. ACM MobiSys'18*, 2018.

[83] Zikun Liu, Gagandeep Singh, Chenren Xu, and Deepak Vasisht. Fire: Enabling reciprocity for fdd mimo systems. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, MobiCom '21, page 628–641, New York, NY, USA, 2021. Association for Computing Machinery.

[84] Guoqiang Lu, Karl D Mann, Lisa Bostrom, Eric W Parsons, Richard Liu, Iris Bujold, Jamie Beuerman, Qingchao Liu, Damitha Ranaweera, Edward Mah, et al. Carrier aggregation scell selection for lte-a, 2017. US Patent 9,756,532.

[85] Yunfei Ma, Nicholas Selby, and Fadel Adib. Minding the billions: Ultra-wideband localization for deployed rfid tags. In *Proc. of ACM MobiCom*, 2017.

[86] Sohrab Madani, Suraj Jog, Jesus O. Lacruz, Joerg Widmer, and Haitham Hassanieh. Practical null steering in millimeter wave networks. In *18th USENIX Symposium on Networked Systems Design and Implementation (NSDI 21)*, pages 903–921. USENIX Association, April 2021.

[87] MATLAB. Channel Estimation. `https://www.mathworks.com/help/lte/ug/channel-estimation.html`, 2018.

[88] George Matsaglia and George PH Styan. Equalities and inequalities for ranks of matrices. *Linear and multilinear Algebra*, 2(3):269–292, 1974.

[89] MIT Technology Review. Moscow Billboard Targets Ads Based on the Car You are Driving. https://www.technologyreview.com/s/603743/moscow-billboard-targets-ads-based-on-the-car-youre-driving/, Mar. 2017.

[90] L Srikar Muppirisetty, Tommy Svensson, and Henk Wymeersch. Spatial wireless channel prediction under location uncertainty. *IEEE Transactions on Wireless Communications*, 15(2):1031–1044, 2015.

[91] Rajalakshmi Nandakumar, Vikram Iyer, Desney Tan, and Shyamnath Gollakota. FingerIO: Using active sonar for fine-grained finger tracking. In *Proc. of ACM CHI*, pages 1515–1525, 2016.

[92] Arvind Narayanan, Xumiao Zhang, Ruiyang Zhu, Ahmad Hassan, Shuowei Jin, Xiao Zhu, Xiaoxuan Zhang, Denis Rybkin, Zhengxuan Yang, Zhuoqing Morley Mao, Feng Qian, and Zhi-Li Zhang. A variegated look at 5g in the wild: Performance, power, and qoe implications. In *Proceedings of the 2021 ACM SIGCOMM 2021 Conference*, SIGCOMM '21, page 610–625, New York, NY, USA, 2021. Association for Computing Machinery.

[93] NVIDIA. Nvidia grid cloud gaming platform. `http://www.nvidia.com/object/cloud-gaming.html`, 2017.

[94] Ookla. United States: Q1-Q2 2019. `https://www.speedtest.net/reports/united-states/`, 2019. Accessed on 07/18/2020.

[95] OpenSignal. USA: Mobile Network Experience Report July 2019. `https://www.opensignal.com/reports/2019/07/usa/mobile-network-experience`, 2019. Accessed on 08/24/2019.

[96] OpenSignal. USA: Mobile Network Experience Report January 2020. `https://www.opensignal.com/reports/2020/01/usa/mobile-network-experience`, 2020. Accessed on 03/28/2020.

[97] OpenSignal. USA: Mobile Network Experience Report July 2020. `https://www.opensignal.com/reports/2020/07/usa/mobile-network-experience`, 2020. Accessed on 07/16/2020.

[98] Guillermo Pocovi, Ilaria Thibault, Troels Kolding, Mads Lauridsen, Rame Canolli, Nick Edwards, and David Lister. On the suitability of lte air interface for reliable low-latency applications. In *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6. IEEE, 2019.

[99] Ravikumar V Pragada, Samian Kaur, Douglas R Castor, Philip J Pietraski, and Arnab Roy. Method and apparatus for integrating different radio access technologies using carrier aggregation, 2016. US Patent 9,294,926.

[100] The Linux Foundation Project. "Magma: A modern mobile core network solution". `https://magmacore.org/`.

[101] Qifan Pu, Sidhant Gupta, Shyam Gollakota, and Shwetak Patel. Whole-home gesture recognition using wireless signals. In *Proc. of ACM MobiCom*, 2013.

[102] Feng Qian, Bo Han, Qingyang Xiao, and Vijay Gopalakrishnan. Flare: Practical viewport-adaptive 360-degree video streaming for mobile devices. In *Proc. ACM MobiCom'18*, 2018.

[103] Lili Qiu, Yin Zhang, Feng Wang, Mi Kyung Han, and Ratul Mahajan. A general model of wireless interference. In *In Proc. of ACM MOBICOM*, 2007.

[104] Lenin S. Ravindranath, Calvin Newport, Hari Balakrishnan, and Samuel Madden. Improving Wireless Network Performance Using Sensor Hints . In *Proc. of NSDI*, Boston, MA, March 2011.

[105] Patchava Raviteja, Khoa T Phan, and Yi Hong. Embedded pilot-aided channel estimation for otfs in delay–doppler channels. *IEEE transactions on vehicular technology*, 68(5):4906–4917, 2019.

[106] Patchava Raviteja, Khoa T Phan, Yi Hong, and Emanuele Viterbo. Interference cancellation and iterative detection for orthogonal time frequency space modulation. *IEEE Transactions on Wireless Communications*, 17(10):6501–6515, 2018.

[107] Souvik Sen, Naveen Santhapuri, Romit Roy Choudhury, and Srihari Nelakuditi. Successive interference cancellation: A back-of-the-envelope perspective. In *Proceedings of the 9th ACM SIGCOMM Workshop on Hot Topics in Networks*, Hotnets-IX, New York, NY, USA, 2010. Association for Computing Machinery.

[108] Wenqian Shen, Linglong Dai, Jianping An, Pingzhi Fan, and Robert W Heath. Channel estimation for orthogonal time frequency space (otfs) massive mimo. *IEEE Transactions on Signal Processing*, 67(16):4204–4217, 2019.

[109] Shu Shi, Varun Gupta, Micheal Hwang, and Rittwik Jana. Mobile VR on Edge Cloud: A Latency-Driven Design. In *Proc. ACM MMSys'19*, 2019.

[110] D. Shutin and G. Kubin. Tracking and prediction of multipath components in wireless mimo channels. In *Proc. of VTC*, pages 329–333, 2007.

[111] L. Sun, S. Sen, D. Koutsonikolas, and K. Kim. WiDraw: enabling hands-free drawing in the air on commodity WiFi devices. In *Proc. of ACM MobiCom*, 2015.

[112] GD Surabhi, Rose Mary Augustine, and Ananthanarayanan Chockalingam. Peak-to-average power ratio of otfs modulation. *IEEE Communications Letters*, 23(6):999–1002, 2019.

[113] Zhaowei Tan, Yuanjie Li, Qianru Li, Zhehui Zhang, Zhehan Li, and Songwu Lu. Enabling mobile vr in lte networks: How close are we? In *ACM SIGMETRICS'18*, 2018.

[114] Zhaowei Tan, Yuanjie Li, Qianru Li, Zhehui Zhang, Zhehan Li, and Songwu Lu. Supporting mobile vr in lte networks: How close are we? *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 2(1):1–31, 2018.

[115] Yanchao Tang. The research on lte coverage solutions on high-speed railway. *Designing Techniques of Posts and Telecommunications*, 12:20–23, 2014.

[116] Space Exploration Technologies. Spacex non-geostationary satellite system attachment a: technical information to supplement schedule s. `https://licensing.fcc.gov/myibfs/download.do?attachment_key=1158350`, Nov. 2016.

[117] Telecom Infra Project. OpenRAN. `https://telecominfraproject.com/openran/`, 2022.

[118] TNW. Bezos and Musk's satellite internet could save Americans $30B a year. `https://thenextweb.com/podium/2019/08/24/bezos-and-musks-satellite-internet-could-save-americans-30b-a-year/`, Aug. 2019.

[119] David Tse and Pramod Viswanath. *Fundamentals of wireless communication*. Cambridge university press, 2005.

[120] Vasisht, Deepak and Kumar, Swarun and Rahul, Hariharan and Katabi, Dina. Eliminating Channel Feedback in Next-Generation Cellular Networks. In *Proceedings of the 2016 ACM SIGCOMM Conference*, pages 398–411. ACM, 2016.

[121] Cheng-Xiang Wang, Ammar Ghazal, Bo Ai, Yu Liu, and Pingzhi Fan. Channel measurements and models for high-speed train communication systems: A survey. *IEEE communications surveys & tutorials*, 18(2):974–987, 2015.

[122] Jing Wang, Yufan Zheng, Yunzhe Ni, Chenren Xu, Feng Qian, Wangyang Li, Wantong Jiang, Yihua Cheng, Zhuo Cheng, Yuanjie Li, et al. An active-passive measurement study of tcp performance over lte on high-speed rails. In *The 25th Annual International Conference on Mobile Computing and Networking*, pages 1–16, 2019.

[123] Jue Wang, Deepak Vasisht, and Dina Katabi. RF-IDraw: virtual touch screen in the air using RF signals. In *Proc. of ACM SIGCOMM*, 2014.

[124] Wei Wang, Alex X Liu, and Ke Sun. Device-free gesture tracking using acoustic signals. In *Proc. of MobiCom*, pages 82–94, 2016.

[125] Teng Wei and Xinyu Zhang. mTrack: high precision passive tracking using millimeter wave radios. In *Proc. of ACM MobiCom*, 2015.

[126] Teng Wei, Anfu Zhou, and Xinyu Zhang. Facilitating Robust 60GHz Network Deployment by Sensing Ambient Reflectors. In *14th USENIX Symposium on Networked Systems Design and Implementation (NSDI'17)*, pages 213–226, 2017.

[127] Wikipedia. Orthogonal frequency division multiplexing. `http://en.wikipedia.org/wiki/Orthogonal_frequency-division_multiplexing`.

[128] Wikipedia. Coherence time (communications systems). `https://en.wikipedia.org/wiki/Coherence_time_(communications_systems)`, 2019.

[129] Wikipedia. High-speed rail in China. `https://en.wikipedia.org/wiki/High-speed_rail_in_China`, 2019.

[130] Ian C Wong and Brian L Evans. Spc03-4: Exploiting spatio-temporal correlations in mimo wireless channel prediction. In *Proc. of IEEE Globecom*, pages 1–5, 2006.

[131] Lixia Xiao, Shuo Li, Ying Qian, Da Chen, and Tao Jiang. An overview of otfs for internet of things: Concepts, benefits and challenges. *IEEE Internet of Things Journal*, 2021.

[132] Hang Yu, Oscar Bejarano, and Lin Zhong. Combating inter-cell interference in 802.11ac-based multi-user mimo networks. In *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*, MobiCom '14, page 141–152, New York, NY, USA, 2014. Association for Computing Machinery.

[133] Sangki Yun, Yi chao Chen, and Lili Qiu. Turning a mobile device into a mouse in the air. In *Proc. of ACM MobiSys*, May 2015.

[134] Sangki Yun, Yi-Chao Chen, Huihuang Zheng, Lili Qiu, and Wenguang Mao. Strata: Fine-grained acoustic-based device-free tracking. In *Proc. of MobiSys*, 2017.

[135] Lili Zhang, Richard Stirling-Gallacher, and Guorong Li. Cell measurement reporting method and user equipment, 2020. US Patent 10,616,815.

[136] Wenjie Zhou, Tarun Bansal, Prasun Sinha, and Kannan Srinivasan. Bbn: Throughput scaling in dense enterprise wlans with bind beamforming and nulling. In *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*, MobiCom '14, page 165–176, New York, NY, USA, 2014. Association for Computing Machinery.

[137] Wenjie Zhou, Tanmoy Das, Lu Chen, Kannan Srinivasan, and Prasun Sinha. Basic: Backbone-assisted successive interference cancellation. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, MobiCom '16, page 149–161, New York, NY, USA, 2016. Association for Computing Machinery.

[138] Yanzi Zhu, Yibo Zhu, Ben Y. Zhao, and Haitao Zheng. Reusing 60 GHz radios for mobile radar imaging. In *Proc. of ACM MobiCom*, 2015.