

UNIVERSITY OF PADUA  
DEPARTMENT OF INFORMATION  
ENGINEERING – DEI



NATIONAL TAIWAN UNIVERSITY  
GRADUATE INSTITUTE OF COMMUNICATION  
ENGINEERING – GICE

Master's Thesis for the Double Degree Programme in  
**ICT for Internet and Multimedia**  
*and*  
**Communication Engineering**

**Scalable Cell-Free ISAC for Drone Surveillance  
with Hybrid Signaling and Dynamic TDD**

CANDIDATE

**Matteo Peri**

*Student ID (Unipd): 2156345*

*Student ID (NTU): R14942151*

SUPERVISOR

**Prof. Michele Zorzi**

*University of Padua*

CO-SUPERVISOR

**Prof. Hsuan-Jung Su**

*National Taiwan University*

ACADEMIC YEAR 2025–2026

09 JUNE 2026



國立臺灣大學電機資訊學院電信工程學研究所  
碩士論文

Graduate Institute of Communication Engineering  
College of Electrical Engineering and Computer Science  
National Taiwan University  
Master's Thesis

應用於無人機監控之可擴展無蜂巢式整合感測與通訊：結合混  
合信號與動態分時雙工

Scalable Cell-Free ISAC for Drone Surveillance with Hybrid  
Signaling and Dynamic TDD

黎宇紘

Matteo Peri

指導教授：蘇炫榮 博士

Advisor: Hsuan-Jung Su, Ph.D.

中華民國 115 年 06 月

June 2026



國立臺灣大學碩士學位論文  
口試委員會審定書  
MASTER'S THESIS ACCEPTANCE CERTIFICATE  
NATIONAL TAIWAN UNIVERSITY

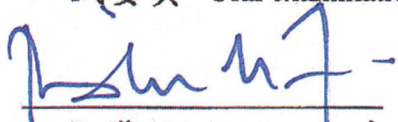
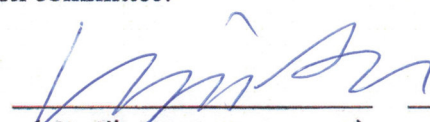
應用於無人機監控之可擴展無蜂巢式整合感測與通訊：結合  
混合信號與動態分時雙工

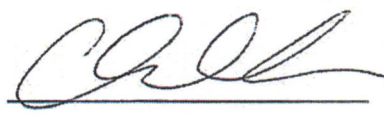
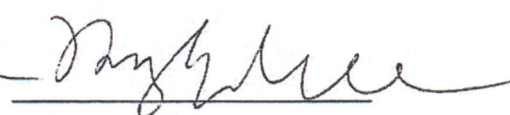
Scalable Cell-Free ISAC for Drone Surveillance with Hybrid  
Signaling and Dynamic TDD

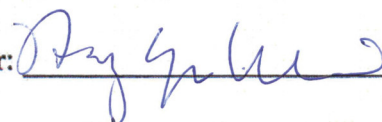
本論文係黎宇紘 (姓名) R14942151 (學號) 在國立臺灣大學電信工程學研究所完成之碩士學位論文，於民國115 年06 月09 日承下列考試委員審查通過及口試及格，特此證明。

The undersigned, appointed by the Graduate Institute of Communication Engineering on 09 (date) 06 (month) 2026 (year) have examined a Master's Thesis entitled above presented by Matteo Peri (name) R14942151 (student ID) candidate and hereby certify that it is worthy of acceptance.

口試委員 Oral examination committee:

   
(指導教授 Advisor) (指導教授 Advisor)

所長 Director: 





## ACKNOWLEDGMENTS

First and foremost, I would like to thank the University of Padua for supporting my academic journey and giving me the opportunity to pursue the Dual Degree programme with National Taiwan University. I am deeply grateful to my supervisors, Professor Hsuan-Jung Su and Professor Michele Zorzi, for their guidance and support throughout my master's studies and this thesis. I would also like to thank National Taiwan University for hosting me and for providing me with such a meaningful academic experience. During my time here, I learned from a new perspective and collaborated with students and researchers from diverse backgrounds. In particular, I would like to thank my friends and colleagues at the CSIP Lab for their support and for creating such a welcoming environment. This campus and Taiwan have captured my heart, and I am truly grateful for the friendships and experiences I have made here, which I hope will remain with me for many years to come.

謝謝臺灣，謝謝你們。

Vorrei poi ringraziare i miei genitori, Luca e Paola, mio fratello Andrea e tutta la mia famiglia, per il loro amore e supporto incondizionati durante tutto il mio percorso di studi, soprattutto in questo lungo periodo di lontananza da casa. Un ringraziamento speciale va anche ai miei amici: Alice, Anna, Edoardo, Francesca e Rita, con cui ho condiviso tanti ricordi della vita universitaria; Andrea, Daniele, Fabio, Manuel, Marialucia, Rei e tanti altri, che mi sono accanto da anni; e Matej, Nela e Yi Rong, per i tanti bei momenti condivisi, i viaggi e il tempo trascorso insieme durante il periodo all'estero. Grazie a tutti voi per aver reso questo percorso così speciale e indimenticabile, e per avermi sostenuto anche nei momenti più difficili.





## 摘要

未來第六代（6G）無線網路預期將同時支援可靠通訊與原生感測，使基於無蜂巢式（Cell-Free）大規模多輸入多輸出（Massive MIMO）架構的整合式感測與通訊（Integrated Sensing and Communication, ISAC），成為低空無人機監視的一種具前景的架構。此任務可視為未來系統中一項具體且高度要求的應用情境：網路必須監控三維空域、偵測小型且具高機動性的空中目標，並在持續服務通訊使用者的同時維持態勢感知。此應用也促使本研究超越一般化的感測問題表述，因為可靠的監視任務需要體素層級搜尋、多基地可視性、受限的前傳負載，以及在通訊限制下運作的系統設計。

然而，將分散式多基地感測與以使用者為中心的無蜂巢式運作結合，會在可擴展性、雙工機制、前傳負載與交叉鏈路干擾（Cross-Link Interference）方面帶來重大挑戰。當感測必須整合至通訊原生協定中，且不能依賴不可擴展的集中式處理或過高的訊令開銷時，這些挑戰將變得更加關鍵。

本論文提出一種基於動態分時雙工（Dynamic TDD, DTDD）的可擴展無蜂巢式 ISAC 架構。在此架構中，分散式接取點（Access Points）會被動態指派為下行照射角色，或上行通訊與感測角色，且感測任務會在明確的通訊限制下執行。所提出的架構結合了混合式下行訊號傳輸策略、有界複雜度的排程與控制架構，以及可擴展的上行接收與分散式同調偵測流程。具體而言，本系統的設計目標是在任何時刻皆維持通訊可行性，同時利用剩餘的空間與功率資源，在受監控空域中支援可靠的體素式無人機監視。

整體而言，本論文顯示，多基地無人機監視能夠以具備明確運作流程與結構可擴展性的方式嵌入無蜂巢式無線網路中，同時保留協同處理、通訊原生探測與分散式感測多樣性的主要優點。研究結果支持以下觀點：無蜂巢式 ISAC 能夠超越概念性表述，並進一步發展為下一代無線監視系統中實用的協定層級架構。

**關鍵詞：** 無蜂巢式 ISAC、整合式感測與通訊、無蜂巢式大規模多輸入多輸出、無人機監視、動態分時雙工、多基地感測、分散式同調偵測、可擴展無線網路、第六代無線網路





## ABSTRACT

Future 6G wireless networks are expected to support both reliable communication and native sensing, making Integrated Sensing and Communication (ISAC) over Cell-Free Massive MIMO a promising architecture for low-altitude drone surveillance. This task is a concrete and demanding mission for future systems: the network must monitor a three-dimensional airspace, detect small and mobile aerial targets, and maintain situational awareness while continuing to serve communication users. This application motivates moving beyond generic sensing formulations, because reliable surveillance requires voxel-wise search, multi-static visibility, bounded fronthaul, and communication-constrained operation.

However, combining distributed multi-static sensing with user-centric Cell-Free operation raises major challenges in scalability, duplexing, fronthaul, and Cross-Link Interference. These challenges become even more critical when sensing must be integrated into communication-native protocols without relying on unscalable centralized processing or excessive signaling overhead.

This thesis proposes a scalable Cell-Free ISAC framework based on Dynamic Time Division Duplex (DTDD), in which distributed Access Points are dynamically assigned to downlink illumination or uplink communication-and-sensing roles, and sensing is performed under explicit communication constraints. The proposed framework combines a hybrid downlink signaling strategy, a bounded-complexity scheduling and control architecture, and a scalable uplink reception and distributed coherent detection pipeline. In particular, the system is designed so that communication feasibility is preserved at all times, while the remaining spatial and power resources are exploited to support reliable voxel-based drone surveillance over the monitored airspace.

Overall, the thesis shows that multi-static drone surveillance can be embedded into Cell-Free wireless networks in a way that is both operationally explicit and structurally scalable, while preserving the main benefits of coordinated processing, communication-native probing, and distributed sensing diversity. The results support the view that Cell-Free ISAC can move beyond conceptual formulations and be developed as a practical protocol-level architecture for next-generation wireless surveillance systems.

**Keywords:** Cell-Free ISAC, Integrated Sensing and Communication, Cell-Free Massive MIMO, Drone Surveillance, Dynamic TDD, Multi-Static Sensing, Distributed Coherent Detection, Scalable Wireless Networks, 6G

# CONTENTS



|                                                                     |      |
|---------------------------------------------------------------------|------|
| MASTER'S THESIS ACCEPTANCE CERTIFICATE                              | v    |
| ACKNOWLEDGMENTS                                                     | vii  |
| 摘要                                                                  | ix   |
| ABSTRACT                                                            | xi   |
| CONTENTS                                                            | xiii |
| LIST OF FIGURES                                                     | xix  |
| LIST OF TABLES                                                      | xx   |
| LIST OF ALGORITHMS                                                  | xx   |
| LIST OF ACRONYMS                                                    | xxi  |
| 1 INTRODUCTION                                                      | 1    |
| 1.1 Historical evolution of mobile connectivity . . . . .           | 1    |
| 1.2 Future challenges and opportunities . . . . .                   | 2    |
| 1.2.1 Integrated Sensing and Communication . . . . .                | 3    |
| 1.2.2 Cell-Free Massive MIMO as an enabling architecture . . . . .  | 4    |
| 1.3 Drone surveillance as a 6G mission profile . . . . .            | 4    |
| 1.3.1 Cell-Free ISAC for Drone Surveillance . . . . .               | 5    |
| 1.3.2 Application scope and Best-Effort Sensing rationale . . . . . | 5    |
| 1.4 Problem statement . . . . .                                     | 7    |
| 1.4.1 Notation and Main symbols . . . . .                           | 8    |
| 1.5 Related works . . . . .                                         | 9    |
| 1.5.1 User-centric Cell-Free Massive MIMO . . . . .                 | 10   |
| 1.5.2 Cell-Free ISAC for Target Detection . . . . .                 | 10   |

|       |                                                                      |    |
|-------|----------------------------------------------------------------------|----|
| 1.5.3 | DTDD and InAI Management in CF Systems . . . . .                     | 12 |
| 1.5.4 | Voxelized sensing and Drone Surveillance in ISAC . . . . .           | 13 |
| 1.5.5 | Hybrid waveforms and Deterministic sensing signals in ISAC . . . . . | 14 |
| 1.5.6 | Summary . . . . .                                                    | 14 |
| 1.6   | Thesis contributions . . . . .                                       | 15 |
| 1.6.1 | Primary novelties . . . . .                                          | 15 |
| 1.6.2 | Adaptations and Extensions of Prior Work . . . . .                   | 17 |
| 1.6.3 | Standard methodology . . . . .                                       | 18 |
| 1.7   | Thesis organization . . . . .                                        | 19 |

## I THEORETICAL FOUNDATIONS OF CELL-FREE ISAC 21

## 2 MASSIVE MIMO AND CELL-FREE FUNDAMENTALS FOR ISAC 23

|       |                                                       |    |
|-------|-------------------------------------------------------|----|
| 2.1   | System Model . . . . .                                | 24 |
| 2.1.1 | Channel Hardening and Favorable Propagation . . . . . | 25 |
| 2.1.2 | TDD Operation and Pilot Transmission . . . . .        | 26 |
| 2.1.3 | Channel Estimation . . . . .                          | 28 |
| 2.2   | Pilot contamination . . . . .                         | 28 |
| 2.3   | Linear processing for uplink and downlink . . . . .   | 30 |
| 2.3.1 | MR Precoding . . . . .                                | 31 |
| 2.3.2 | ZF Precoding . . . . .                                | 31 |
| 2.3.3 | RZF / MMSE Precoding . . . . .                        | 32 |
| 2.4   | Performance Metrics . . . . .                         | 33 |
| 2.4.1 | SINR and achievable Spectral Efficiency . . . . .     | 33 |
| 2.4.2 | Spectral Efficiency gains . . . . .                   | 35 |
| 2.4.3 | Quality of Service constraints . . . . .              | 35 |
| 2.5   | Cell-Free concept and cellular networks . . . . .     | 36 |
| 2.5.1 | Cell-Free operation . . . . .                         | 36 |
| 2.6   | User-centric clustering and AP selection . . . . .    | 37 |
| 2.7   | Centralized and distributed processing . . . . .      | 39 |

|          |                                                           |           |
|----------|-----------------------------------------------------------|-----------|
| 2.7.1    | Centralized processing . . . . .                          | 39        |
| 2.7.2    | Distributed processing . . . . .                          | 40        |
| 2.7.3    | Hybrid processing . . . . .                               | 40        |
| 2.8      | Scalable Cell-Free protocols . . . . .                    | 41        |
| 2.9      | Sensing in CF Architectures . . . . .                     | 41        |
| <b>3</b> | <b>RADAR AND SENSING FUNDAMENTALS</b>                     | <b>43</b> |
| 3.1      | Sensing tasks . . . . .                                   | 44        |
| 3.2      | Detection theory . . . . .                                | 44        |
| 3.3      | Estimation basics . . . . .                               | 46        |
| 3.4      | Clutter and interference . . . . .                        | 48        |
| 3.5      | Multi-static sensing basics . . . . .                     | 49        |
| 3.5.1    | Monostatic Radar . . . . .                                | 49        |
| 3.5.2    | Bistatic Radar . . . . .                                  | 49        |
| 3.5.3    | Multi-static Radar . . . . .                              | 50        |
| 3.6      | Performance metrics . . . . .                             | 51        |
| 3.6.1    | Radar Cross Section . . . . .                             | 51        |
| 3.6.2    | SNR and SCNR . . . . .                                    | 52        |
| 3.6.3    | Resolution . . . . .                                      | 52        |
| 3.6.4    | Other Metrics . . . . .                                   | 53        |
| <b>4</b> | <b>ISAC AND CELL-FREE ISAC</b>                            | <b>55</b> |
| 4.1      | ISAC paradigms . . . . .                                  | 56        |
| 4.2      | Frame structure for ISAC . . . . .                        | 57        |
| 4.3      | Payload-aided sensing . . . . .                           | 58        |
| 4.4      | Sensing metrics landscape . . . . .                       | 60        |
| 4.4.1    | Sensing Mutual Information and Sensing Capacity . . . . . | 60        |
| 4.4.2    | MSE-Based Metrics: CRB and E-LMMSE . . . . .              | 61        |
| 4.4.3    | SNR, SINR, and SCNR proxies . . . . .                     | 61        |
| 4.5      | Communication-sensing trade-off . . . . .                 | 61        |

|                                                             |                                                                |           |
|-------------------------------------------------------------|----------------------------------------------------------------|-----------|
| 4.6                                                         | Cell-Free ISAC system models . . . . .                         | 63        |
| 4.6.1                                                       | Multi-static target detection . . . . .                        | 64        |
| 4.7                                                         | Duplexing options in CF-ISAC . . . . .                         | 66        |
| 4.7.1                                                       | TDD Baseline . . . . .                                         | 66        |
| 4.7.2                                                       | Dynamic Time Division Duplex (DTDD) . . . . .                  | 66        |
| 4.8                                                         | Scalability gaps in existing work . . . . .                    | 68        |
| <br><b>II PROPOSED SCALABLE CELL-FREE ISAC ARCHITECTURE</b> |                                                                | <b>71</b> |
| <br><b>5 PROBLEM FORMULATION AND DESIGN PHILOSOPHY</b>      |                                                                | <b>73</b> |
| 5.1                                                         | Motivation . . . . .                                           | 73        |
| 5.2                                                         | System setting . . . . .                                       | 75        |
| 5.3                                                         | CF-ISAC scalability criterion . . . . .                        | 77        |
| 5.4                                                         | Communication-constrained sensing problem . . . . .            | 78        |
| 5.4.1                                                       | Best-effort sensing . . . . .                                  | 79        |
| 5.4.2                                                       | Solving philosophy and procedure . . . . .                     | 80        |
| <br><b>6 SCHEDULING AND CONTROL ARCHITECTURE</b>            |                                                                | <b>85</b> |
| 6.1                                                         | Role of scheduling in the proposed protocol . . . . .          | 85        |
| 6.2                                                         | Scan semantics . . . . .                                       | 86        |
| 6.2.1                                                       | Adaptive coarse/fine scan logic . . . . .                      | 87        |
| 6.2.2                                                       | Track state, support-preserving remap, and bridge requests . . | 89        |
| 6.2.3                                                       | Bridge-capped requested-voxel scheduler . . . . .              | 91        |
| 6.3                                                         | UE and voxel neighborhoods . . . . .                           | 92        |
| 6.4                                                         | DTDD AP scheduling . . . . .                                   | 94        |
| 6.5                                                         | Pilot coordination . . . . .                                   | 94        |
| 6.6                                                         | Voxel-sequence planning . . . . .                              | 96        |
| 6.6.1                                                       | Deterministic sensing sequence . . . . .                       | 96        |
| 6.6.2                                                       | Coverage-balanced candidate distribution . . . . .             | 98        |
| 6.7                                                         | UL sensing assignment . . . . .                                | 100       |

|          |                                                             |            |
|----------|-------------------------------------------------------------|------------|
| 6.8      | Fronthaul and computational scalability . . . . .           | 102        |
| <b>7</b> | <b>HYBRID DOWNLINK SIGNAL DESIGN</b>                        | <b>103</b> |
| 7.1      | Downlink design objectives . . . . .                        | 104        |
| 7.2      | Protected local payload precoding . . . . .                 | 105        |
| 7.2.1    | Scalar feedback and CPU-tracked payload power loading . . . | 107        |
| 7.2.2    | Protection trade-offs and applicability . . . . .           | 108        |
| 7.3      | Hybrid voxel admission and probing split . . . . .          | 110        |
| 7.4      | Dedicated sensing on remaining voxels . . . . .             | 113        |
| 7.4.1    | Effective illumination scalar returned to the CPU . . . . . | 116        |
| 7.5      | Hybrid downlink transmit signal . . . . .                   | 117        |
| 7.6      | Downlink reception model and metrics . . . . .              | 118        |
| 7.6.1    | Exact instantaneous DL SINR . . . . .                       | 118        |
| 7.7      | Fronthaul and computational scalability . . . . .           | 119        |
| <b>8</b> | <b>UPLINK RECEPTION AND COHERENT DETECTION</b>              | <b>123</b> |
| 8.1      | Uplink received signal under DTDD operation . . . . .       | 124        |
| 8.2      | Distributed uplink protocol . . . . .                       | 126        |
| 8.2.1    | Protection trade-offs and applicability . . . . .           | 129        |
| 8.2.2    | CPU soft fusion for uplink communication . . . . .          | 129        |
| 8.3      | Centralized uplink protocol . . . . .                       | 130        |
| 8.4      | Voxel observations . . . . .                                | 132        |
| 8.4.1    | Rank-one coherent voxel model . . . . .                     | 134        |
| 8.5      | Distributed GLRT . . . . .                                  | 135        |
| 8.5.1    | Detection threshold and sensing metrics . . . . .           | 136        |
| 8.6      | Centralized GLRT . . . . .                                  | 138        |
| 8.7      | Uplink SINR and Spectral Efficiency . . . . .               | 142        |
| 8.8      | Fronthaul and computational scalability . . . . .           | 144        |
| 8.8.1    | Distributed uplink and distributed GLRT . . . . .           | 144        |
| 8.8.2    | Centralized uplink and centralized GLRT . . . . .           | 145        |

|       |                                                          |     |
|-------|----------------------------------------------------------|-----|
| 9     | NUMERICAL RESULTS                                        | 149 |
| 9.1   | Simulation methodology . . . . .                         | 151 |
| 9.1.1 | Surveillance target model . . . . .                      | 151 |
| 9.1.2 | Protocol variants and receiver split . . . . .           | 153 |
| 9.1.3 | Communication and sensing metrics . . . . .              | 154 |
| 9.2   | Sensing-communication trade-off . . . . .                | 155 |
| 9.2.1 | Communication threshold satisfaction . . . . .           | 158 |
| 9.3   | Trade-off under communication load . . . . .             | 161 |
| 9.4   | Residual inter-AP interference protection . . . . .      | 164 |
| 9.4.1 | Downlink precoder sweep . . . . .                        | 164 |
| 9.4.2 | Uplink combiner sweep . . . . .                          | 164 |
| 9.4.3 | Trade-off analysis and operating conditions . . . . .    | 165 |
| 9.5   | Performance over different DL and UL splits . . . . .    | 166 |
| 9.6   | Detection probability over target strength . . . . .     | 169 |
| 9.7   | Track retention under mobility and density . . . . .     | 171 |
| 9.7.1 | Retention under the fixed-heading model . . . . .        | 171 |
| 9.7.2 | Comparison with the random-heading model . . . . .       | 173 |
| 9.8   | Fixed hardware budget . . . . .                          | 175 |
| 10    | CONCLUSIONS AND FUTURE WORK                              | 179 |
| 10.1  | Main conclusions . . . . .                               | 179 |
| 10.2  | Limitations of the present framework . . . . .           | 181 |
| 10.3  | Future work . . . . .                                    | 183 |
|       | BIBLIOGRAPHY                                             | 187 |
| A     | ROC CURVES FOR PROTOCOL VARIANTS                         | 191 |
| B     | ADAPTIVE GRID CONVERGENCE                                | 195 |
| B.1   | Convergence from different initializations . . . . .     | 195 |
| B.2   | Steady-state grid size versus effective budget . . . . . | 197 |

# LIST OF FIGURES



|            |                                                                |     |
|------------|----------------------------------------------------------------|-----|
| Figure 2.1 | Massive MIMO processing pipeline . . . . .                     | 25  |
| Figure 2.2 | Channel hardening and favorable propagation . . . . .          | 27  |
| Figure 2.3 | Typical TDD frame structure . . . . .                          | 27  |
| Figure 2.4 | Effects of pilot contamination on uplink SINR . . . . .        | 30  |
| Figure 2.5 | Comparison of beamforming strategies . . . . .                 | 34  |
| Figure 2.6 | Cell-Free Massive MIMO system architecture . . . . .           | 38  |
| Figure 2.7 | User-centric clustering in Cell-Free Massive MIMO . . . . .    | 39  |
| Figure 2.8 | Processing split in different CF Massive MIMO architectures    | 42  |
| Figure 3.1 | Radar sensing pipeline from received signal to target tracking | 45  |
| Figure 3.2 | ROC curve of coherent Neyman-Pearson detection in AWGN .       | 47  |
| Figure 3.3 | Monostatic, bistatic, and multistatic radar geometries . . . . | 50  |
| Figure 4.1 | Frame-structure gallery for ISAC . . . . .                     | 59  |
| Figure 4.2 | Pareto boundary of the communication-sensing tradeoff . . .    | 63  |
| Figure 4.3 | Cell-Free ISAC multi-static architecture . . . . .             | 64  |
| Figure 4.4 | Duplexing timelines in CF-ISAC . . . . .                       | 67  |
| Figure 5.1 | Protocol execution architecture . . . . .                      | 82  |
| Figure 6.1 | Control architecture of the scheduling layer . . . . .         | 87  |
| Figure 8.1 | Distributed UL receiver architecture . . . . .                 | 137 |
| Figure 9.1 | Representative simulation scenario . . . . .                   | 152 |
| Figure 9.2 | Sensing-communication trade-off over SINR threshold . . . .    | 156 |
| Figure 9.3 | Communication SINR-threshold satisfaction . . . . .            | 159 |
| Figure 9.4 | Trade-off under increasing communication load . . . . .        | 162 |
| Figure 9.5 | Effect of protected DL precoding under residual InAI . . . .   | 165 |
| Figure 9.6 | Effect of protected UL combining under residual InAI . . . .   | 166 |

|             |                                                           |     |
|-------------|-----------------------------------------------------------|-----|
| Figure 9.7  | Communication SE and track retention over the UL/DL split | 168 |
| Figure 9.8  | Detection probability over RCS                            | 170 |
| Figure 9.9  | Target-retention heatmap under motion and density         | 172 |
| Figure 9.10 | Performance over a fixed total antenna budget             | 176 |
| Figure A.1  | Full ROC curves for the protocol variants                 | 192 |
| Figure B.1  | Adaptive-grid convergence and steady-state budget law     | 196 |

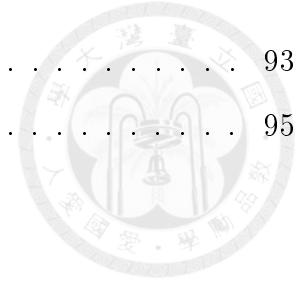
## LIST OF TABLES

|           |                                                   |     |
|-----------|---------------------------------------------------|-----|
| Table 1.1 | Evolution of mobile networks                      | 3   |
| Table 1.2 | Main system and control notation                  | 9   |
| Table 1.3 | Feature comparison with closely related works     | 15  |
| Table 2.1 | Illustrative coherence examples                   | 25  |
| Table 3.1 | Typical RCS values for various targets            | 53  |
| Table 6.1 | Dominant online cost of the scheduling layer.     | 102 |
| Table 7.1 | Hybrid layers of the proposed downlink design     | 104 |
| Table 8.1 | Scalability summary of the UL and detection layer | 147 |
| Table 9.1 | Protocol variants used in the results             | 154 |
| Table 9.2 | Retention difference by drone speed               | 175 |
| Table 9.3 | Retention difference by drone count               | 175 |

## LIST OF ALGORITHMS

|               |                                  |    |
|---------------|----------------------------------|----|
| Algorithm 6.1 | Adaptive coarse/fine grid update | 90 |
|---------------|----------------------------------|----|

|               |                                  |    |
|---------------|----------------------------------|----|
| Algorithm 6.2 | Scan request scheduler . . . . . | 93 |
| Algorithm 6.3 | DTDD AP Scheduling . . . . .     | 95 |



## LIST OF ACRONYMS

|             |                               |
|-------------|-------------------------------|
| <b>1G</b>   | First Generation              |
| <b>2.5G</b> | Second and a half Generation  |
| <b>2G</b>   | Second Generation             |
| <b>3G</b>   | Third Generation              |
| <b>4G</b>   | Fourth Generation             |
| <b>5G</b>   | Fifth Generation              |
| <b>6G</b>   | Sixth Generation              |
| <b>AI</b>   | Artificial Intelligence       |
| <b>AoA</b>  | Angle of Arrival              |
| <b>AP</b>   | Access Point                  |
| <b>AWGN</b> | Additive White Gaussian Noise |
| <b>BS</b>   | Base Station                  |
| <b>CDMA</b> | Code Division Multiple Access |
| <b>CF</b>   | Cell-Free                     |
| <b>CFAR</b> | Constant False Alarm Rate     |
| <b>CLI</b>  | Cross-Link Interference       |
| <b>CPU</b>  | Central Processing Unit       |
| <b>CRB</b>  | Cramér-Rao Bound              |
| <b>CRLB</b> | Cramér-Rao Lower Bound        |
| <b>CSI</b>  | Channel State Information     |



|                     |                                            |
|---------------------|--------------------------------------------|
| <b>DFT</b>          | Discrete Fourier Transform                 |
| <b>DL</b>           | Downlink                                   |
| <b>DoF</b>          | Degrees of Freedom                         |
| <b>DTDD</b>         | Dynamic Time Division Duplex               |
| <b>E-LMMSE</b>      | Ergodic Linear Minimum Mean Squared Error  |
| <b>FDMA</b>         | Frequency Division Multiple Access         |
| <b>GLRT</b>         | Generalized Likelihood Ratio Test          |
| <b>i.i.d.</b>       | independent and identically distributed    |
| <b>InAI</b>         | Inter-AP Interference                      |
| <b>InUI</b>         | Inter-User Interference                    |
| <b>IoT</b>          | Internet of Things                         |
| <b>IP</b>           | Internet Protocol                          |
| <b>ISAC</b>         | Integrated Sensing and Communication       |
| <b>L-MMSE</b>       | Linear Minimum Mean Squared Error          |
| <b>LP-MMSE</b>      | Local Partial MMSE                         |
| <b>LS</b>           | Least Squares                              |
| <b>LTE</b>          | Long Term Evolution                        |
| <b>LTE-Advanced</b> | Long Term Evolution-Advanced               |
| <b>MIMO</b>         | Multiple Input Multiple Output             |
| <b>MMIMO</b>        | Massive MIMO                               |
| <b>MMSE</b>         | Minimum Mean-Squared Error                 |
| <b>MR</b>           | Maximum Ratio                              |
| <b>MSE</b>          | Mean Squared Error                         |
| <b>MTI</b>          | Moving Target Indicator                    |
| <b>OFDM</b>         | Orthogonal Frequency-Division Multiplexing |

|             |                                         |
|-------------|-----------------------------------------|
| $P_D$       | Probability of Detection                |
| $P_{FA}$    | Probability of False Alarm              |
| <b>QoS</b>  | Quality of Service                      |
| <b>RCS</b>  | Radar Cross Section                     |
| <b>RMS</b>  | Root Mean Square                        |
| <b>ROC</b>  | Receiver Operating Characteristic       |
| <b>RZF</b>  | Regularized Zero-Forcing                |
| <b>SCNR</b> | Signal-to-Clutter-and-Noise Ratio       |
| <b>SE</b>   | Spectral Efficiency                     |
| <b>SINR</b> | Signal-to-Interference-plus-Noise Ratio |
| <b>SMI</b>  | Sensing Mutual Information              |
| <b>SMS</b>  | Short Message Service                   |
| <b>SNR</b>  | Signal-to-Noise Ratio                   |
| <b>SSB</b>  | Synchronization Signal Block            |
| <b>TDD</b>  | Time Division Duplex                    |
| <b>TDMA</b> | Time Division Multiple Access           |
| <b>TRI</b>  | Target-Reflected Interference           |
| <b>UAV</b>  | Unmanned Aerial Vehicle                 |
| <b>UE</b>   | User Equipment                          |
| <b>UL</b>   | Uplink                                  |
| <b>ULA</b>  | Uniform Linear Array                    |
| <b>ZF</b>   | Zero-Forcing                            |





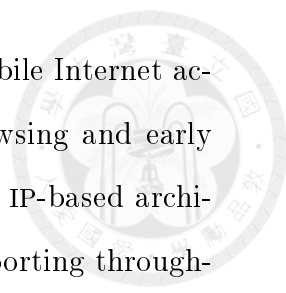


# 1 | INTRODUCTION

Wireless communication technologies have undergone rapid evolution over the past decades, transforming the way we connect, communicate, and interact with the world around us. Mass-adoption of mobile devices, the proliferation of data-intensive applications, and the emergence of new use cases have driven a technological revolution in the telecommunications landscape, leading to the development of countless new technologies and paradigms. These advancements will drive the global mobile data traffic to grow at an unprecedented rate in the coming years, forecasting 9.5 billion mobile subscriptions by 2031 [1].

## 1.1 HISTORICAL EVOLUTION OF MOBILE CONNECTIVITY

From the mid-1980s onward, cellular networks have evolved through successive generations, each introducing significant advancements in terms of capacity, latency, reliability, and user experience [2]. First Generation (1G) systems were built for analog voice, with only minimal data capability and several practical shortcomings, including weak security, low efficiency, and difficult handover due to the lack of a unified standard. Second Generation (2G) marked the shift to digital communications, adopting multiple-access schemes such as Time Division Multiple Access (TDMA) and Code Division Multiple Access (CDMA); this improved voice quality and capacity and introduced services like Short Message Service (SMS) and basic data connectivity on the order of tens of kbit/s, e. g., 64 kbit/s, with Global System for Mobile Communications (GSM) becoming the dominant standard [3].



With Third Generation (3G), cellular systems moved toward mobile Internet access, targeting data rates around 2 Mbit/s and enabling web browsing and early multimedia streaming. Fourth Generation (4G) then introduced an IP-based architecture to increase Spectral Efficiency (SE) and reduce latency, supporting throughput around 1 Gbit/s downlink and 500 Mbit/s uplink and consolidating technologies such as Orthogonal Frequency-Division Multiplexing (OFDM), Multiple Input Multiple Output (MIMO), and coordinated transmission/reception within standards like Long Term Evolution-Advanced (LTE-Advanced) and Worldwide Interoperability for Microwave Access (WiMAX) [4]. Building on this, Fifth Generation (5G) aims not only at even higher peak rates—up to 10 Gbps—but also at improved reliability, lower latency, higher energy efficiency, and massive connectivity. A comprehensive summary of the key characteristics and enabling technologies across all mobile network generations is provided in Table 1.1 [5, 6].

## 1.2 FUTURE CHALLENGES AND OPPORTUNITIES

As we look toward the future, several challenges of current wireless technologies become apparent. The increasing requirements for data rates, latency, and connectivity density strain existing network architectures and spectrum resources. Moreover, emerging applications such as augmented reality, autonomous vehicles, and massive Internet of Things (IoT) deployments demand new capabilities that current systems may struggle to provide [7]. While still in development, Sixth Generation (6G) is envisioned to push wireless communication further, targeting data rates in the order of hundreds of Gbit/s, sub-millisecond latency, and ubiquitous connectivity. Beyond traditional communication metrics, 6G aims to integrate sensing capabilities, AI-native network architectures, and enhanced energy efficiency to support applications such as Extended Reality (XR), holographic communications, and pervasive IoT. This evolution will require novel technologies and paradigms, including terahertz communications, intelligent surfaces, and Integrated Sensing and Communication (ISAC) systems [8].

| GEN. | ERA          | MAIN FEATURE                                                                    | KEY ENABLERS                                                                  |
|------|--------------|---------------------------------------------------------------------------------|-------------------------------------------------------------------------------|
| 1G   | 1980s        | Analog cellular voice; macro-cells; basic mobility and power control.           | Analog Frequency Modulation (FM); FDMA; circuit-switched core.                |
| 2G   | 1990s        | Digital voice + SMS; improved SE and security; low-rate data.                   | TDMA/FDMA and CDMA; digital modulation/coding; circuit-switched.              |
| 2.5G | 1999–2005    | Packet data over 2G infrastructure; best-effort mobile Internet.                | GPRS/EDGE overlay; packet core introduction; improved coding/modulation.      |
| 3G   | 2001–2010    | Mobile Internet; higher throughput; better mobility support.                    | WCDMA / UMTS; HSPA / HSPA+; fast power control; packet-centric evolution.     |
| 4G   | 2010–2019    | Broadband all-IP; low latency; high capacity and robust mobility.               | OFDMA / Single Carrier-FDMA; MIMO; LTE / LTE-Advanced.                        |
| 5G   | 2019–present | eMBB/URLLC/mMTC; flexible deployments; edge native services; slicing.           | 5G New Radio (NR); MMIMO + beamforming; mmWave; network slicing.              |
| 6G   | ~2030        | IMT-2030 vision: AI-native networking; ubiquitous coverage; extreme data rates. | Sub-THz + wideband waveforms; RAN automation; RIS; ISAC; CF; MMIMO evolution. |

Table 1.1: Evolution of mobile networks from 1G to 6G: high-level services, capabilities, and enabling technologies.

### 1.2.1 *Integrated Sensing and Communication*

Sensing capabilities are expected to play a crucial role in 6G networks, enabling new applications and services that leverage the ability to perceive and understand the environment. By integrating sensing functionalities into communication systems, 6G networks can provide enhanced situational awareness, enable new forms of human-machine interaction and support advanced applications such as autonomous driving, smart cities, and the low-altitude economy [9]. In this context, ISAC has emerged as a promising paradigm that combines communication and sensing functionalities within a unified framework, allowing for efficient resource utilization and improved performance for both tasks [10].

### 1.2.2 Cell-Free Massive MIMO as an enabling architecture

Another emerging technology that holds great promise for meeting future connectivity demands is Cell-Free (CF) Massive MIMO (MMIMO). By distributing a large number of Access Points across a wide area and coordinating their transmissions, CF-MMIMO can provide more uniform coverage, high Spectral Efficiency, and improved reliability, making it well-suited for the dense and heterogeneous environments envisioned in 6G networks [11]. The same architectural features that benefit communication, such as distributed infrastructure, macro-diversity, and coordination, can also be leveraged to enable *multi-static* sensing, where multiple spatially separated transmitters/receivers observe the same target from different perspectives, potentially improving detection robustness and localization accuracy within an ISAC framework [10].

## 1.3 DRONE SURVEILLANCE AS A 6G MISSION PROFILE

The rapid growth of Unmanned Aerial Vehicles (UAVs) is enabling new services (e. g., inspection, delivery, emergency response), but also raises safety, security, and airspace management concerns. Drone surveillance is therefore a representative and timely *mission profile* for 6G sensing-enabled networks, especially in the context of the low-altitude economy [9].

From a signal processing and network design perspective, drone surveillance is a challenging task with several key requirements:

- *Volumetric search in 3D*: targets may appear anywhere within an airspace volume, motivating scan-and-refine procedures (coarse-to-fine search).
- *Low observability*: small targets may have limited Radar Cross Section (RCS) and be affected by clutter and interference, stressing detection performance.
- *Multi-target and mobility*: multiple UAVs may be present and moving, requiring target association and tracking over time.

- *Coexistence with communication:* sensing must be performed while guaranteeing communication Quality of Service (QoS) for users, leading to explicit sensing–communication trade-offs.

These requirements naturally motivate an ISAC approach, where the same infrastructure and radio resources are jointly designed to serve communication User Equipment (UE) while providing sensing-based situational awareness [10].

### 1.3.1 *Cell-Free ISAC for Drone Surveillance*

CF-MMIMO offers several benefits for drone surveillance when combined with ISAC [10, 11]. Multi-static diversity is achieved through distributed Access Points (APs) that provide multiple spatial perspectives of the same target, increasing robustness to blockages and unfavorable geometries. The deployment also ensures uniform coverage, reducing cell-edge effects and improving reliability across large areas. Additionally, coordinated beamforming enables cooperation that provides spatial selectivity and interference mitigation.

However, CF-ISAC also introduces non-trivial challenges [10]. Scalability of computational power and fronthaul capacity is a primary concern, as coordination and processing must scale with the number of APs and UEs. Furthermore, coherent multi-static sensing requires stringent synchronization across distributed APs, with operation highly sensitive to timing and frequency misalignments. Finally, dynamic Uplink (UL)/Downlink (DL) scheduling in flexible duplexing modes can introduce Cross-Link Interference (CLI) that demands careful resource coordination and power control.

### 1.3.2 *Application scope and Best-Effort Sensing rationale*

The framework developed in this thesis is designed for a class of low-altitude UAV monitoring scenarios in which a communication network is the primary deployed infrastructure and drone surveillance is a secondary, value-added service. Understanding this scope is essential for the design philosophy developed in the following

chapters, and in particular for the choice of not imposing a minimum sensing guarantee alongside the communication requirements.

The intended deployment context is characterized by three properties that hold simultaneously. First, the communication service is the primary and non-negotiable obligation of the network. Ground users depend on the network for reliable connectivity, and no sensing goal can be allowed to reduce their QoS below a guaranteed floor. Second, the sensing requirement is statistical rather than deterministic: the relevant operational objective is to maintain persistent situational awareness over a series of revisit cycles, not to guarantee detection at every instant. A target that is not detected in one slot can be re-observed in a subsequent one, and the quality of surveillance is governed by the average detection reliability over time. Third, the consequence of a single missed detection is limited: an individual undetected target in one slot does not trigger a safety-critical response, because the system is designed to recover through re-observation in subsequent slots.

These properties are jointly satisfied by several classes of civilian UAV monitoring applications, including the following representative examples.

- *Registered-UAV corridor monitoring.* In emerging UAV traffic management frameworks for the low-altitude economy, commercially operated drones are increasingly required to carry electronic identification and to file flight plans [9]. A CF network serving the same coverage area can function as a passive verification layer, checking compliance with declared trajectories and flagging unexpected entries, without requiring dedicated radar hardware. An occasional missed detection followed by re-detection at the next scheduled revisit carries negligible operational impact in this context.
- *Smart-city and airspace awareness.* In urban or semi-urban environments where drones are used for inspection, delivery, or monitoring tasks, a communication network can provide background drone detection as a supplementary service [9].

- *Event-venue and logistics-hub monitoring.* Outdoor facilities with transient UAV presence may request supplementary detection coverage over their service area. The expected target density is moderate, most platforms are cooperative or registered, and a transient miss recoverable within a few revisit cycles has low operational cost.

The common thread is that all three settings are communication-primary, statistically tolerant of individual missed detections, and non-safety-critical at the single-slot level. For this class of applications, a *best-effort sensing* approach is not only acceptable but the appropriate design choice: it respects the primacy of communication, exploits the available sensing opportunity whenever resources permit, and relies on a scan-and-revisit architecture to achieve persistent surveillance quality.

By contrast, applications that require a guaranteed minimum detection probability in every slot—such as the protection of restricted airspace around airports, the monitoring of critical infrastructure, or active threat-response operations—lie outside the scope of this design. For those applications, a hard sensing floor constraint would need to be imposed, at the cost of either reduced communication performance or dedicated sensing resources; this is discussed as a future direction in Section 10.3.

## 1.4 PROBLEM STATEMENT

This thesis addresses the design of a scalable CF-ISAC framework for multi-static drone surveillance. The key design goal is to jointly optimize sensing and communication performance under practical constraints. For drone surveillance, sensing performance is primarily characterized by detection metrics (e.g., Probability of Detection ( $P_D$ ) at a fixed Probability of False Alarm ( $P_{FA}$ )), while communication performance is quantified by standard rate/QoS metrics (e.g., minimum UE rate or SE constraints). The system must also be *scalable*, meaning that computation and fronthaul/backhaul requirements must remain feasible as the number of UE, UAVs and APs grows [12].

A hybrid signaling strategy is considered, where deterministic probing is embedded into payload-carrying DL streams and dedicated sensing beams are activated only when residual spatial and power resources are available. This flexible signaling approach allows for explicit control of the sensing budget, but also introduces new design challenges in terms of voxel assignment, sequence reuse, resource allocation, and interference management that must be addressed to ensure coexistence with communication [13, 14].

A Dynamic Time Division Duplex (DTDD) scheduling strategy is employed, allowing UL/DL directions to be flexibly reconfigured within each coherence time interval. This approach enables adaptive resource allocation and better sensing-communication coexistence; however, it introduces CLI that requires careful management through beamforming strategies [15].

#### 1.4.1 Notation and Main symbols

To improve readability across the protocol development, this subsection collects the notation conventions and the main symbols used in the thesis. The detailed definition of each quantity is still given when the corresponding model or algorithm is introduced, while Table 1.2 is intended as a global map of the recurring notation families.

Unless otherwise stated, regular font letters,  $x$ , denote scalars, boldface lowercase letters,  $\mathbf{x}$ , denote vectors, bold uppercase letters,  $\mathbf{X}$ , denote matrices, and calligraphic uppercase letters,  $\mathcal{X}$ , denote sets. The operators  $(\cdot)^T$ ,  $(\cdot)^*$ ,  $(\cdot)^H$ , and  $(\cdot)^\dagger$  denote transpose, complex conjugate, Hermitian transpose, and Moore-Penrose pseudoinverse, respectively. Moreover,  $|\mathcal{X}|$  denotes the cardinality of a set,  $\|\cdot\|$  is the Euclidean norm,  $\text{tr}(\cdot)$  is the trace,  $\text{diag}(\cdot)$  and  $\text{blkdiag}(\cdot)$  denote diagonal and block-diagonal constructions,  $\mathbb{E}\{\cdot\}$  is expectation, and  $\mathcal{CN}(\boldsymbol{\mu}, \mathbf{C})$  denotes a circularly symmetric complex Gaussian distribution with mean  $\boldsymbol{\mu}$  and covariance matrix  $\mathbf{C}$ .

| SYMBOL                                                                                         | DEFINITION                                                                                                                                    |
|------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|
| $M, N, K, V$                                                                                   | Number of distributed APs, antennas per AP, active UEs, and candidate/requested voxels.                                                       |
| $t, \tau, T_{\text{slot}}, \tau_p, T$                                                          | Slot index, payload-symbol index, slot/coherence-block length, pilot-phase length, and payload-window length $T = T_{\text{slot}} - \tau_p$ . |
| $\mathcal{A}, \mathcal{A}_d[t], \mathcal{A}_u[t]$                                              | Full AP set, slot-dependent DL/UL AP partitions used by DTDD.                                                                                 |
| $\mathcal{U}_d[t], \mathcal{U}_u[t]$                                                           | Active DL and UL UE sets in slot $t$ .                                                                                                        |
| $\mathcal{C}_k, \mathcal{D}_j[t], \mathcal{U}_m[t]$                                            | User-centric candidate cluster of UE $k$ , capped DL service set of AP $j$ , and locally served UL set of AP $m$ .                            |
| $\mathcal{V}[t], \mathcal{V}_{\text{scan}}[t], \mathcal{V}_{\text{inf}}[t]$                    | Voxel grid, requested scan set, and informative voxel set in slot $t$ .                                                                       |
| $I[v, t], Q_{\text{eff}}[t], Q_{\text{trk}}[t]$                                                | Binary illumination flag, number of informative voxels, and number of track-maintenance scan requests.                                        |
| $\mathcal{V}_m^{\text{sens}}[t], B_u^{\text{max}}, \mathcal{M}_v[t]$                           | Bounded sensing list of UL AP $m$ , its post-processing cap and scheduled receiver set.                                                       |
| $s_v[t], d_{s_v[t]}[\tau], \mathcal{B}_v[t]$                                                   | Deterministic sequence index of voxel $v$ , corresponding waveform, and set of DL APs illuminators for voxel $v$ .                            |
| $\mathbf{h}_{jn}[t], \hat{\mathbf{h}}_{jn}[t]$                                                 | AP-UE channels and their pilot-based estimates.                                                                                               |
| $\mathcal{N}_j^{\text{IA}}, \mathcal{N}_m^{\text{DL}}, L_{\text{IA}}, \mu_{mj}^{\text{store}}$ | Dominant nuisance-neighbor lists, their size, and stored-list weights.                                                                        |
| $P_j^{\text{max}}, P_u^{\text{max}}, N_0$                                                      | Per-AP transmit-power budget, UL user power budget, noise power.                                                                              |
| $\Gamma_n^{\text{min}}, \Gamma_k^{\text{min}}$                                                 | Minimum DL/UL SINR requirements for UE $n$ and $k$ .                                                                                          |

Table 1.2: Main system and control notation used in the proposed DTDD CF-ISAC protocol.

Slot-level quantities are indexed by  $t$ , while payload-symbol quantities inside a slot are indexed by  $\tau$ . Superscripts are used to distinguish processing branches or implementations, for example  $(\cdot)^{\text{DL}}$ ,  $(\cdot)^{\text{UL}}$ ,  $(\cdot)^{\text{hyb}}$ ,  $(\cdot)^{\text{ded}}$ ,  $(\cdot)^{\text{a}}$ , and  $(\cdot)^{\text{c}}$  for Downlink, Uplink, hybrid probing, dedicated sensing, antenna-domain, and centralized quantities.

## 1.5 RELATED WORKS

The proposed framework integrates four partly independent research threads: scalable user-centric CF MMIMO; target detection and waveform design in CF-ISAC sys-

tems; DTDD and Inter-AP Interference (InAI) management; and volumetric sensing for drone surveillance. This section reviews the most directly related prior work in each thread, identifies what has been adapted or taken as a starting point for the present framework, and explains where the thesis departs from the existing literature.

### 1.5.1 *User-centric Cell-Free Massive MIMO*

Cell-Free Massive MIMO was introduced by Ngo et al. [16] as an alternative to conventional cellular Massive MIMO, placing a large number of distributed APs across a wide area and coordinating them to jointly serve all UEs without cell boundaries. The original proposal assumed a fully centralized processing model, which does not scale to large networks since per-AP computation and fronthaul exchange grow with the total number of UEs [12]. Bjornson and Sanguinetti [12] resolved this by introducing an explicit scalability criterion: each AP may only perform operations whose complexity is independent of the total number of UEs; showing that local Local Partial MMSE (LP-MMSE) precoding and local Linear Minimum Mean Squared Error (L-MMSE) combining within user-centric clusters satisfy it.

The communication layer of the present thesis is built directly on Bjornson and Sanguinetti [12] and Demir et al. [17]: user-centric clusters, pilot-based local channel estimation, local LP-MMSE payload precoding, local L-MMSE combining are adopted without modification. The scalability criterion of Bjornson and Sanguinetti [12] is extended in Chapter 5 to cover the joint AP-UE-voxel case, in which per-AP processing must remain bounded not only in the total number of UEs but also in the total number of candidate voxels.

### 1.5.2 *Cell-Free ISAC for Target Detection*

The integration of radar-type sensing into CF MMIMO has lately attracted significant research attention. Behdad et al. [18] study single-target detection in a fully centralized CF MMIMO network, where transmit APs serve UEs in the DL while op-

tionally steering an additional beam toward a single target at a fixed known location. A maximum-a-posteriori-ratio-test detector is derived, with power allocation maximizing the sensing SINR subject to minimum per-UE SINR constraints. They compare two ISAC configurations: communication beams only, and an added dedicated sensing beam with a fixed power budget. The paper demonstrates that the dedicated beam improves detection performance for targets with low RCS at the cost of communication resources.

The two-configuration comparison, payload-only sensing versus payload-plus-dedicated-beam, directly motivates the layered design philosophy of this thesis. However, the DL waveform in the present thesis uses three layers rather than two, exploiting also sensing signals superimposed to the UE streams, and the target is not at a fixed known location but must be discovered and tracked. Furthermore, this thesis considers a scalable user-centric CF architecture, whereas Behdad et al. [18] assume fully centralized processing.

Demirhan and Alkhateeb [19] propose a joint sensing and communication beam-forming design for a centralized CF-ISAC system with multiple targets at fixed known locations. They formulate a max-min fair communication SINR optimization subject to per-target sensing gain constraints, showing that joint design achieves near-communication-optimal SINR while approaching sensing-optimal gain. This work provides further evidence that dedicated sensing beams in a CF architecture are a practically valuable design element. However, this paper operates in a fully centralized Time Division Duplex (TDD) framework, with no scan management.

Elfiatoure et al. [20] study multiple-target detection in a CF MMIMO ISAC system, where APs steer probing signals toward predefined static sensing zones and AP modes are selected jointly with power control to maximize detection performance. This is the closest CF-ISAC prior work to the multi-target setting of this thesis. The work confirms that assigning different APs to different sensing regions while still serving communication UEs is feasible, and motivates the multi-voxel scan design. Nevertheless, the sensing zones are fixed spatial regions defined prior to operation,

with no per-slot adaptation. The thesis extends the multi-zone concept to a dynamic, slot-varying voxelized search problem with explicit scan management.

Finally, Galappaththige et al. [10] provide a comprehensive survey of the CF-ISAC field, explicitly identifying multi-target detection, fronthaul scalability, and interference management as open research directions. This survey is used throughout the thesis as a broad reference for contextualizing the proposed framework in the literature.

### 1.5.3 DTDD and InAI Management in CF Systems

Operating CF MMIMO with DTDD, where some APs transmit DL signals while others receive UL signals in the same time-frequency resource, was studied from a communication perspective by Mohammadi et al. [21]. They formulated joint AP mode assignment and power control as a mixed-integer problem maximizing sum SE subject to per-UE rate requirements, showing that a virtually full-duplex architecture built from half-duplex hardware roughly doubles the 95%-likely sum SE of conventional half-duplex TDD. This work established the key structural insight that CLI, InAI from DL APs to UL APs, and Inter-User Interference (InUI) from UL users to DL users, can be managed at the network level through mode scheduling. The virtually full-duplex CF architecture of Mohammadi et al. [21] is the communication-layer heritage for the DTDD mode used in this thesis.

Chowdhury et al. [15] present the first analysis of DTDD in a CF-ISAC setting and constitute the single closest prior work to this thesis. They propose a joint sensing and bidirectional communication framework where DL APs transmit payload beams toward served UEs alongside a dedicated sensing beam aimed at a single fixed target, while UL APs simultaneously receive UE payloads and collect sensing echoes. Both a centralized and a distributed Generalized Likelihood Ratio Test (GLRT) are derived for single-target detection, treating UL UE signals as part of the equivalent sensing noise. The sum UL-DL SE is analyzed including the effects of InAI and InUI, and the paper demonstrates that DTDD can approximately double the 90%-likely

sum SE compared with conventional TDD [15]. For DL precoding, a Regularized Zero-Forcing (RZF) design is used for the communication streams alongside target-centric and user-centric sensing precoders. For UL combining, a fixed SINR-optimal Minimum Mean-Squared Error (MMSE) combiner is derived whose interference covariance absorbs the residual InAI as a static term. AP mode assignment follows local traffic demand and is corrected to guarantee a nearby UL AP and a nearby DL AP for the single known fixed target location.

The DTDD CF-ISAC system model and the side-by-side centralized/distributed GLRT comparison structure are all adopted from Chowdhury et al. [15] as the substrate infrastructure for this thesis. However, the proposed framework extends and departs from Chowdhury et al. [15] in the four primary novelties described in Section 1.6: a full scan-and-track protocol layer, a three-layer hybrid DL waveform, an InAI-protected DL precoder with adaptive directional penalty toward UL AP locations, and an InAI-protected UL combiner with adaptive activation scalar. Furthermore, the present framework enforces the user-centric scalability criterion of Bjornson and Sanguinetti [12], which Chowdhury et al. [15] do not address.

#### 1.5.4 *Voxelized sensing and Drone Surveillance in ISAC*

Coordinated beam sweeping over a voxelized three-dimensional surveillance volume for low-altitude drone detection was proposed by Jopanya et al. [22] in a multi-static MIMO-ISAC network. Multiple APs simultaneously steer a sensing beam toward each voxel of the surveillance grid during the beam-sweeping phase of the 5G Synchronization Signal Block (SSB), with a precoder designed to enforce orthogonality between the sensing beam and the SSB so as to protect UE signal quality. The voxelized surveillance region model is the spatial abstraction of this thesis, directly inspired by Jopanya et al. [22] and grounded in classical radar scan-volume partitioning [23, 24]. While the scheme in Jopanya et al. [22] performs a fixed sequential sweep with no per-slot adaptation, the thesis embeds the voxelized region into a full adaptive scan-and-track protocol layer integrated with a DTDD CF-ISAC stack.

The scan-and-track concepts used in the scheduling layer, such as the coarse-to-fine grid, revisit-time budgeting, track initiation and confirmation, have been already introduced in classical radar systems [23, 24]. They show that jointly processing both components reduces estimation error significantly compared with a pilot-only design, and provide a performance analysis under imperfect Channel State Information (CSI).

### 1.5.5 *Hybrid waveforms and Deterministic sensing signals in ISAC*

Several recent works have studied the use of communication waveforms for sensing and the design of deterministic signal components superimposed on communication streams.

Xu et al. [13] analyze sensing performance in a MIMO-ISAC system that exploits both the deterministic pilot component and the random data payload of the communication frame. They show that jointly processing both components significantly reduces estimation error relative to a pilot-only design, under imperfect CSI. Xie et al. [14] study the same hybrid signal from an information-theoretic perspective, deriving a closed-form expression for the sensing mutual information of a waveform comprising deterministic pilots and random data payloads. Both works motivate the payload-attached probing branch of the hybrid DL waveform in this thesis, in which a deterministic Discrete Fourier Transform (DFT)-based sequence is superimposed onto an active payload beam.

### 1.5.6 *Summary*

The analysis above shows that the key building blocks of the proposed framework exist separately in the literature. While most works address one aspect of the problem, none combine all of these elements. Table 1.3 provides a compact feature comparison with the five most closely related works.

| FEATURE                    | [18] | [15]    | [20]    | [22] | [13] | THIS WORK |
|----------------------------|------|---------|---------|------|------|-----------|
| CF MMIMO architecture      | Yes  | Yes     | Yes     | –    | –    | Yes       |
| DTDD multiplexing          | –    | Yes     | –       | –    | –    | Yes       |
| UL communication users     | –    | Yes     | –       | –    | –    | Yes       |
| Voxelized 3D region        | –    | –       | Partial | Yes  | –    | Yes       |
| Adaptive scan-and-track    | –    | –       | –       | –    | –    | Yes       |
| Superimposed probing       | –    | –       | –       | –    | Yes  | Yes       |
| Dedicated sensing beam     | Yes  | Yes     | –       | Yes  | –    | Yes       |
| InAI-protected DL precoder | –    | –       | –       | –    | –    | Yes       |
| InAI-protected UL combiner | –    | Partial | –       | –    | –    | Yes       |
| Scalable user-centric      | –    | –       | –       | –    | –    | Yes       |

Table 1.3: Feature comparison between the proposed framework and the five most closely related prior works. “Partial” indicates a related but incomplete treatment; “–” indicates the feature is absent.

## 1.6 THESIS CONTRIBUTIONS

The prior work reviewed in Section 1.5 establishes the building blocks on which this thesis is constructed, but leaves the central problem of designing a scalable, protocol-level DTDD CF-ISAC architecture for multi-target drone surveillance open. The contributions are organized below into three tiers: main novelties, extensions of existing work, and the application of established methodology.

### 1.6.1 Primary novelties

- *Voxelized scan-and-track protocol layer.* The thesis introduces the scan-and-track protocol layer for CF-ISAC. The monitored airspace is partitioned into a Cartesian voxel grid: a coarse grid for discovery and a local fine grid for post-detection localization. The Central Processing Unit (CPU) maintains explicit scan pointers that advance only when a requested voxel is confirmed as informative, i.e., when it receives at least the minimum DL illumination and UL observation support. Detections trigger track initiation, and active tracks are maintained across slots through cyclic scan lists that keep the network revisiting the relevant airspace region. No prior CF-ISAC work implements such a layer: Behdad et al. [18] and Chowdhury et al. [15] both assume a single tar-

get at a fixed known location, while Elfiatoure et al. [20] probe a set of static predefined zones without any per-slot adaptation. The voxelized region model is inspired by Jopanya et al. [22] and the classical radar scan-volume literature [23, 24], but the integration of adaptive scan planning, pointer advancement, bridge tracking, and multi-track maintenance into a DTDD CF-ISAC scheduling and waveform stack is entirely original.

- *Three-layer hybrid downlink waveform.* The DL waveform combines three sensing layers in every slot. The *payload layer* carries user data on locally constructed protected precoder directions and is always active; it constitutes the communication baseline and defines the spatial degrees of freedom already committed to users. The *payload-attached probing layer* superimposes a per-voxel deterministic DFT-based sequence onto a payload beam when that beam's spatial direction provides sufficient sensing coverage of a requested voxel, reusing an active communication stream for target illumination without any additional beamformer. The *dedicated sensing layer* generates sensing beams in the complement of the served-UE subspace, using the residual spatial degrees of freedom and the remaining per-AP power budget, targeting voxels that could not be matched to any payload stream. The concept of superimposing a deterministic component on a communication payload is motivated by single-node ISAC studies [13, 14], which do not address the three-layer construction or the CF setting. The simultaneous use of all three layers in a DTDD CF-ISAC slot, together with the communication-constrained admission logic that determines which layer each voxel candidate enters, is a contribution of this thesis.
- *InAI-protected downlink LP-MMSE precoder.* The standard local LP-MMSE precoder [12, 17] is augmented with a direction-specific InAI penalty. Each DL AP stores a bounded list of the most vulnerable neighboring UL APs identified from path-loss and residual coupling statistics. The penalty is a weighted sum of rank-one matrices pointing toward the physical locations of those UL

neighbors. Its scalar strength is adaptive: it grows with the active InAI threat and falls back to zero when there is no active threat, preserving the LP-MMSE gain. The LP-MMSE structure itself is standard [12, 17]; the direction-specific penalty toward UL AP physical locations and the slot-adaptive scalar strength are new. Chowdhury et al. [15] use an RZF precoder for the communication streams whose regularizer is a scalar multiple of the identity and does not account for the directions of neighboring UL APs; their approach to managing InAI relies exclusively on scheduling and on the receiver-side covariance model.

- *InAI-protected uplink L-MMSE combiner.* The local L-MMSE combiner is augmented with a complementary direction-specific penalty pointing toward the physical locations of the most dangerous active DL neighbor APs. An adaptive scalar activates this penalty only when the actual residual InAI-to-noise threat at the UL AP exceeds a reference level, avoiding unnecessary modification of the useful-signal suppression when the leakage is negligible. This is the receiver-side counterpart of the protected precoder, ensuring that both sides of the DTDD link have threat-adapted InAI mitigation. Chowdhury et al. [15] derive a fixed SINR-optimal MMSE combiner whose interference covariance absorbs the InAI as a static term; the adaptive activation scalar and the direction-specific penalty shape are new.

### 1.6.2 Adaptations and Extensions of Prior Work

- *DTDD AP mode assignment extended to a voxelized sensing task.* The scheduling layer follows the traffic-based AP mode assignment logic of Chowdhury et al. [15], in which each AP is assigned the direction corresponding to its higher local traffic demand. This is augmented with a per-voxel feasibility correction: for every requested voxel, the scheduler verifies that at least one eligible UL AP and at least one eligible DL AP lie within the voxel's sensing neighborhood, and flips the minimum number of APs necessary to restore feasibility.

- *Scalability criterion extended to the ISAC case.* The scalability criterion of Bjornson and Sanguinetti [12] is extended to the joint communication and sensing task. In the proposed framework, per-AP operations must remain bounded in both the number of served UEs and the number of candidate voxels. The formal criterion is stated in Chapter 5 and is verified for every algorithmic component of the protocol.
- *Dual centralized/distributed GLRT applied per-voxel.* The side-by-side formulation of a distributed scalar-GLRT and a centralized raw-uplink GLRT follows the same comparative philosophy introduced by Chowdhury et al. [15]. The thesis adapts this structure to the per-voxel coherent detection problem, in which the sensing signature is a known DFT-based sequence, and applies Constant False Alarm Rate (CFAR) thresholding via the central chi-square null distribution of the test statistic.

### 1.6.3 Standard methodology

The following components are based on well-established signal-processing and CF MMIMO techniques and are not novel contributions to this work. The GLRT detector and CFAR thresholding follow classical statistical signal processing, with the null and alternative distributions of the test statistic derived from central and noncentral chi-square laws [23, 25, 26]. User-centric cluster construction, pilot assignment, and local channel estimation follow Bjornson and Sanguinetti [12] and Demir et al. [17]. The coherent accumulation of echo energy from multiple AP illuminators through a shared per-voxel sequence is a standard principle in distributed coherent detection [23]. The matched-filter UL sensing front-end and the DFT-based waveform generation are standard in radar signal processing [23, 24]. The coarse-to-fine grid concept, while used in classical radar scan-and-track [23, 24], requires original adaptation in this thesis to function within the DTDD AP scheduling and communication-feasibility constraints.

## 1.7 THESIS ORGANIZATION

The thesis is organized as follows. Chapter 2 reviews the communication foundations used in the thesis, from co-located Massive MIMO to scalable user-centric CF MMIMO. Chapter 3 introduces the sensing background required for drone surveillance, including detection, tracking, RCS, clutter, Signal-to-Clutter-and-Noise Ratio (SCNR), CFAR, and voxel-wise statistical reasoning. Chapter 4 connects communication and sensing through ISAC and CF-ISAC, with emphasis on multi-static sensing, waveform sharing, duplexing, and DTDD operation.

Chapter 5 formulates the considered DTDD CF-ISAC drone-surveillance problem, defining the voxelized sensing task, communication constraints, scalability criterion, and protocol decomposition. Chapter 6 presents the scheduling and control layer, including adaptive grid management, track and bridge handling, voxel-request selection, AP-mode assignment, sequence reuse, and candidate voxel distribution. Chapter 7 develops the hybrid DL design, covering protected payload precoding, payload-to-voxel matching, superimposed probing, dedicated sensing beams, and the final DL signal model. Chapter 8 describes UL reception and sensing detection, including adaptive InAI-protected local UL combining under residual DTDD interference, deterministic signature extraction, distributed coherent GLRT, CPU-side scalar fusion, and a centralized raw-uplink receiver.

Chapter 9 reports the numerical results, comparing the protocol variants in terms of SE,  $P_D$ ,  $P_{FA}$ , tracking performance, InAI robustness, target strength, load, motion, and antenna-budget distribution. Chapter 10 concludes the thesis and outlines future research directions. Finally, Appendix A reports the complete Receiver Operating Characteristic (ROC) curves for the protocol variants and Appendix B provides validation of the adaptive grid management.





## Part I

### THEORETICAL FOUNDATIONS OF CELL-FREE ISAC

In the first part of the thesis, we review the theoretical foundations required to position the proposed Cell-Free ISAC framework for drone surveillance. This part covers the fundamentals of Massive MIMO, CF architectures and scalability, sensing principles, and ISAC frame structures. It also surveys the state of the art in CF ISAC, with particular attention to multi-static sensing, distributed processing and duplexing strategies.





## 2 | MASSIVE MIMO AND CELL-FREE FUNDAMENTALS FOR ISAC

Massive MIMO has emerged as one of the key enabling technologies for modern wireless communication systems, allowing a Base Station (BS) equipped with dozens or even hundreds of antennas to simultaneously serve many UEs over the same time-frequency resources. By exploiting spatial multiplexing and adaptive beamforming, Massive MIMO can substantially improve SE, link reliability, and interference management [27, 28, 29]. At the same time, the evolution from conventional co-located arrays to geographically distributed deployments has led to the Cell-Free Massive MIMO paradigm, in which many distributed APs, coordinated by one or multiple CPUs, jointly serve the users in a user-centric manner. This distributed architecture preserves the main signal processing principles of Massive MIMO, while introducing new opportunities in terms of macro-diversity, coverage uniformity, and scalability, together with new challenges in coordination, synchronization, and fronthaul design.

This chapter provides the theoretical foundation needed for the rest of the thesis by covering both the classical Massive MIMO framework and its Cell-Free extension. We begin with the co-located Massive MIMO setting, which offers the natural starting point to introduce the basic UL/DL system model, TDD operation, and pilot-based channel acquisition. On this basis, we discuss channel estimation and pilot contamination, highlighting their impact on system performance and on the partitioning of the coherence interval between training and data transmission. We then review the main linear multi-user processing techniques and introduce the key performance metrics adopted throughout the thesis, such as achievable SE, SINR, and relevant QoS considerations. Finally, the discussion is extended to Cell-Free Massive MIMO,

showing how the same core principles evolve when the antenna array is replaced by a distributed network of cooperative APs. In this way, the chapter moves from the canonical Massive MIMO model to the distributed architecture that is most relevant for the ISAC framework studied in this thesis.

## 2.1 SYSTEM MODEL

In the classical co-located scenario, a single BS equipped with  $N$  antennas serves  $K$  single-antenna UEs in the same time-frequency resource. For analytical tractability, assume a block-fading channel model: the wireless channels remain approximately constant over a *coherence interval* of  $\tau_c$  channel uses, determined by the channel coherence time  $T_c$  and bandwidth  $B_c$ , and then change independently to new random values. Table 2.1 reports illustrative mobility examples to show the typical order of magnitude of  $\tau_c$ .

Throughout this interval, the channel from UE  $i$  to the  $n$ -th antenna can be modeled as a complex Gaussian random variable. In matrix-vector form, the channel of UE  $i$  to all  $N$  antennas is represented by an  $N \times 1$  vector  $\mathbf{h}_i$ . We adopt the common Rayleigh fading model:

$$\mathbf{h}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_i), \quad (2.1)$$

where  $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  denotes the complex Gaussian distribution with mean  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$ , and  $\mathbf{R}_i$  is the  $N \times N$  spatial correlation matrix capturing the channel's large-scale fading (path loss, shadowing) and correlation between antenna elements. In many analytical examples, one assumes uncorrelated channels with  $\mathbf{R}_i = \beta_i \mathbf{I}_N$ , where  $\beta_i$  is the average channel gain of UE  $i$ .

Figure 2.1 provides an overview of the Massive MIMO processing chain used throughout this chapter, from pilot transmission and CSI estimation to UL/DL processing and final performance metrics over one coherence block.

| SCENARIO         | $f_c$ [GHz] | $v$ [km/h] | $T_c$ [ms] | $\sigma_\tau$ [ $\mu$ s] | $B_c$ [kHz] | $\tau_c \approx T_c B_c$ |
|------------------|-------------|------------|------------|--------------------------|-------------|--------------------------|
| Walking          | 3.5         | 3          | 43         | 1.0                      | 200         | 8600                     |
| Cycling          | 3.5         | 15         | 8.7        | 1.0                      | 200         | 1740                     |
| Car              | 3.5         | 50         | 2.6        | 1.0                      | 200         | 520                      |
| Highway car      | 3.5         | 100        | 1.3        | 1.0                      | 200         | 260                      |
| High-speed train | 3.5         | 250        | 0.52       | 1.0                      | 200         | 104                      |

Table 2.1: Illustrative coherence-parameter examples. Here,  $f_c$  is the carrier frequency,  $v$  is the UE speed,  $T_c$  is the coherence time,  $\sigma_\tau$  is the RMS delay spread,  $B_c$  is the coherence bandwidth, and  $\tau_c$  is the coherence block size. Illustrative formulas are  $f_D \approx \frac{vf_c}{c}$ ,  $T_c \approx \frac{0.423}{f_D}$  for Clarke/Jakes 50% temporal correlation and  $B_c \approx \frac{1}{5\sigma_\tau}$  for about 50% frequency correlation [30].

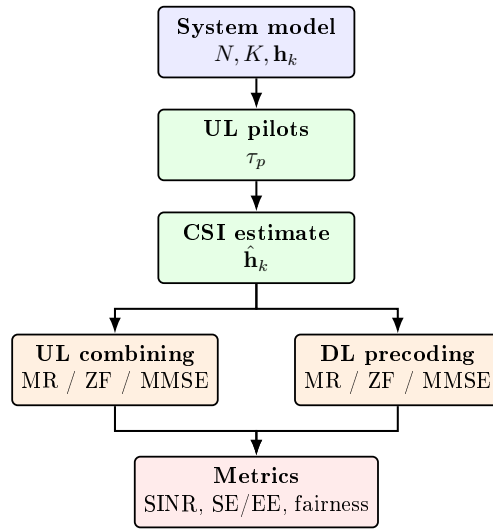


Figure 2.1: Massive MIMO in one picture: compact pipeline from pilots and CSI estimation to UL/DL processing and final metrics over one coherence block.

### 2.1.1 Channel Hardening and Favorable Propagation

Under independent Rayleigh fading, as  $N$  grows large, two useful phenomena occur: *Channel Hardening* and *Favorable Propagation* [17]. Channel hardening refers to the effect that the effective channel gain  $\frac{|\mathbf{h}_i|^2}{N}$  converges to a deterministic value (specifically,  $\beta_i$  in the uncorrelated case) as  $N \rightarrow \infty$ , meaning that the channel behaves more like a non-fading channel with reduced variability. This simplifies channel estimation and improves the effectiveness of linear processing, enabling efficient beamforming [10].

Favorable propagation refers to the phenomenon that the channel vectors of different UEs become nearly orthogonal as  $N$  increases, which allows the base station

to spatially separate UEs with simple linear processing, reducing InUI and achieving high SE [31].

These properties underpin many benefits of Massive MIMO, such as simplified power control and interference averaging. Figure 2.2 illustrates these phenomena, showing how the variance of the normalized channel gain decreases and the mean normalized inner product between two UEs's channels approaches zero as  $N$  increases.

### 2.1.2 TDD Operation and Pilot Transmission

Massive MIMO typically employs TDD operation to exploit *reciprocity* between UL and DL channels [27]. In TDD, the UL and DL use the same frequency band in different time slots, and the physical channel response is ideally identical in both directions, allowing channel estimation to be done in the UL and reused for DL beamforming. Each coherence interval of  $\tau_c$  symbols is divided into three phases: an UL *pilot training* phase of  $\tau_p$  symbols, an UL data transmission phase of  $\tau_u$  symbols, and a DL data transmission phase of  $\tau_d$  symbols [12], as shown in Figure 2.3.

The relative lengths of these phases depend on the traffic and duplexing strategy; for example, in a purely UL scenario  $\tau_d$  might be zero, or vice versa. During the pilot phase, each active UE transmits a known pilot signal that is used by the BSs to estimate the channel. To allow the receiver to distinguish different UEs' channels, the pilot signals of different UEs should be orthogonal. In practice, we have a set of  $\tau_p$  mutually orthogonal pilot sequences of length  $\tau_p$ . When  $K > \tau_p$ , multiple UEs are forced to share the same pilot sequence. This situation causes *pilot contamination*, a phenomenon explored in detail in Section 2.2.

At the receiver side, the BS correlates the received signal with the known pilot sequences to obtain estimates of the channel. To describe this process, let  $\mathcal{S}_t \subset \{1, \dots, K\}$  denote the subset of UEs assigned to pilot sequence  $t \in \{1, \dots, \tau_p\}$ .

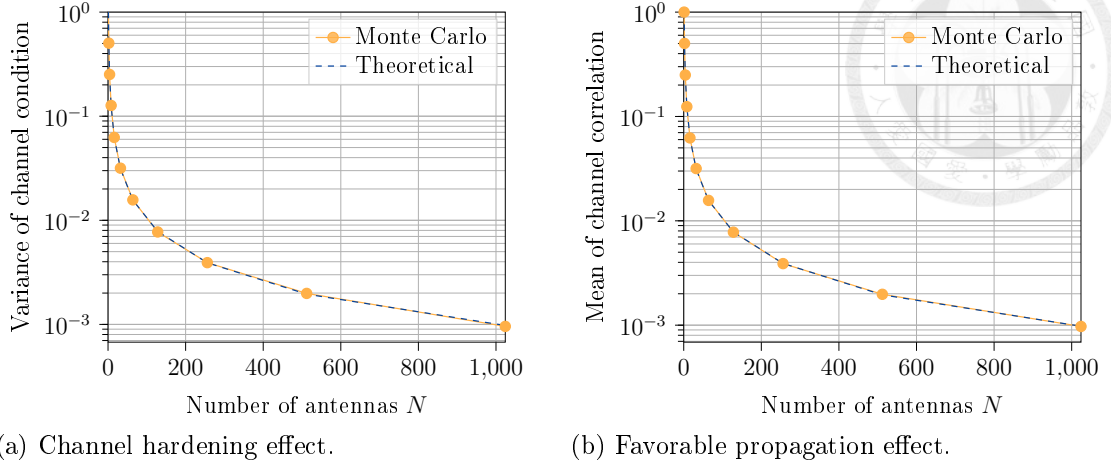


Figure 2.2: Illustration of channel hardening and favorable propagation phenomena in Massive MIMO as the number of antennas  $N$  increases.

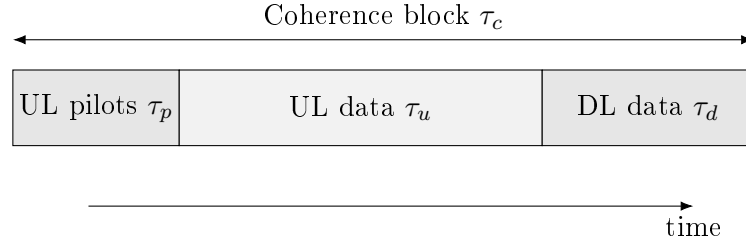


Figure 2.3: Typical TDD frame structure within a coherence interval, showing pilot training and UL/DL data phases.

Ideally,  $|\mathcal{S}_t| = 1 \forall t$ , but in scenarios with more UEs than pilots,  $\exists t : |\mathcal{S}_t| > 1$ . The received pilot signal at the array, after correlating with pilot  $t$ , can be expressed as:

$$\mathbf{y}_t = \sum_{i \in \mathcal{S}_t} \sqrt{\tau_p p_i} \mathbf{h}_i + \mathbf{n}_t, \quad (2.2)$$

where  $p_i$  is the transmit power used by UE  $i$ ,  $\mathbf{h}_i$  is the  $N \times 1$  channel of UE  $i$ , and  $\mathbf{n}_t \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_N)$  is additive noise at the receiver (with  $\sigma^2$  the noise variance per antenna). The factor  $\sqrt{\tau_p}$  appears because the receiver correlates across  $\tau_p$  symbols, effectively giving a processing gain that amplifies the signal relative to noise. From (2.2), we see that if multiple UEs share the pilot,  $\mathbf{y}_t$  contains a superposition of those UEs' channels, making it challenging to separate them.

### 2.1.3 Channel Estimation

Given the observation  $\mathbf{y}_t$  in (2.2), the receiver computes an estimate  $\hat{\mathbf{h}}_k$  of the channel for each UE  $k$ . A simple Least Squares (LS) estimator ignores prior statistics and treats  $\mathbf{y}_t$  as the noisy composite channel: for a UE  $k \in \mathcal{S}_t$ , an LS estimate is  $\hat{\mathbf{h}}_k^{\text{LS}} = \frac{1}{\sqrt{\tau_p p_k}} \mathbf{y}_t$ . This yields

$$\hat{\mathbf{h}}_k^{\text{LS}} = \mathbf{h}_k + \sum_{i \in \mathcal{S}_t, i \neq k} \sqrt{\frac{p_i}{p_k}} \mathbf{h}_i + \frac{1}{\sqrt{\tau_p p_k}} \mathbf{n}_t. \quad (2.3)$$

The interference term (the sum over  $i \neq k$ ) represents contamination of the  $k$ -th channel estimate by other UEs sharing the pilot. While the LS estimate is unbiased in expectation ( $\mathbb{E} [\hat{\mathbf{h}}_k^{\text{LS}}] = \mathbf{h}_k$ ), it has a large variance due to this interference [32]. A better approach is to use the MMSE estimator, which leverages the known second-order statistics  $\mathbf{R}_i$ . The MMSE estimate of  $\mathbf{h}_k$  for  $k \in \mathcal{S}_t$  is given by:

$$\hat{\mathbf{h}}_k = \sqrt{\tau_p p_k} \mathbf{R}_k \left( \sum_{i \in \mathcal{S}_t} \tau_p p_i \mathbf{R}_i + \sigma^2 \mathbf{I}_N \right)^{-1} \mathbf{y}_t, \quad (2.4)$$

which can be derived from the Bayesian MMSE estimation principle [12, 32]. In the special case of uncorrelated channels ( $\mathbf{R}_i = \beta_i \mathbf{I}$ ), the expression reduces to a scalar weighing of the LS estimate:

$$\hat{\mathbf{h}}_k = \frac{\sqrt{\tau_p p_k} \beta_k}{\sum_{i \in \mathcal{S}_t} \tau_p p_i \beta_i + \sigma^2} \mathbf{y}_t. \quad (2.5)$$

Intuitively, the MMSE estimator deweights  $\mathbf{y}_t$  by the interference-plus-noise covariance, thus suppressing the contributions of other UEs' channels according to their large-scale strengths. The quality of the channel estimate depends on the transmit power  $p_k$ , the level of pilot reuse interference and the pilot length  $\tau_p$ .

## 2.2 PILOT CONTAMINATION

Accurate CSI at the transmitter and receiver is critical for Massive MIMO to deliver its promised beamforming gains. In TDD Massive MIMO, the principal bottleneck in acquiring CSI is *pilot contamination*. Pilot contamination leads to InUI that cannot be eliminated, because the interference is induced by channel estimation errors that

align with the desired signal. The following details the impact of pilot contamination and strategies to mitigate it.

When two or more UEs share the same pilot sequence, their channel estimates become linearly dependent. Referring to the MMSE estimate (2.4), if  $i \neq k$  but  $i \in \mathcal{S}_t$ ,  $\hat{\mathbf{h}}_k$  will be an approximate linear combination of  $\mathbf{h}_k$  and the contaminating UEs' channels  $\mathbf{h}_i$ . This has two main consequences for the system [27, 28, 29]:

1. The variance of the channel estimation error increases because the presence of other UEs' pilot signals effectively raises the interference-plus-noise floor during estimation.

2. The channel estimates of UEs sharing a pilot become correlated with each other. This means that when the transmitter uses  $\hat{\mathbf{h}}_k$  to beamform to UE  $k$ , it will inadvertently direct some energy toward UE  $i$ . Such interference scales with the array size in a similar way to the desired signal. This fundamentally limits the asymptotic SINR and achievable rate for UEs experiencing pilot contamination.

Figure 2.4 illustrates these effects empirically, showing how both the desired signal and coherent interference scale with  $N$ , causing the SINR to saturate. The severity of pilot contamination depends on how many UEs share a pilot and their channel disparities. In cellular networks, pilot contamination often refers to interference from UEs in neighboring cells using the same pilot sequences. In practice, systems often have  $K \gg \tau_c$ , especially in massive IoT scenarios. Increasing  $\tau_p$  improves CSI accuracy but reduces the fraction of time available for data, which in turn lowers SE. This trade-off is often referred to as *coherence block budgeting*: system designers must allocate the coherence interval between training and data optimally [33].

Advanced techniques to mitigate pilot contamination include intelligent pilot assignment (e.g., assigning the same pilot only to UEs that are geographically well-separated), large-scale fading precoding across cells [34], and blind or data-aided channel estimation methods that reduce reliance on long orthogonal pilot sequences [35, 36].

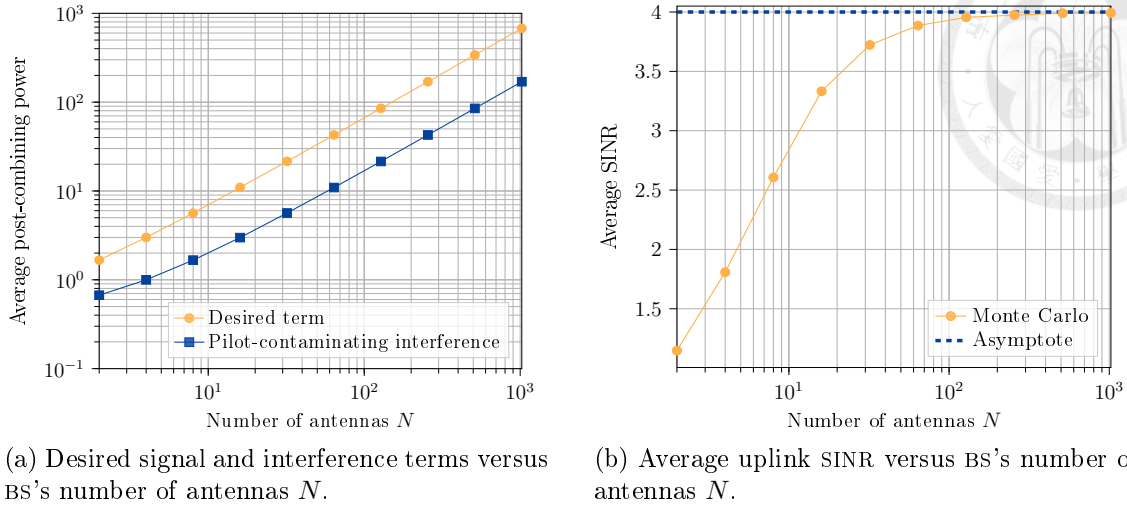


Figure 2.4: Illustration of pilot contamination effects on uplink SINR with MR combining. Two single-antenna users experience i.i.d. Rayleigh fading and reuse the same pilot. The L-MMSE estimate  $\hat{h}_1$  is contaminated by UE2, and the Maximum Ratio (MR) combiner is  $w = \hat{h}_1 / \|\hat{h}_1\|$ . Left: both desired term  $\mathbb{E}(p_u |w^H \hat{h}_1|^2)$  and interference term  $\mathbb{E}(p_u |w^H \hat{h}_2|^2)$  scale approximately linearly with  $N$ , so increasing  $N$  alone does not eliminate the interference. Right: the average uplink SINR saturates as  $N$  increases due to coherent pilot-contaminating interference.

## 2.3 LINEAR PROCESSING FOR UPLINK AND DOWNLINK

Once the channel estimates are obtained, the BS can perform multi-user transmission and reception. In Massive MIMO, a popular and practical approach is to use *linear processing*, whereby each UE's data is encoded/decoded using linear filters based on the CSI. Linear processing strikes a favorable balance between performance and complexity, and in the large- $N$  regime it can approach the optimal performance thanks to channel hardening and favorable propagation [16, 32]. The following describes the three canonical linear processing schemes: Maximum Ratio (MR), Zero-Forcing (ZF), and MMSE processing from a DL precoding perspective.

Let  $s_k$  denote the information symbol for UE  $k$ , with unit power  $\mathbb{E}[|s_k|^2] = 1$ , and mutually uncorrelated. The BS applies a complex *precoding vector*  $\mathbf{w}_k \in \mathbb{C}^N$  to the symbol  $s_k$  and transmits the superposition

$$\mathbf{x} = \sum_{k=1}^K \sqrt{\rho_k} \mathbf{w}_k s_k, \quad (2.6)$$

where  $\rho_k$  is the transmit power allocated to UE  $k$ . Typically, the precoding vectors are normalized such that their power is unity:  $|\mathbf{w}_k|^2 = 1 \forall k$ , and  $\rho_k$  then represents the actual transmit power for UE  $k$ . The received signal at UE  $j$  is

$$y_j = \mathbf{h}_j^H \mathbf{x} + n_j = \sqrt{\rho_j} \mathbf{h}_j^H \mathbf{w}_j s_j + \sum_{k \neq j} \sqrt{\rho_k} \mathbf{h}_j^H \mathbf{w}_k s_k + n_j, \quad (2.7)$$

where  $n_j \sim \mathcal{CN}(0, \sigma^2)$  is the noise at UE  $j$ . The term  $\mathbf{h}_j^H \mathbf{w}_j$  is the complex inner product between the channel of UE  $j$  and its precoding vector, representing the desired beamforming gain for UE  $j$ . The terms for  $k \neq j$  are interference: UE  $j$  hears the components of other UEs' signals weighted by  $\mathbf{h}_j^H \mathbf{w}_k$ . A good precoder will make  $\mathbf{h}_j^H \mathbf{w}_j$  large to maximize signal power for  $j$ , while making  $\mathbf{h}_j^H \mathbf{w}_k$  small for  $k \neq j$  to reduce interference to  $j$ . The three linear precoding strategies are presented in the following.

### 2.3.1 MR Precoding

Maximum Ratio precoding simply aligns each precoding vector with the estimated channel of the target UE, i.e.,  $\mathbf{w}_k^{\text{MR}} = \hat{\mathbf{h}}_k$ . MR is asymptotically optimal in noise-limited scenarios and has low complexity. However, it does nothing to suppress multi-user interference; when multiple UEs' channels are not orthogonal, MR can lead to significant interference. In a network with pilot contamination, MR will even amplify the coherent interference from UEs sharing pilots, since it treats the contaminated channel estimate as truth. Despite being suboptimal, MR processing is widely studied because of its simplicity and it provides a useful performance baseline [16].

### 2.3.2 ZF Precoding

Zero-Forcing precoding is designed to *eliminate* multi-user interference under perfect CSI conditions. The ZF precoder chooses  $\{\mathbf{w}_k\}$  such that  $\mathbf{h}_j^H \mathbf{w}_k = 0$  for all  $j \neq k$ .

In matrix form, if we assemble the estimated channel vectors into an  $N \times K$  matrix  $\hat{\mathbf{H}} = [\hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2, \dots, \hat{\mathbf{h}}_K]$ , then one ZF solution is

$$\mathbf{W}^{\text{ZF}} = \hat{\mathbf{H}}(\hat{\mathbf{H}}^H \hat{\mathbf{H}})^{-1}, \quad (2.8)$$

where  $\mathbf{W}^{\text{ZF}} = [\mathbf{w}_1^{\text{ZF}}, \dots, \mathbf{w}_K^{\text{ZF}}]$  and the matrix inverse exists as long as  $N \geq K$  and the channel estimates are full rank [29, 32]. In practice, the columns are then scaled to meet the power constraint. With ZF precoding, if the channel estimates were perfect and  $N \geq K$ , the interference terms in (2.7) ideally vanish, i. e.,  $\mathbf{h}_j^H \mathbf{w}_k^{\text{ZF}} \approx 0$  for  $j \neq k$ . However, ZF comes at the cost of higher computational complexity and increased sensitivity to CSI errors. If the channel estimates  $\hat{\mathbf{H}}$  are imperfect, as they always are in practice, then some interference remains. In particular, pilot contamination undermines ZF: if  $\hat{\mathbf{h}}_k$  contains a component of  $\mathbf{h}_j$  (for  $j \neq k$  sharing the pilot), the ZF precoder will inadvertently cancel part of UE  $k$ 's signal in an attempt to zero out interference, or vice versa, leading to degraded performance. In summary, ZF tends to outperform MR in interference-limited scenarios, e. g., cell-edge users in a cellular system or dense networks [32].

### 2.3.3 RZF / MMSE Precoding

A natural compromise between MR and ZF is to account for both noise and interference in the precoder design. This leads to the linear MMSE precoder, also known as RZF in Massive MIMO contexts. The idea is to minimize the Mean Squared Error (MSE) between transmitted and received signals for each UE. The RZF precoding matrix can be written as

$$\mathbf{W}^{\text{MMSE}} = \hat{\mathbf{H}}(\hat{\mathbf{H}}^H \hat{\mathbf{H}} + \alpha \mathbf{I}_K)^{-1}, \quad (2.9)$$

where  $\alpha$  is a regularization factor that depends on the noise and interference level. When  $\alpha \rightarrow 0$ , (2.9) reduces to ZF; when  $\alpha \rightarrow \infty$ , it behaves like MR. MMSE precoding outperforms MR and ZF in general, especially when the system is neither purely noise-limited nor completely interference-limited.

Figure 2.5 provides a directional view of the three beamforming techniques. For a two-UE toy Uniform Linear Array (ULA) with  $N = 16$  elements,  $\theta_1 = -20^\circ$  and  $\theta_2 = 30^\circ$ , the plotted quantity is the normalized beampattern for UE 1,  $B(\theta) = |\mathbf{w}_1^H \mathbf{a}(\theta)|^2$ .

## 2.4 PERFORMANCE METRICS

This section defines metrics to evaluate the system performance. The primary performance metric for communication in Massive MIMO is the SE per UE, measured in bits/s/Hz. We also consider the SINR at the receivers as an intermediate metric, since SE is typically a logarithmic function of SINR under Gaussian signaling. Additionally, we discuss how QoS constraints, such as guaranteeing a minimum rate to each UE or ensuring fairness, come into play in the context of power control.

### 2.4.1 SINR and achievable Spectral Efficiency

For a given instantaneous channel realization and a fixed precoding/combining scheme, the SINR for UE  $k$  can be derived from the received signal expressions. For example, on the DL with linear precoding as in (2.7), the SINR for UE  $j$  is

$$\text{SINR}_j = \frac{\rho_j |\mathbf{h}_j^H \mathbf{w}_j|^2}{\sum_{k \neq j} \rho_k |\mathbf{h}_j^H \mathbf{w}_k|^2 + \sigma^2}. \quad (2.10)$$

This ratio captures the power of the desired signal versus the power of interference plus noise. In an UL scenario with linear combining, a similar formula can be written: if  $\mathbf{v}_j$  is the receive combiner for UE  $j$ , then

$$\text{SINR}_j^{(\text{ul})} = \frac{p_j |\mathbf{v}_j^H \mathbf{h}_j|^2}{\sum_{k \neq j} p_k |\mathbf{v}_j^H \mathbf{h}_k|^2 + \sigma^2 |\mathbf{v}_j|^2}, \quad (2.11)$$

where  $p_k$  is the transmit power of UE  $k$  on uplink. These formulas assume the use of linear detectors treating interference as additional Gaussian noise. Under those

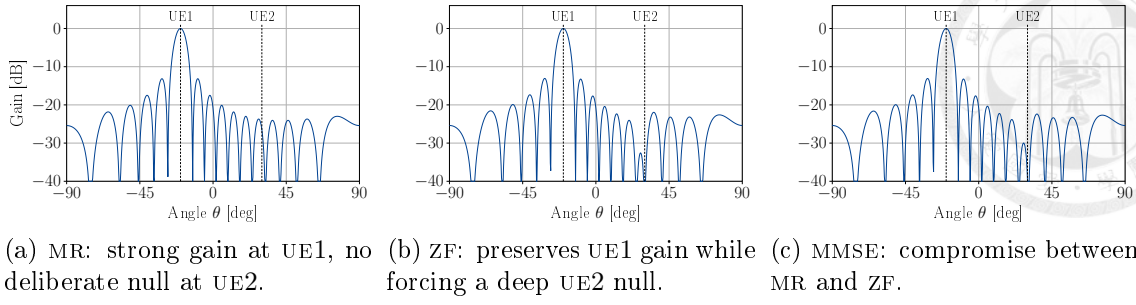


Figure 2.5: Comparison of beamforming strategies for the same two-UE ULA setup. MR provides array gain but leaks interference, ZF suppresses multiuser interference via nulling, and MMSE trades interference suppression against noise amplification through regularization.

assumptions, which are standard for performance analysis, an achievable SE for UE  $j$  is given by the Shannon capacity formula:

$$\text{SE}_j = \log_2 \left( 1 + \text{SINR}_j \right) \quad (\text{bits/s/Hz per UE}). \quad (2.12)$$

This is a per-channel-use efficiency; if the system bandwidth is  $B$  Hz, and assuming Gaussian codebooks and sufficiently long codes across fading realizations, the user's throughput is  $B \times \text{SE}_j$  bits/s.

Because channel estimates are not perfect and vary from block to block, one common approach is to consider the *ergodic Spectral Efficiency*, where we average over the distribution of channel estimates and noise, under the assumption of coding over many coherence blocks. A typical result is that the uplink Spectral Efficiency for UE  $k$  can be written in closed form as

$$\text{SE}_k^{(\text{ul})} = \left( 1 - \frac{\tau_p}{\tau_c} \right) \mathbb{E} \left[ \log_2 \left( 1 + \text{SINR}_k^{(\text{ul})} \right) \right], \quad (2.13)$$

where the expectation is with respect to the random channels and the pre-factor  $(1 - \frac{\tau_p}{\tau_c})$  accounts for the fraction of time used for data [33]. In many cases, techniques like *use-and-then-forget bounding* are applied to get a deterministic approximate SINR that yields a rigorous lower bound on  $\text{SE}_k$  [28, 29]. This treats the mean of the desired signal as the useful signal and regards all deviations, including those caused by channel estimation errors, as interference.

### 2.4.2 Spectral Efficiency gains

Under favorable conditions, Massive MIMO enables the sum SE to grow linearly with  $N$  and  $K$ . For example, with MR combining and i.i.d. Rayleigh channels,

$$\text{SINR}_k^{(\text{ul})} \xrightarrow{N \rightarrow \infty} \frac{N p_k \beta_k^2}{\sum_{i \neq k} p_i \beta_k \beta_i + \sigma^2}. \quad (2.14)$$

Thus, in practical regimes, Massive MIMO with sufficiently many antennas can yield very high SINRs for each UE, translating to high per-UE rates. Moreover, the ability to serve  $K$  UEs at once means the system throughput scales with  $K$  as well. However, increasing  $K$  also exacerbates pilot contamination and requires more power or more sophisticated precoding to manage interference.

### 2.4.3 Quality of Service constraints

In multi-user systems, network design must also consider fairness and QoS constraints. For instance, some UEs, perhaps in bad channel conditions, might achieve very low rates if all resources are allocated to maximize the sum rate, which often favors UEs with strong channels. Two common QoS-related performance metrics are:

*Minimum guaranteed rate:* We may require that each UE achieves at least a target rate  $R_{\min}$ . Massive MIMO can help improve the minimum UE rates dramatically by exploiting spatial diversity and coherent beamforming [16].

*Max-min fairness:* A design strategy where the objective is to maximize the smallest UE's rate:  $\max_{\rho_k} (\min_k R_k)$ . The solution often equalizes the UEs' SINRs at an optimal level [16], ensuring absolute fairness.

In the framework proposed in this thesis, we will primarily focus on the sum SE as the main performance metric, while ensuring a minimum SINR threshold for each UE to guarantee a baseline QoS.

## 2.5 CELL-FREE CONCEPT AND CELLULAR NETWORKS

The discussion so far has focused on the classical co-located Massive MIMO architecture, where all antennas belong to the same physical BS. For the rest of the thesis, however, the more relevant operating point is a distributed network in which many geographically separated APs cooperate over fronthaul. This motivates the transition to Cell-Free Massive MIMO, which retains the signal-processing intuition of Massive MIMO while fundamentally changing the architecture, the cooperation model, and the scalability bottlenecks.

Unlike a conventional cellular architecture, where each UE is served by a single BS with inevitable cell-edge performance degradation, a CF network deploys a large number of distributed APs that cooperate to jointly serve all UEs without cell boundaries [16, 31]. In essence, the antennas are spread across the environment and connected via fronthaul to one or multiple CPUs for coordination. This arrangement offers pronounced *macro-diversity* gains: each UE is likely to be near at least a few APs, ensuring strong channels to those APs and mitigating the traditional notion of cell edges. At the same time, by coordinating transmissions, the distributed APs can coherently combine signals from multiple locations, increasing the overall signal quality.

In the sections that follow, we explore how Cell-Free compares to traditional cellular networks, delve into user-centric clustering and AP selection strategies, and examine both centralized and distributed processing architectures. We also review the key protocols that enable CF networks to scale efficiently in real deployments. We then highlight the distinctive features of CF Massive MIMO and demonstrate their relevance to sensing and ISAC applications, establishing the foundation for the theoretical and experimental contributions presented in the subsequent chapters.

### 2.5.1 Cell-Free operation

One of the defining features of CF Massive MIMO is the absence of cell boundaries: in the canonical CF implementation with  $M$  APs, each having  $N$  antennas,

all  $MN$  antennas in the network cooperate as a single distributed array serving  $K$  UEs. Ideally, all APs simultaneously serve all UEs in a jointly coherent fashion. The operating regime of CF Massive MIMO often assumes many more distributed antennas than UEs ( $MN \gg K$ ), which, together with favorable propagation and channel hardening inherited from Massive MIMO, allows simple linear processing to achieve near-optimal performance. Studies like Ngo et al. [16] have shown that CF networks can indeed offer a more uniform rate distribution compared to cellular networks.

Figure 2.6 illustrates a generic CF Massive MIMO architecture, with  $M$  distributed APs connected to a CPU and jointly serving  $K$  UEs in the area. Despite its conceptual advantages, the fully centralized Cell-Free implementation comes with practical challenges in terms of fronthaul signaling and computational scalability. Much of the research in this area has focused on how to retain the benefits of the Cell-Free approach while addressing these implementation challenges.

## 2.6 USER-CENTRIC CLUSTERING AND AP SELECTION

User-centric clustering has therefore been proposed as a solution, whereby each UE is served by a subset of the most relevant APs [12]. In a user-centric design, each UE finds and activates its own cluster of serving APs, and these clusters can overlap for different UEs. Figure 2.7 illustrates this principle, where each UE activates a different serving AP set and multiple APs are shared across overlapping UEs clusters.

User-centric clustering capitalizes on the fact that far-away APs contribute little power and cause mostly interference. By letting each UE be served by only a subset of APs, one can greatly reduce the computational load and fronthaul messaging *without* significantly compromising performance [31]. Indeed, it has been observed that a well-chosen subset of APs can attain performance close to the ideal full cooperation case [12, 17, 37].

There are several typical policies for AP selection in user-centric clustering. One simple policy is *distance-based*, where each UE is served by the  $N_{\text{AP}}$  nearest APs. Another is *gain-based*: each UE is served by all APs whose average channel gain to

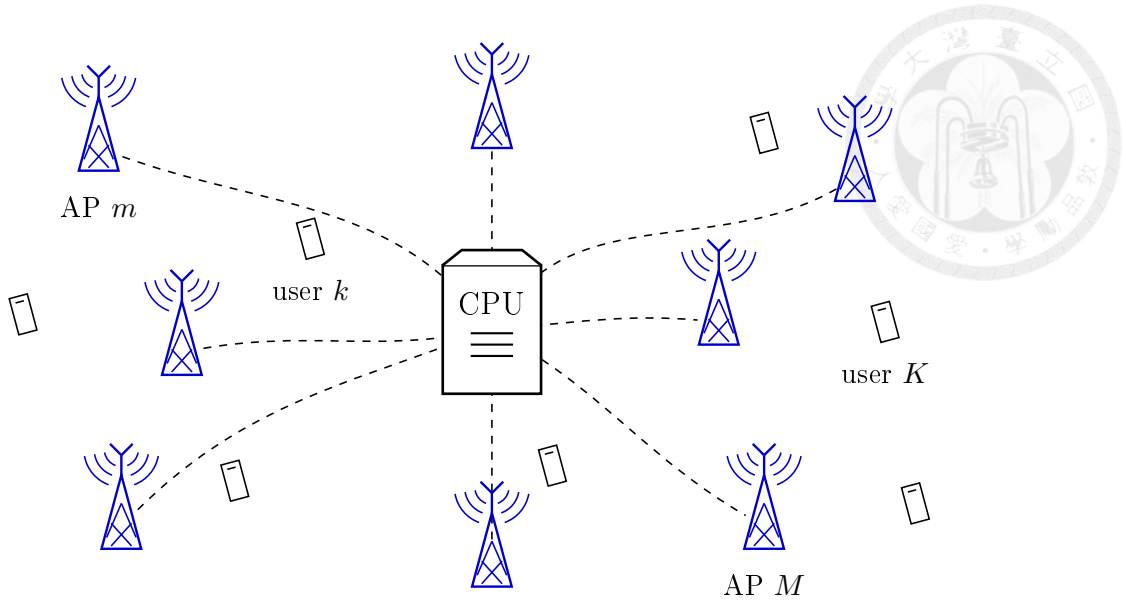


Figure 2.6: Cell-Free Massive MIMO system with distributed APs, a central CPU, and multiple UEs. All APs cooperate via fronthaul links to serve the UEs without cell boundaries.

that UE is within a certain margin of the strongest gain. More dynamic or optimized clustering methods have also been proposed, but the simple geographic or gain-based rules already capture most of the benefit [12]. Importantly, user-centric clustering can be leveraged to mitigate pilot contamination. Since an interfering UE  $j$  that is far from UE  $k$  will not be served by the same APs as user  $k$ , they could be assigned the same pilot sequence without causing significant contamination at those APs [12]. This is a smarter reuse of pilot resources based on the geographical UE separation and is one of the reasons Cell-Free systems can in principle support more UEs than the number of orthogonal pilot sequences  $\tau_p$  by appropriate pilot assignment strategies.

It is worth noting that the user-centric clusters are dynamic, as they can be updated as UEs move or traffic conditions change. This raises practical considerations such as how to hand over a UE from one set of APs to another as it moves, and how to efficiently inform the network nodes of clustering decisions. These issues, however, can be handled on a large-scale fading time scale, since the cluster selection is typically based on slowly-varying quantities. Research has shown that user-centric clustering, combined with proper power control, can maintain the CF network's feature of nearly uniform UE throughput while keeping complexity scal-

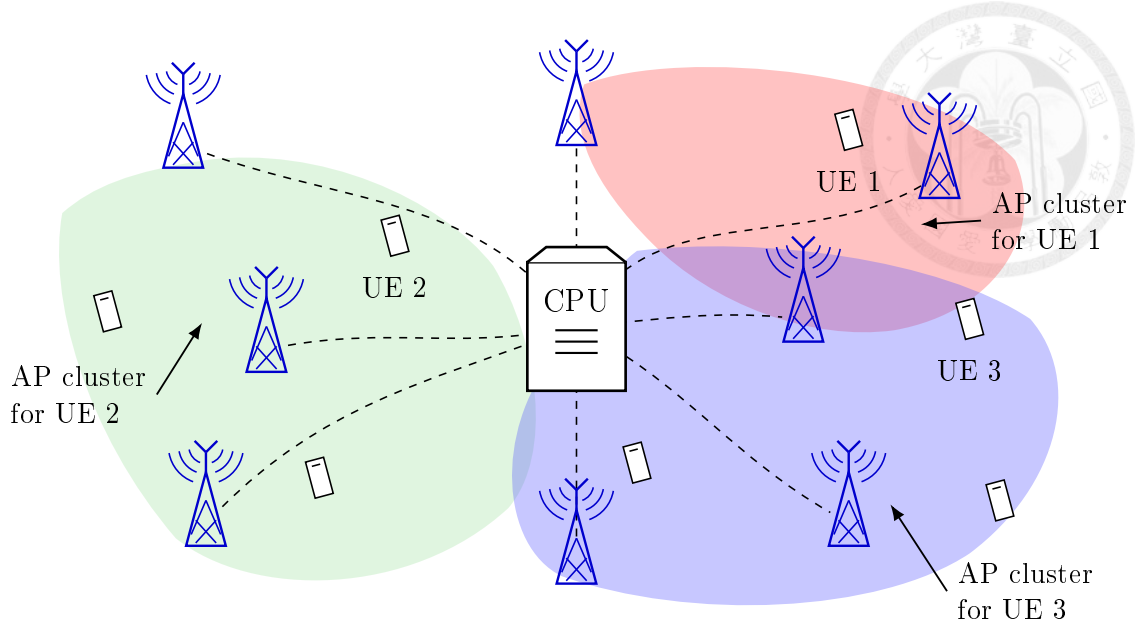


Figure 2.7: User-centric clustering in Cell-Free Massive MIMO. Each UE activates a personalized serving set of APs, and the resulting clusters overlap.

able. For instance, Interdonato et al. [37] demonstrated that serving each UE with as few as 5-10 APs is often enough to achieve most of the gains of full cooperation. Thus, user-centric clustering is a cornerstone in making CF Massive MIMO practical.

## 2.7 CENTRALIZED AND DISTRIBUTED PROCESSING

Another fundamental design aspect of CF Massive MIMO is where the heavy signal processing tasks, such as channel estimation, combining of UL signals, precoding of DL signals and power control are carried out. The research has identified two extreme paradigms: *Centralized processing* and *Distributed processing* [12, 17].

### 2.7.1 Centralized processing

In a centralized architecture, a single CPU collects either the raw signals or detailed CSI from all APs and computes the optimal combining/precoding for the network as a whole. In the DL, centralized processing entails the CPU computing all the transmit signals for every AP, e.g., by computing a precoding vector for each AP based on global CSI. This approach maximizes the performance, as it fully exploits

the array gain and interference cancellation potential of the distributed system [38]. However, fully centralized processing also maximizes the fronthaul overhead and computational burden at the CPU, since it must continuously receive a large volume of data from all APs and perform large-dimensional linear algebra operations in real-time.

### 2.7.2 *Distributed processing*

In contrast, distributed processing means that the APs perform most of the signal processing locally, using only locally obtained CSI, and the CPU aggregates more limited information, such as partially processed signals or final decoded data. In distributed UL processing, each AP estimates local channels and applies local combining. It forwards  $K$  scalar per UE to the CPU instead of  $N$ -antenna raw signals, and the CPU sums the  $M$  contributions per UE to recover the data. It has been shown to yield performance close to fully centralized MMSE processing in many cases, while drastically reducing the amount of data moved over the fronthaul [12, 37, 39].

Similarly, for the downlink, distributed processing means the CPU might only send the data payloads, and then each AP uses its local channel estimates to beam-form the signal toward the UEs. The APs thus require only local CSI, and the InUI is managed in a decentralized way. The benefit of distributed processing is a significant reduction in required fronthaul capacity and greater parallelization of processing tasks. The trade-off is a potential performance loss compared to fully centralized optimal filtering, since the lack of global CSI means suboptimal combining/precoding. Nevertheless, extensive evaluations have found that cleverly designed distributed schemes can achieve near-optimal SE in CF systems [12, 38].

### 2.7.3 *Hybrid processing*

In practice, real implementations may adopt a hybrid of these two extremes. Some processing steps might be centralized and others distributed. A common approach is to do as much as possible in a distributed manner and only centralize what

truly benefits from network-wide coordination. For example, channel estimation and initial combining could be done at the APs, whereas final decoding, which might benefit from combining signals across APs, could be done at the CPU.

Figure 2.8 summarizes the processing split choices using the same physical AP-CPU topology and highlights the resulting fronthaul payloads.

## 2.8 SCALABLE CELL-FREE PROTOCOLS

A critical question for CF Massive MIMO is how to *scale up* the network to hundreds of APs and UEs while keeping the overhead and complexity manageable. Therefore, a lot of research has focused on *scalable protocols* for CF Massive MIMO, which aim to preserve the performance benefits while reducing the internal overhead.

For instance, Bjornson and Sanguinetti [12] demonstrated a scalable user-centric design which employs a local MMSE combining that considers interference only from overlapping clusters. This design achieved near-optimal SE while keeping complexity and overhead growing roughly linearly, with the number of APs.

Overall, the protocols and architectural features discussed in this section aim to ensure that CF Massive MIMO can be implemented in large-scale networks without an explosion of complexity. By intelligently selecting who cooperates with whom, processing data as locally as possible, and using the available resources flexibly, one can retain the advantages of the Cell-Free approach in a practical system.

## 2.9 SENSING IN CF ARCHITECTURES

The Cell-Free Massive MIMO architecture provides fertile ground for ISAC applications. Its inherent macro-diversity and cooperative processing, originally intended to improve communication performance, also translate into spatial diversity and cooperation for sensing. A network of many APs can therefore function as a high-performance radar system that is *multi-static*, flexible, and robust.

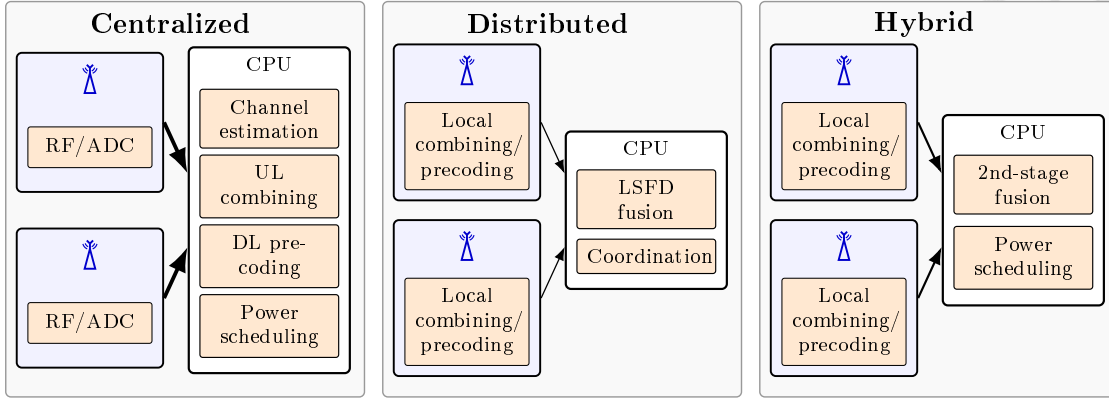


Figure 2.8: Processing split in Cell-Free Massive MIMO across the three designs. Centralized processing sends near-raw baseband information to the CPU; distributed processing keeps channel estimation and local combining/precoding at APs and forwards low-dimensional summaries; hybrid processing combines local dimensionality reduction at APs with second-stage fusion and global coordination at the CPU.

In later chapters, these characteristics will be exploited when designing and evaluating ISAC techniques. For instance, we will consider how targets can be detected and localized by a distributed CF network and how communication and sensing functionalities can be balanced. The key message is that a CF system can enable *ubiquitous sensing coverage* much like it enables ubiquitous connectivity, thereby unlocking new capabilities for 6G wireless networks.



# 3 | RADAR AND SENSING FUNDAMENTALS

This chapter provides a comprehensive background on radar and sensing theory, laying the groundwork for subsequent ISAC-specific discussions. Radar is a sensing modality that uses electromagnetic waves to detect objects and estimate key parameters such as range, velocity, and Angle of Arrival (AoA) [40, 41]. By transmitting radio signals and processing echoes reflected from targets, radar systems can determine target presence, localize objects, and track them over time [10]. Radar technology has been pivotal in applications ranging from defense, e. g., for aircraft and missile detection, to civilian domains like automotive collision avoidance and weather monitoring, due to its all-weather operation and long-range capabilities [40].

This chapter focuses exclusively on radar sensing fundamentals, first distinguishing the three core tasks: detection, estimation, and tracking, while leaving communication aspects aside. It then introduces detection theory through binary hypothesis testing and likelihood-ratio tests, it addresses estimation of key target parameters—range, Doppler velocity, and angle—using signal models and matched filtering, and evaluates accuracy via metrics such as MSE and bounds such as the Cramér-Rao Bound (CRB). The chapter also discusses clutter and interference, including SCNR and Doppler processing for moving targets. Finally, it compares mono-static, bi-static, and multi-static radar configurations, highlighting the geometric and spatial-diversity advantages of multi-static sensing for Cell-Free MIMO systems, and defines the main performance metrics.

### 3.1 SENSING TASKS

Radar sensing tasks can be broadly grouped into detection, estimation, and tracking:

**DETECTION.** This is the process of determining whether a target is present in the received signals. It is a binary decision problem that uses statistical hypothesis testing to determine whether observed echoes are target returns or noise/clutter [25]. Detection yields a discrete decision and precedes further processing.

**ESTIMATION.** Estimation measures the parameters of one or more detected targets. It is a continuous inference task: from the received signals, the radar infers the numerical values of target parameters. Accuracy is typically quantified by metrics such as MSE, while bounds such as the CRB set theoretical limits for unbiased estimators under a given signal model and Signal-to-Noise Ratio (SNR) [41].

**TRACKING.** Tracking continuously monitors a target's trajectory by updating its estimated state as new measurements arrive. The radar associates detections across scans and smooths measurement noise to maintain a stable track. Tracking relies on reliable detections and accurate instantaneous estimates [42].

Figure 3.1 summarizes these tasks as a practical radar processing pipeline from raw returns to recursive target-state updates.

### 3.2 DETECTION THEORY

Radar detection involves deciding between two hypotheses: the null hypothesis  $H_0$  that no target is present, and the alternate hypothesis  $H_1$  that a target is present and reflecting the radar signal. A simple analytical model for a radar detection problem is:

$$H_0 : y(t) = n(t) , \quad (3.1)$$

$$H_1 : y(t) = \alpha s(t) + n(t) , \quad (3.2)$$

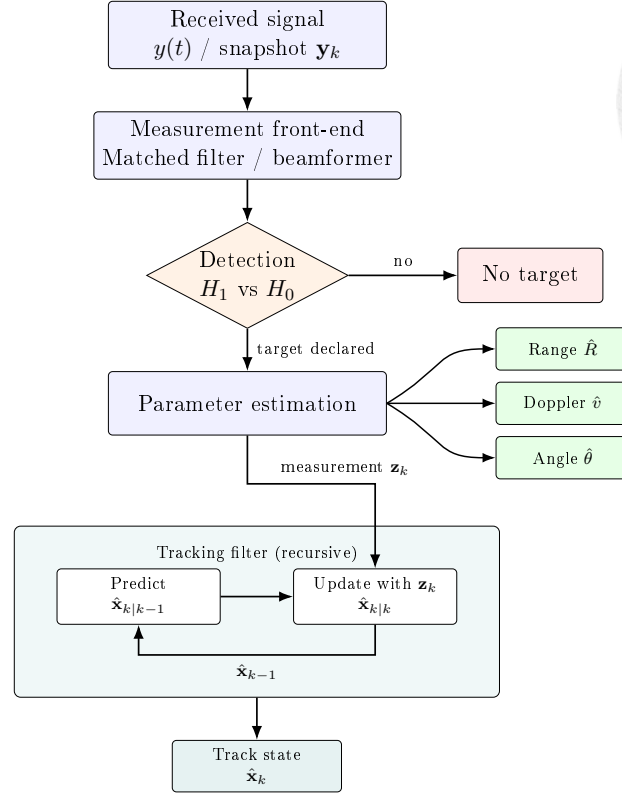


Figure 3.1: Radar sensing pipeline from received signal processing to target tracking. A measurement front-end converts raw returns into detection statistics; detection acts as a binary gate, and only declared targets proceed to parameter estimation. The tracking stage is shown as a recursive predict-update filter driven by measurements  $\mathbf{z}_k$  and producing a state estimate  $\hat{\mathbf{x}}_k$  over time.

where  $s(t)$  is the transmitted radar waveform assumed known at the receiver,  $\alpha$  is a complex amplitude proportional to the target's reflection coefficient, incorporating path loss and RCS, and  $n(t)$  is additive noise. The optimal decision rule to distinguish  $H_1$  from  $H_0$  in a statistical sense is given by the Neyman-Pearson likelihood ratio test [25]. One computes the likelihood ratio

$$\Lambda(\mathbf{y}) = \frac{p(\mathbf{y} | H_1)}{p(\mathbf{y} | H_0)}, \quad (3.3)$$

and compares it to a threshold  $\eta$ : decide  $H_1$  if  $\Lambda(\mathbf{y}) > \eta$ , otherwise decide  $H_0$ . The threshold  $\eta$  is set according to the desired  $P_{\text{FA}}$ .

Two key detector performance metrics are the  $P_{\text{D}}$  and the  $P_{\text{FA}}$  [25]. Here,  $P_{\text{D}}$  is the probability of correctly declaring a target present when it is present, whereas

$P_{\text{FA}}$  is the probability of declaring a target present when no target is present. These are defined as

$$P_{\text{FA}} = \Pr \{ \text{decide } H_1 \mid H_0 \}, \quad P_{\text{D}} = \Pr \{ \text{decide } H_1 \mid H_1 \}. \quad (3.4)$$

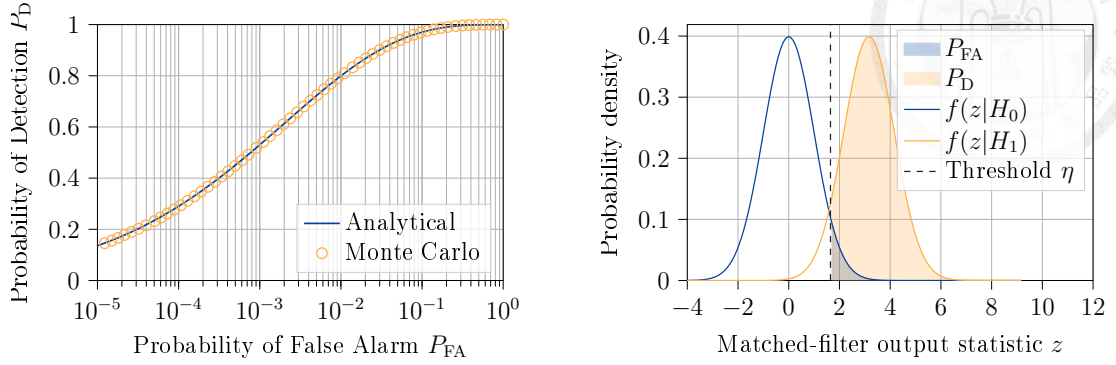
By lowering the detection threshold  $\eta$ , the radar can achieve a higher  $P_{\text{D}}$  at the expense of a higher  $P_{\text{FA}}$ ; conversely, a higher threshold reduces false alarms but also makes genuine targets harder to detect. The trade-off is typically visualized via the ROC curve, which plots  $P_{\text{D}}$  against  $P_{\text{FA}}$  as the threshold  $\eta$  is varied [25], as illustrated in Figure 3.2. The corresponding area under the curve is then used as a threshold-independent summary of the detector performance. Radar designers often choose  $\eta$  to meet a specified  $P_{\text{FA}}$  (e.g.,  $10^{-2}$  per observation) and then evaluate the corresponding  $P_{\text{D}}$  at that operating point. In practice, maintaining a constant  $P_{\text{FA}}$  across varying interference conditions leads to the concept of Constant False Alarm Rate (CFAR) detectors, which dynamically adjust the threshold based on background noise/clutter estimates [41].

### 3.3 ESTIMATION BASICS

Beyond detection of a target's presence, radars must estimate target parameters such as distance to the target, relative speed, and AoA. Estimation is typically performed on the signals after detection to extract these quantitative parameters.

**RANGE ESTIMATION.** The range  $R$  to a target is related to the round-trip propagation delay  $\tau$  of the radar signal by  $R = c\frac{\tau}{2}$ , where  $c$  is the speed of light. In practice the received signal is correlated with a replica of the transmitted waveform, and the time at which the correlation output peaks gives the time delay  $\hat{\tau}$  [41]. The accuracy of range estimation depends on the signal bandwidth: a larger bandwidth yields a narrower correlation peak and thus finer range resolution.

**VELOCITY ESTIMATION.** A moving target causes a Doppler frequency shift in the reflected signal. For a monostatic radar, a target with radial velocity  $v$  induces



(a) For a threshold  $\eta$ , the false-alarm and detection probabilities are  $P_{FA} = Q(\eta)$  and  $P_D = Q(\eta - \sqrt{2\text{SNR}})$ , yielding the closed-form ROC  $P_D = Q(Q^{-1}(P_{FA}) - \sqrt{2\text{SNR}})$ .

(b) Trade-off between  $P_{FA} = Q(\eta)$  (right tail, red) and  $P_D = 1 - Q(\eta)$  (right tail, blue), where  $\eta$  is the detection threshold.

Figure 3.2: Coherent Neyman-Pearson detection in AWGN after matched filtering and noise whitening, with  $\text{SNR} = 7$  dB. In (a), the analytical ROC is compared against Monte-Carlo points. Higher SNR yields a steeper ROC curve, achieving better detection performance at lower false alarm rates. In (b), the test-statistic densities under  $H_0$  and  $H_1$ , the threshold  $\eta$ , and the shaded regions corresponding to  $P_{FA}$  and  $P_D$  are shown. As  $\eta$  decreases, both  $P_{FA}$  and  $P_D$  change, demonstrating the fundamental detection trade-off.

a Doppler shift of approximately  $f_D = \frac{2v}{\lambda}$ , where  $\lambda$  is the radar wavelength [40]. Thus, by measuring the Doppler frequency  $f_D$  of the echo, one can estimate the radial velocity as  $\hat{v} \approx \frac{\lambda}{2} f_D$ . Doppler processing also serves to reject stationary clutter, as static objects will have  $f_D \approx 0$  and can be filtered out, leaving moving targets for detection.

**ANGLE OF ARRIVAL ESTIMATION.** By comparing the phase or amplitude of the incoming signal across multiple antenna elements, the radar can infer the direction from which the signal arrived. A basic method is beamforming or scanning: the array is electronically steered to look in different directions, and the direction that yields the strongest echo indicates the target's angle [40].

### 3.4 CLUTTER AND INTERFERENCE

Clutter refers to unwanted echoes from environmental objects that are not the intended targets [41]. Clutter returns can be much stronger than thermal noise and can thus mask the presence of real targets or cause false detections. Clutter echoes often have non-Gaussian amplitude statistics that depend on the nature of the reflecting surfaces and the radar resolution cell. The severity of clutter is characterized by the Signal-to-Clutter-and-Noise Ratio (SCNR), where both clutter and noise are present.

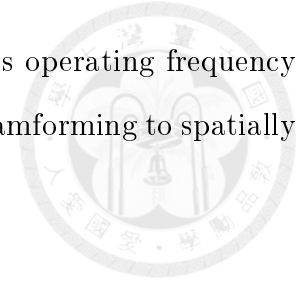
A variety of signal processing techniques are employed to mitigate clutter and improve target detection. One foundational approach is Moving Target Indicator (MTI), which exploits the fact that many clutter sources are stationary or slow-moving. By using a high-pass Doppler filter or subtracting returns from pulse-to-pulse, MTI cancels out echoes that do not change between pulses [40]. However, clutter is not always perfectly stationary, and some residual leakage often persists.

Another crucial technique is the use of CFAR detectors, which estimate the local clutter-plus-noise level in real time and adjust the detection threshold so that a preset  $P_{\text{FA}}$  is maintained despite variations in interference power [41]. In regions of heavy clutter the threshold rises, reducing false alarms, while in cleaner regions the threshold can be lower to catch weaker targets.

Other clutter-mitigation methods include clutter mapping, which consists of recording the locations and amplitudes of permanent echoes and subtracting them out in subsequent radar returns. This is particularly effective for fixed installations where the clutter environment is stable over time, and will be implemented in the proposed Cell-Free ISAC framework in Chapter 8.

In addition to natural clutter, radar systems may encounter deliberate or accidental interference. For example, other radars operating on the same frequency can cause mutual interference, and in military contexts, intentional jamming signals may be transmitted by adversaries to blind the radar. Unlike clutter, interference signals may have different waveforms or modulation. Mitigating interference often

requires different strategies: for instance, changing the radar's operating frequency to avoid continuous interference, time gating, and adaptive beamforming to spatially suppress incoming interference sources [40].



## 3.5 MULTI-STATIC SENSING BASICS

The spatial configuration of transmitters and receivers significantly impacts sensing performance and capabilities. The three primary configurations are monostatic, bistatic, and multi-static radar, each with distinct geometric arrangements and implications for detection and estimation [40, 41]. Figure 3.3 summarizes these geometries and highlights the key propagation paths and angles.

### 3.5.1 *Monostatic Radar*

In a monostatic radar, the transmitter and receiver are co-located, often even sharing the same antenna via a duplexer that switches between transmit and receive modes [40]. Monostatic radars are straightforward to implement because the geometry is simple and the radar has direct knowledge of the transmitted signal for reception. Monostatic radar returns often include strong self-clutter, e.g., ground reflections directly beneath the radar, which must be handled via MTI or other means.

### 3.5.2 *Bistatic Radar*

In a bistatic configuration, the transmitter and receiver are at different locations, separated by some distance [40]. The transmitter illuminates the target, and the receiver listens for the scattered energy. Bistatic radar introduces additional geometric complexity: the propagation paths from transmitter to target and target to receiver form a bistatic angle. A target's effective RCS can vary with bistatic angle, sometimes offering a larger reflection to the receiver than it would to a co-located transmitter, but other times smaller [41]. However, bistatic radars face significant challenges in synchronization and signal processing. If using an independent trans-

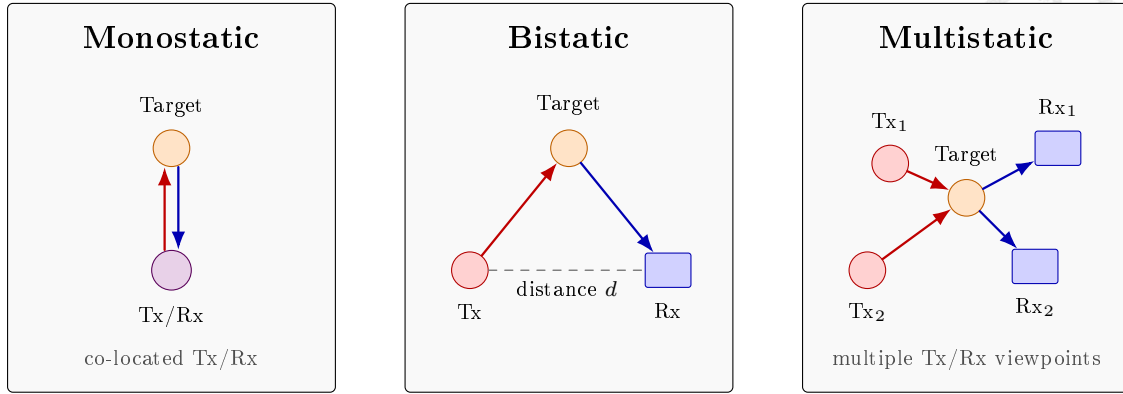


Figure 3.3: Geometry comparison for monostatic, bistatic, and multistatic sensing. In monostatic radar, Tx and Rx are co-located. In bistatic radar, Tx and Rx are separated by a distance  $d$ . In multistatic radar, multiple Tx/Rx nodes observe the same target under different aspect angles, enabling spatial diversity.

mitter, the receiver does not inherently know the exact transmit signal timing, and it must have a method to synchronize [41].

### 3.5.3 Multi-static Radar

In a multi-static network, several radar units work jointly to sense the environment. Multi-static radar offers many potential benefits thanks to spatial diversity: each radar node can provide a different view of the target, and by combining information, the system as a whole can achieve better performance than any individual radar. For instance, a target that has a weak echo in one angle might produce a stronger echo at a different angle; multi-static radars can thus detect low-RCS or dynamically oriented targets more reliably [10]. Also, multiple receivers can triangulate a target's position, estimating the target's location and velocity with higher precision. Multi-static configurations can also improve tracking stability, since a target is less likely to drop out of all radar beams simultaneously.

However, the advantages of multi-static radar come with increased complexity. The system must coordinate among multiple nodes, which requires a communication network and often a CPU to fuse data. Achieving tight phase coherence is even more challenging, as it requires calibration of oscillators and compensation of propagation delays between the nodes. The data fusion aspect can be approached in two ways:

1. *Centralized*: each receiver forwards raw or pre-processed echoes to a CPU for joint detection/estimation. This can approach best performance, but requires high fronthaul bandwidth and central compute.
2. *Distributed*: each radar makes local detections/estimates and only shares summaries for fusion. This is more scalable and robust, but can sacrifice performance if local information is limited.

In the context of Cell-Free Massive MIMO ISAC systems, the sensing function inherently operates in a multi-static fashion. This is essentially a networked radar: it capitalizes on the spatial diversity and the increased baseline for improved localization accuracy [10].

Geometry plays a crucial role: the precision in target localization will depend on how the APs are positioned relative to the target. The principles of multi-static radar thus directly inform CF ISAC design, but additional challenges arise due to the dual use for communication and the necessity to synchronize with communication slots. Nevertheless, the multi-static perspective makes clear that CF ISAC has the potential for superior sensing performance: by coherently combining reflections at multiple APs, the system can achieve an array gain and resilience that a traditional monostatic radar would not [10].

## 3.6 PERFORMANCE METRICS

Several quantitative metrics are used to characterize radar performance. Each metric captures a different aspect of the radar's ability to detect and estimate target parameters under various conditions. This section defines the key performance metrics that will be used throughout this thesis, including RCS, SNR and SCNR.

### 3.6.1 Radar Cross Section

As introduced earlier, the RCS  $\sigma$  of a target is a measure of how strongly the target reflects radar energy back toward the radar. Intuitively,  $\sigma$  has units of area ( $\text{m}^2$ ) and

can be thought of as the equivalent size of an isotropic reflector that would return the same echo power. Formally, RCS is defined by the far-field ratio of scattered power per unit solid angle in the direction of the radar to the incident power density on the target [41]. Mathematically:

$$\sigma = \lim_{r \rightarrow \infty} 4\pi r^2 \frac{P_{\text{scatter}}}{P_{\text{incident}}}, \quad (3.5)$$

where  $P_{\text{incident}}$  is the power density of the signal arriving at the target, and  $P_{\text{scatter}}$  is the power per unit solid angle scattered back toward the radar at distance  $r$ . The RCS of a target depends on its size, shape, material, and orientation relative to the radar, as well as the transmit signal's frequency and polarization. Table 3.1 provides order-of-magnitude RCS values for various objects [40].

### 3.6.2 SNR and SCNR

The SNR at the output of the radar's signal processor is the key parameter that determines  $P_D$  and  $P_{FA}$  via the detection theory discussed in Section 3.2.

In the presence of clutter, the relevant metric extends to include clutter power. SCNR is defined as

$$\text{SCNR} = \frac{\text{Signal power}}{\text{Clutter power} + \text{Thermal Noise}}. \quad (3.6)$$

SCNR essentially replaces SNR when clutter dominates; a target might have high SNR but still be hard to detect if the clutter power is much higher. Many of the techniques in clutter suppression aim to improve the SCNR by reducing the clutter power through filtering or cancelation.

### 3.6.3 Resolution

Resolution is often defined as the minimum separation at which two identical targets can be resolved as distinct. Range resolution  $\Delta R$  is determined by the pulse duration or signal bandwidth. If  $\tau_p$  is the pulse width, a rough estimate is  $\Delta R \approx \frac{c\tau_p}{2}$ . In terms of signal bandwidth  $B$ , this can be written as  $\Delta R \approx \frac{c}{2B}$  [41].

| TARGET                    | APPROX. RCS ( $\sigma$ )       |
|---------------------------|--------------------------------|
| Insect                    | $10^{-5} \text{ m}^2$          |
| Small bird                | $10^{-2} \text{ m}^2$          |
| Human (adult)             | $1 \text{ m}^2$                |
| Fighter aircraft          | $5 \text{ m}^2$ (front aspect) |
| Sedan automobile          | $20 \text{ m}^2$               |
| Large commercial airliner | $100 \text{ m}^2$              |
| Large ship (metal hull)   | $10^6 \text{ m}^2$             |

Table 3.1: Typical RCS values for various targets (order-of-magnitude) [40]. Actual RCS can vary with aspect angle and frequency.

Doppler resolution is set by the coherent integration time:  $\Delta f_D \approx 1/T$ , which for monostatic radars translates to a velocity resolution of  $\Delta v \approx \frac{\lambda}{2} \frac{1}{T}$ .

Angular resolution depends on the antenna aperture size and beamwidth. To resolve two targets in angle, their separation must exceed the beamwidth or else their echoes will appear in the same beam. Phased arrays with many elements achieve very narrow beams, hence high angular resolution.

#### 3.6.4 Other Metrics

In the context of ISAC, there are also emerging metrics such as sensing capacity or mutual information of the sensing channel, which attempt to quantify how much information about the target or environment is gained, analogous to communication capacity [10]. However, the classical metrics of  $P_D$ ,  $P_{FA}$ , resolution, and SNR/SCNR remain the foundational measures by which radar sensing performance is judged and compared.





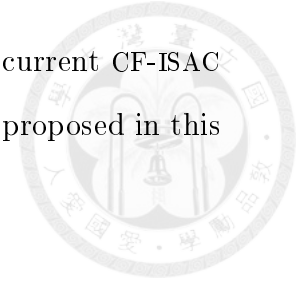
# 4 | ISAC AND CELL-FREE ISAC

The convergence of sensing and communication in wireless networks, commonly referred to as Integrated Sensing and Communication (ISAC), has been identified as a foundational technology for future beyond-5G and 6G systems [43]. ISAC envisions a unified framework where the same waveform, hardware, and time-frequency resources simultaneously serve radar sensing and data transmission [43]. This dual-role operation can dramatically improve SE and reduce hardware costs, and it enables new applications ranging from autonomous driving and smart cities to extended reality and environmental monitoring [43]. Accordingly, ISAC has been incorporated as a key feature in emerging 6G physical layer designs.

This chapter delves into the modern ISAC paradigm and its extension to CF MMIMO networks. It begins by categorizing the principal approaches to ISAC, then covers how ISAC signals are organized in a transmission frame, and next the landscape of sensing performance metrics used in ISAC is surveyed, including newly proposed information- and error-based metrics that mirror communication metrics. Formulations of the *communication-sensing trade-off* are also discussed, describing how researchers optimize the balance between high communication rates and accurate sensing.

The chapter's second half focuses on CF-ISAC, where distributed APs jointly serve UEs and jointly perform sensing. Typical system models in the CF-ISAC literature are outlined, and established strategies for multi-static target detection and resource allocation are then reviewed. The chapter proceeds to compare duplexing options such as conventional TDD and emerging schemes like DTDD, which allow concurrent

UL and DL operation in different APs. Finally, scalability issues in current CF-ISAC designs are identified, which motivate the more advanced solutions proposed in this work.



## 4.1 ISAC PARADIGMS

ISAC can be realized via different design paradigms, distinguished by whether sensing functionality is built *on top of* existing communication signals or via dedicated radar waveforms. ISAC can be broadly classified into the following three categories [43].

**COMMUNICATION-CENTRIC:** Priority is given to maintaining communication performance, while sensing capabilities are embedded by opportunistically reusing the communication signals. This paradigm ensures compatibility with existing networks and does not require additional spectrum.

**RADAR-CENTRIC:** This approach treats high-performance sensing as the primary design driver and allows the communication data to be superimposed or embedded within radar waveforms [43]. In these *radar-first* designs, one might start with a classical radar signal and then modulate information bits onto it in a way that minimally distorts the radar's characteristics.

**JOINT DESIGN:** Joint design approaches formulate optimization problems to find waveforms, beamforming vectors, or resource allocations that balance the needs of both. For example, the transmit signal can be a superposition of a dedicated radar waveform and a communication signal, with their relative power or precoder chosen to optimize a weighted sum of communication rate and sensing accuracy.

We will later see in Section 4.5 how such trade-offs are formulated mathematically. Importantly, all these paradigms still share the common goal of a single platform performing two functions; the classification simply guides whether we start from a communications perspective, a radar perspective, or a co-design perspective.

## 4.2 FRAME STRUCTURE FOR ISAC

Contemporary wireless systems transmit in frames that contain deterministic *pilot* signals for synchronization and channel estimation and random *data payload* signals carrying user information, as discussed in Section 2.1.2. These pilot sequences are carefully designed to have favorable autocorrelation properties, enabling accurate channel estimation and, in an ISAC context, good range-Doppler resolution for sensing targets.

Early ISAC proposals therefore treated pilots as radar *probe signals*, leveraging the fact that the transmitter can correlate received echoes with the known pilot pattern to detect targets. Indeed, a number of works demonstrated that standard communication reference signals can double as radar signals [43]. One concrete example is the use of the 5G SSB burst for sensing: the SSB contains a deterministic signal that can be exploited to illuminate the environment, and by processing its echoes one can detect objects such as low-altitude drones [22].

However, relying solely on pilots limits the sensing performance, because pilots occupy only a small fraction of each frame and are usually not optimized for sensing. The next step in frame design is therefore to integrate sensing functionality during data payload periods. Two main strategies have emerged: *time orthogonal* and *simultaneous* arrangements. In a time-orthogonal framing, certain subslots or subframes are reserved for purely sensing transmissions ensuring minimal mutual interference, but at the cost of dedicating some fraction of time exclusively to sensing. The more integrated approach embeds sensing signals into the same time-frequency slots as data. One method is to superimpose an additional sensing waveform onto the data signals:

$$x(t) = x_{\text{data}}(t) + x_{\text{sensing}}(t), \quad (4.1)$$

where  $x_{\text{sensing}}(t)$  might be a known radar-centric sequence that is added to the communication waveform. This superposition approach was analyzed in a recent study by Behdad et al. [18], who considered two scenarios:

- Using only the downlink communication beams for sensing.
- Adding a dedicated sensing beam with some allocated power alongside the communication beams.



The latter approach explicitly inserts extra sensing symbols in the frame and steers a beam towards a potential target. This yields a truly integrated frame structure: deterministic pilots, random data, *and possibly additional sensing pulses* all coexist. Taken together, these designs span the main frame-level resource allocation modes for ISAC, from pilot-only sensing to dedicated sensing bursts and fully integrated superimposed operation, as illustrated in Figure 4.1. Among the designs discussed above, this latter fully integrated approach is the one adopted in the proposed CF-ISAC framework of this thesis. In particular, we combine payload-aided sensing with dedicated sensing signals to deliver robust multi-static sensing while preserving communication performance.

A key design issue for ISAC frame structure is managing interference between the two functionalities. If sensing and communication signals overlap in time-frequency, one must ensure that the communication receivers are not excessively impaired by the sensing signals, and vice versa for the sensing receiver processing the echoes. Techniques such as orthogonal waveform design or precoding can be used to mitigate this interference. For example, a precoder can be designed to project the sensing waveform onto the nullspace of the multi-user communication channel, effectively *nulling out* the sensing signal at the UEs while still illuminating the target.

### 4.3 PAYLOAD-AIDED SENSING

Payload-aided sensing refers to repurposing the communication data symbols to assist in sensing tasks. This way, sensing can be performed opportunistically during most of the communication operation. Recent research has quantified the gains: for example, Xu et al. [13] show that by leveraging data payload signals for target

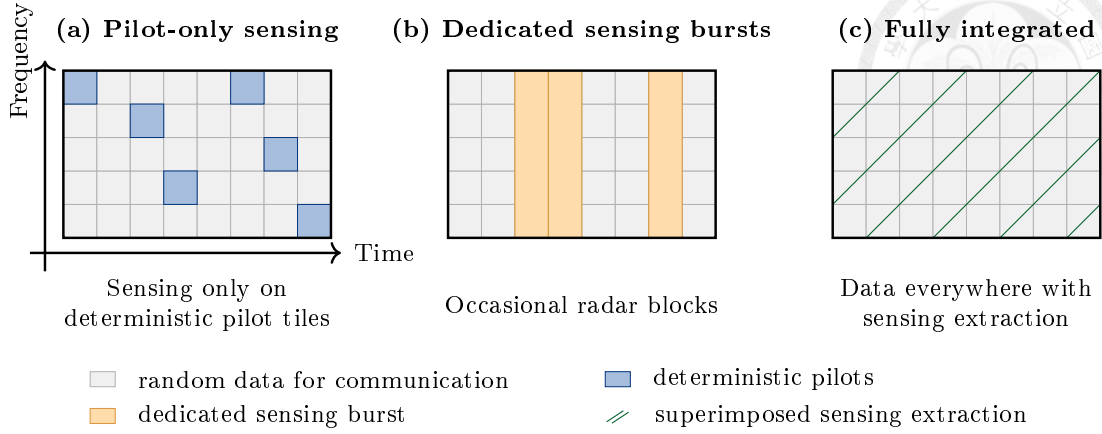


Figure 4.1: Frame-structure gallery for ISAC on an OFDM-like time-frequency grid: (a) pilot-only sensing, (b) dedicated sensing bursts, and (c) fully integrated/superimposed operation. The three panels visualize how sensing and communication share resources under each strategy.

sensing, the sensing error can be reduced by up to 5.6 dB compared to a conventional pilot-only approach.

However, using random data for sensing is not as straightforward as using known pilots. The primary challenge is that the data sequence may be unknown to the sensing receiver. In a bi-static or multi-static scenario, extra coordination overhead and latency are introduced because the data payload must be shared over back-haul. Moreover, even if the data sequence is known, it is not designed for good autocorrelation properties.

Another cost of using payload for sensing is potential degradation of communication performance or increased receiver complexity. One approach is joint signal processing: treat the target echo as just another channel path in the communication receiver, so that it can be estimated and subtracted if necessary. In summary, payload-aided sensing greatly boosts the sensing *duty cycle* and energy, thereby improving detection and estimation capabilities. The trade-off is the added complexity of handling random signals.

## 4.4 SENSING METRICS LANDSCAPE

Measuring sensing performance in an ISAC system requires metrics that can meaningfully be compared or traded off with communication metrics. Historically, radar sensing performance has been characterized by metrics quite different from those in communications, such as the metrics discussed in Section 3.2 and Section 3.6. In this section, several important newly proposed sensing metrics will be defined and discussed.

### 4.4.1 *Sensing Mutual Information and Sensing Capacity*

The idea of Sensing Mutual Information (SMI) is to treat the radar sensing process in an information-theoretic way: one considers the random variable representing the target and the random observations, and computes the mutual information [44]. The SMI measures in bits how much information the observations carry about the target. SMI is appealing because it is unbounded and accumulative, since more signal energy or bandwidth yields higher SMI, and it has a solid theoretical footing which allows unification with communication metrics [45].

Xie et al. [44] derived a closed-form expression for the SMI achieved by random ISAC waveforms in a Gaussian target-return channel. They further established connections between SMI and traditional metrics like estimation MSE and the Bayesian Cramér-Rao Lower Bound (CRLB). By dividing SMI by the observation time, one can define a sensing rate in bits per second that quantifies the maximal rate at which target information is obtained [46].

It is important to note that maximizing SMI does not always exactly correspond to maximizing  $P_D$  or minimizing estimation error, but in many cases they are aligned [18]. Therefore, some works choose SMI as the optimization objective for beamforming or power allocation [45].

#### 4.4.2 *MSE-Based Metrics: CRB and E-LMMSE*

When the sensing task is not just to detect a target but to estimate its parameters, the accuracy is often measured by the MSE or related benchmarks. A fundamental lower bound on estimation accuracy is the CRLB, which gives the minimum achievable variance of an unbiased estimator for a given parameter. In ISAC contexts, one might talk about the CRLB for target range estimation or the CRLB for target RCS estimation. However, CRLB is a bound; to use it as a metric, one typically ensures the estimator is efficient, i.e. its MSE attains the CRLB.

#### 4.4.3 *SNR, SINR, and SCNR proxies*

Finally, a very common practice in both radar and ISAC papers is to use various SNR-like ratios as surrogate metrics for sensing. In ISAC, one may define a sensing SNR for a given target, which is essentially the power of the target echo divided by the noise power at the sensing receiver. Behdad et al. [18], for instance, formulate a power allocation algorithm aiming to maximize the worst-case sensing SNR across the network while maintaining communication SINR requirements. The use of SNR or SINR in sensing is convenient because it converts a sensing design problem into a form similar to a communication design problem.

One must be cautious, however, because a high SNR might also amplify interference from clutter if not properly accounted for, which is why SCNR is used instead of pure SNR in some cases. This is why many ISAC works report communication SE versus sensing SCNR trade-offs, or guarantee a certain minimum communication SINR while maximizing the sensing SCNR.

## 4.5 COMMUNICATION-SENSING TRADE-OFF

An integrated ISAC system inherently faces a multi-objective optimization problem: improving the sensing performance usually requires using more energy or Degrees of Freedom (DoF) for sensing, which might detract from communication resources, and

vice versa. Formulating and solving this trade-off is a central theme in ISAC research. The trade-off between communication and sensing objectives can be handled through several methodological approaches:

- *Pareto boundary exploration:* A Pareto optimal trade-off curve can be obtained by finding all operating points where it is impossible to improve one objective without degrading the other. For instance, one might plot the maximum achievable  $P_D$  for a target as a function of the communication sum-rate in a network. An illustrative Pareto-front example is shown in Figure 4.2. Pareto analysis is insightful because it reveals the fundamental limits of the system.
- *Weighted-sum optimization:* In practice, a common way to produce Pareto-optimal designs is to solve a weighted sum objective:

$$\max_{\mathbf{x}} \quad \lambda U_{\text{comm}}(\mathbf{x}) + (1 - \lambda) U_{\text{sense}}(\mathbf{x}) , \quad (4.2)$$

where  $U_{\text{comm}}$  is the utility for communication,  $U_{\text{sense}}$  is the utility for sensing, and  $0 \leq \lambda \leq 1$  is a weight parameter. By varying  $\lambda$ , one can trace out different trade-off points. The advantage of the weighted-sum method is that it yields a single objective function that can often be tackled with standard optimization techniques, e.g., convex optimization or gradient-based methods.

- *Constrained optimization:* Another approach is to turn one objective into a constraint and optimize the other. For instance:

$$\begin{aligned} \max_{\mathbf{x}} \quad & U_{\text{sense}}(\mathbf{x}) \\ \text{subject to} \quad & U_{\text{comm}}(\mathbf{x}) \geq R_0 , \end{aligned} \quad (4.3)$$

where  $R_0$  is a minimum required communication rate. This reflects a scenario where communications have a strict QoS requirement, and within that feasibility set we want to maximize sensing performance. Many ISAC resource allocation problems are naturally cast this way. For example, Behdad et al. [18] set a minimal SINR target for every UE and then maximize the detection probability. We will review this concept in more detail in Chapter 5, where we will

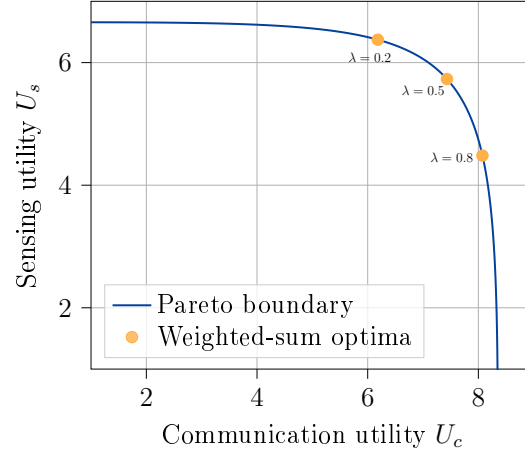


Figure 4.2: Pareto boundary of the communication-sensing tradeoff for a simple Multiple-Input Single-Output (MISO) ISAC model. Communication utility  $U_c(\mathbf{w})$  and sensing utility  $U_s(\mathbf{w})$  are traded off for varying  $\lambda \in [0, 1]$ . The plot shows the resulting Pareto boundary and three possible beamforming solutions for different  $\lambda$  values.

formulate our own CF-ISAC design problem as a constrained optimization that maximizes sensing performance subject to communication SINR constraints.

## 4.6 CELL-FREE ISAC SYSTEM MODELS

We now turn to the specific case of Cell-Free Massive MIMO networks that integrate sensing functionality, abbreviated here as CF-ISAC. In the context of ISAC, two sets of APs may be defined at any given time: a subset acting as transmitters, illuminating targets and serving downlink UEs, and a subset acting as receivers, listening for uplink signals and target echoes. The corresponding multi-static architecture and representative interference paths are summarized in Figure 4.3. All APs share CSI via the central processor, which means the network has knowledge of the channels from each AP to each UE and, importantly, any sensing-related channels. In many models, a target is treated as an object with a certain RCS that reflects the transmitted signals. If the target's location is known, e. g., within a “hotspot” area of interest, one can assume that the mean path to that location is known, but the small-scale fading and reflection coefficients may be unknown and need estimation. Some other

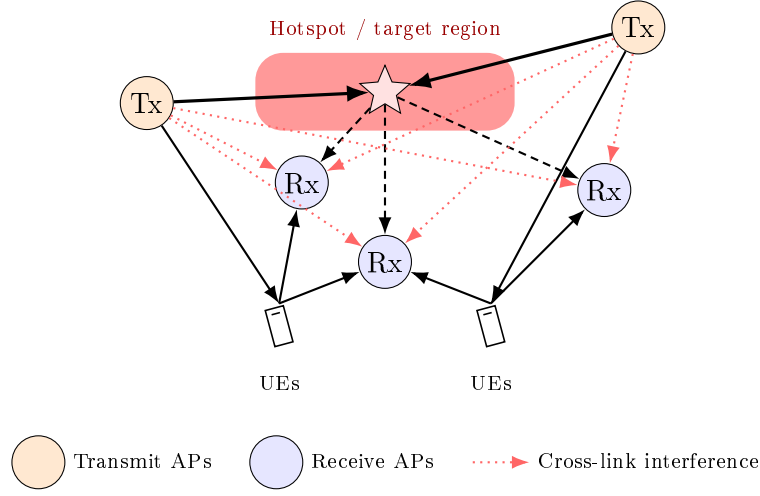


Figure 4.3: CF ISAC multi-static architecture. A subset of distributed APs transmits DL multi-user signals and optional sensing illumination towards a hotspot, while another subset receives UL transmissions and target echoes. Dotted arrows illustrate representative CLI paths during simultaneous transmit/receive operation.

studies assume a grid of possible target locations, and try to estimate the reflection at each grid point.

The DL communication signals to UEs contribute to the sensing process, but the UL communication signals from UEs can act as interference to sensing if APs are listening for targets at the same time. Therefore, models often include self-interference: for example, interference from a target-reflecting signal into a UE's receiver, or vice versa. In a Cell-Free setup, interference among APs is also considered. Chowdhury et al. [15] define CLI in two forms:

- *InAI*: when a transmitting AP's signal leaks into a nearby receiving AP.
- *InUI*: when a transmitting AP serving a DL UE interferes with an UL UE's transmission.

#### 4.6.1 Multi-static target detection

A major benefit of Cell-Free networks for sensing is the *multi-static* radar configuration. This spatial diversity can drastically improve detection robustness and

coverage. However, coordinating multiple transmit and receive APs for sensing requires careful signal design and processing. Two key research problems arise:

- How to perform target detection using the distributed measurements.
- How to allocate transmit power and optimize beamformers across APs to maximize target detection while satisfying communication QoS requirements.

For target detection, a common strategy is the GLRT, which is suitable when certain channel or target parameters are unknown. Chowdhury et al. [15] derive both centralized and distributed GLRTs for detecting a target in a CF-ISAC setup with DTDD. The centralized GLRT assumes all AP signals are brought to a CPU which performs a joint hypothesis test, while the distributed GLRT allows each receiving AP to process its local received signal and forward only a compressed summary to the CPU. The outputs are then fused for a global decision. As noted earlier, the complexity of a fully centralized detection grows rapidly with the number of APs and antennas. Thus, research is directed at finding clever ways to do near-optimal detection in a distributed fashion.

Regarding power allocation, a key challenge in multi-static sensing is to decide which APs should actively transmit towards the target and which should remain silent or only serve UEs. If too many APs transmit a strong sensing signal, it could interfere at UEs or create a lot of InAI. Some works, like Cheng et al. [47], have posed problems of minimizing total transmit power subject to a certain sensing SINR and communication SINR constraints. This yields a sparse solution where only a subset of APs, perhaps those with good geometry to the target, are actively emitting sensing waveforms. Such an approach can enhance energy efficiency and reduce interference, a concept similar to user-centric Cell-Free processing where only a cluster of APs handle a given UE.

Coordinated beamforming also plays a crucial role. Demirhan and Alkhateeb [19] formulate a joint beamforming design for communication and sensing, with an objective of max-min fairness in UE SINR while also achieving a target sensing

performance. They underscore that careful beamforming can reduce the inherent conflict between the two objectives.



## 4.7 DUPLEXING OPTIONS IN CF-ISAC

The duplexing strategy in CF-ISAC has a profound impact on performance, since it determines whether the network can simultaneously transmit to UEs/illuminate targets and listen for echoes. Two main options are considered: conventional TDD, and DTDD, which leverages network-level coordination to achieve simultaneous transmission and reception. The contrast between these two modes, in terms of temporal resource usage and interference exposure, is illustrated in Figure 4.4.

### 4.7.1 TDD Baseline

In a traditional TDD system, the network alternates between DL and UL in a synchronized manner for all APs. Under this mode, a straightforward way to perform sensing is during the DL phase: the APs transmit a known sensing waveform or their communication signals, then the echo from the target will arrive during the subsequent UL phase and can be picked up by APs at that time. Some works have indeed treated the UL phase as a “sensing reception” phase in which no UL UE signals are sent so that any backscattered signals can be collected. This avoids interference between communication and sensing but at the cost of dedicating UL slots to sensing-only, which reduces communication capacity.

### 4.7.2 Dynamic Time Division Duplex (DTDD)

In DTDD, different APs are allowed to operate in UL or DL mode simultaneously on the same time-frequency resources. Essentially, the network is partitioned into a transmit set and a receive set concurrently, rather than all doing one function. The partition can be adapted dynamically based on traffic demand and sensing needs.

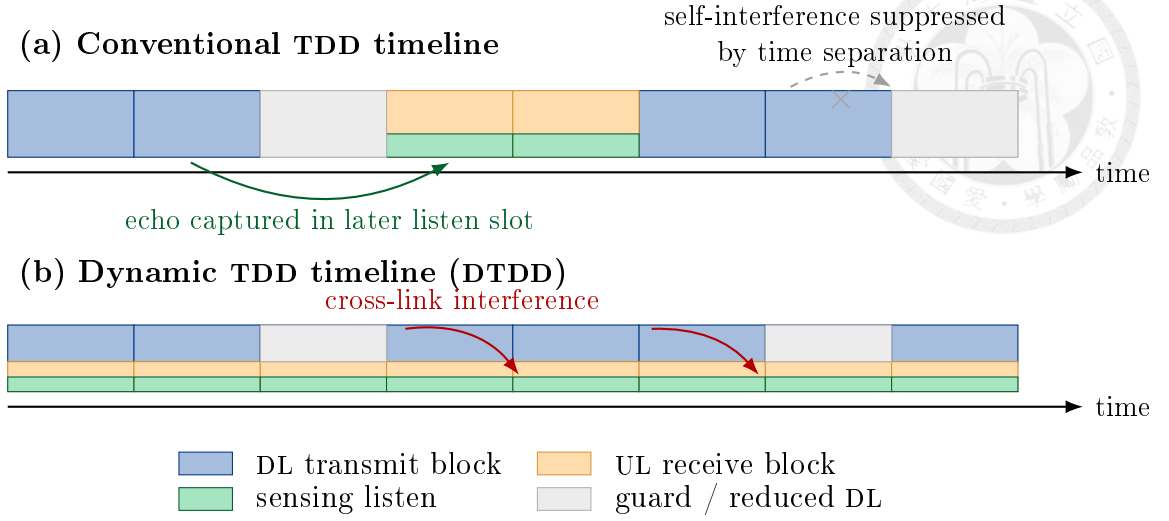


Figure 4.4: Duplexing timeline comparison in CF-ISAC. In conventional TDD, transmit and receive phases are separated in time, so same-AP self-interference is largely avoided and sensing echoes are captured in later receive slots. In Dynamic TDD, DL-transmitting and UL-receiving APs coexist in the same slots, enabling simultaneous communication and sensing but introducing CLI.

Chowdhury et al. [15] demonstrated several benefits of DTDD in CF-ISAC. First, it enables truly concurrent target sensing and UL communication. Second, since APs are half-duplex, DTDD avoids the self-interference problem that would plague a single node trying full duplex. Third, DTDD retains the channel reciprocity benefits for those APs which do UL and then DL for a given UE, as long as the channel coherence is maintained.

However, DTDD introduces new interference that must be managed: the aforementioned CLI. These cross-links can degrade both communication and sensing: for instance, a strong DL signal from a close-by AP could mask a weak target echo at a receiving AP. To mitigate CLI, various techniques are proposed: clever AP scheduling, power control, estimating the cross-link channels to enable interference-aware combining, or even smart beamforming design. In fact, one could envision designing the DL precoders not only to null interference at UEs but also to minimize leakage toward known receiving AP locations. Chowdhury et al. [15] found that with appropriate measures, DTDD can double the 90%-likely sum UL-DL SE compared to traditional TDD in a CF-MIMO setup, while still using simple half-duplex radios.

## 4.8 SCALABILITY GAPS IN EXISTING WORK

While the research so far demonstrates the potential of CF-ISAC in terms of raw performance, there are notable gaps when it comes to scalability. Scalability refers to the ability to handle a growing number of APs, UEs, and possibly targets, without an explosion in complexity, overhead, or degraded performance. Several limitations in current literature point to the need for more scalable solutions:

- *Centralized processing bottleneck:* Most CF-ISAC studies assume a fully centralized processing model, where a CPU has access to all measurements from all APs. This is theoretically optimal but raises concerns in a practical deployment with perhaps hundreds of APs.
- *Computational complexity:* The computational complexity at the CPU or APs themselves can become prohibitive. Similarly, joint beamforming optimization problems typically have complexity that grows polynomially with the number of APs and UEs.
- *Multiple target scalability:* Many existing works consider a single target or single radar task at a time. Scaling to multiple targets might require multiple sensing beams or more sophisticated waveform design, which has not been deeply studied in Cell-Free contexts.

Another scalability issue lies in channel estimation for sensing. Some works assume the target location or channel is known *a priori* to focus on detection of its reflection coefficient. Research that begins to bridge this often simplifies by treating the target return as one additional “UE” with unknown channel. Fully scalable approaches to sense a large environment with many resolution cells using a Cell-Free network remain an open field.

A further gap is how to reduce complexity via decentralization or clustering. One promising idea is to partition the network into clusters of APs that handle sensing and communication locally, similar to user-centric CF approaches. However, cluster-based ISAC design introduces questions on how to choose the clusters and

how to adapt them to moving targets or changing traffic. Finally, from an algorithmic complexity perspective, solving the joint design problems (beamforming, power allocation, scheduling) optimally might be nondeterministic-polynomial-time (NP)-hard when scaled up. Chowdhury et al. [15] indeed proposed a low-complexity AP scheduling algorithm based on local traffic and target SNR considerations, instead of exhaustive search which would be prohibitive for large-scale deployments. This kind of approach sacrifices some optimality but makes the problem tractable.





## Part II

### PROPOSED SCALABLE CELL-FREE ISAC ARCHITECTURE

In the second part of the thesis, we develop the original contribution of this work: a communication-constrained, scalable Cell-Free ISAC architecture for multi-static drone surveillance. Starting from the problem formulation and design philosophy, we define a protocol in which DTDD AP scheduling, scan control, hybrid down-link signaling, and uplink coherent detection are integrated into a unified execution framework. This part concludes with performance evaluation, discussion of the main sensing-communication trade-offs, and the final conclusions of the thesis.





# 5 | PROBLEM FORMULATION AND DESIGN PHILOSOPHY

The previous chapters have established the technological ingredients that make CF MMIMO ISAC a credible candidate for future wireless networks. In the specific context of drone surveillance, these ingredients become particularly compelling. A low-altitude target must be detected over a three-dimensional search volume while the same network continues to provide communication service to conventional UEs. Recent works have shown both the promise of coordinated volumetric sensing for drone detection and the benefits of combining DTDD with CF MMIMO ISAC [15, 22].

This chapter first introduces the application scenario and the operating assumptions adopted throughout the remainder of the thesis. It then formulates the communication-constrained sensing problem that guides the design, states the scalability criterion used to evaluate whether a protocol is compatible with large CF deployments, and explains why the proposed solution is organized as a sequence of bounded subproblems rather than as a monolithic online optimization.

## 5.1 MOTIVATION

Unlike a conventional point-to-point communication task, surveillance requires the network to reason over a search region rather than over a single receiver location. In practice, the target may appear anywhere inside a three-dimensional volume, may occupy unfavorable coverage regions, and may need to be revisited over time to maintain track continuity. These characteristics motivate a scan-based sensing architecture in which the area of interest is partitioned into elementary spatial cells and progressively explored, refined, and updated [22, 23].

A CF architecture is attractive for this task because it naturally provides multi-static sensing and, at the same time, remains well suited for communication, since CF MMIMO is specifically designed to provide more uniform service quality than cellular architectures through macro-diversity and cooperation among many simple APs [11, 16].

The setting considered in this thesis is more demanding than a static sensing problem. The network must sense and communicate simultaneously in a DTDD mode, which is appealing because it allows the network to use DL APs as illuminators and UL APs as sensing receivers without requiring full-duplex radio hardware at every node. However, it also introduces new coupling mechanisms, such as CLI, InAI, and InUI between opposite transmission directions [15].

Of these coupling mechanisms, InAI, which is the energy leakage from actively transmitting DL APs into simultaneously receiving UL APs, occupies a structurally distinct role in the proposed solution. A straightforward response would be to model the resulting leakage as additive noise at the UL receiver and absorb it into the interference covariance. The framework proposed here goes further: it actively suppresses the InAI on both sides of the coupling by shaping the DL payload beams to reduce their energy leakage toward the physically nearest vulnerable UL neighbors, and by introducing a complementary directional penalty into the local UL receive covariance to suppress the dominant residual arrival directions. The interference power injected into UL receivers is thereby reduced before it enters any receive chain, rather than being accepted as given and then processed away. However, this dual-sided mitigation reduces the useful beamforming or combining gain available for communication, requires knowledge of the inter-AP interference geometry and residual coupling statistics, demands storage of spatial direction descriptors and coupling weights at each AP, and adds a small computational overhead. The mechanism, its cost, and the operating conditions under which it is most beneficial are analyzed in detail in Section 7.2 for the transmit side and in Section 8.2 for the receive side, while numerical evidence is reported in Section 9.4.

The central question of the thesis is therefore how sensing can be added in a way that remains communication-feasible, structurally scalable, and operationally explicit. The answer developed in the following chapters is grounded in a *best-effort sensing* philosophy: communication constraints are enforced unconditionally in every slot, and sensing exploits only the resources that remain. The class of UAV monitoring scenarios for which this approach is appropriate is characterized in Section 1.3.2.

## 5.2 SYSTEM SETTING

The contribution part of the thesis considers a user-centric CF MMIMO network composed of  $M$  distributed APs, a CPU, a set  $\mathcal{U}_d[t]$  of active DL UEs, and a set  $\mathcal{U}_u[t]$  of active UL UEs in slot  $t$ . Each AP is equipped with  $N$  antennas and operates in half duplex at the hardware level, but the network as a whole supports DTDD by partitioning the APs into a DL set  $\mathcal{A}_d[t]$  and an UL set  $\mathcal{A}_u[t]$  at each slot, with

$$\mathcal{A}_d[t] \cup \mathcal{A}_u[t] = \mathcal{A}, \quad \mathcal{A}_d[t] \cap \mathcal{A}_u[t] = \emptyset, \quad (5.1)$$

where  $\mathcal{A} \triangleq \{1, \dots, M\}$  denotes the full AP index set. The distributed APs are coordinated by the CPU, which distributes the slot-level decisions required by the protocol.

The sensing task is defined over a voxelized surveillance region. Let  $\mathcal{V}[t]$  denote the set of voxels induced by a Cartesian grid descriptor, and let  $\mathcal{V}_{\text{scan}}[t] \subseteq \mathcal{V}[t]$  denote the subset requested for scan in slot  $t$ . The protocol deliberately distinguishes between requested and informative voxels. A voxel may be requested by the CPU because it belongs to the current search frontier or because it must be revisited for track maintenance, but it is counted as informative only if the network configuration of that slot actually provides the minimum support required for illumination and observation. The use of this distinction is consistent with the radar view that requesting a cell for observation and obtaining an actually informative observation

are not equivalent events, especially in scan-and-track procedures constrained by geometry and resource allocation [23, 24].

Communication and sensing are integrated in a hybrid downlink architecture. The term hybrid is used here in an operational sense: the network combines payload-carrying communication beams with deterministic sensing components and, when possible, with additional dedicated sensing beams. For each active DL AP  $j$ , the local transmission resources are organized as follows.

- *Communication layer*: AP  $j$  serves its DL UEs set  $\mathcal{D}_j[t]$  through adaptive protected payload precoders. These streams are the primary objects of the downlink, since the protocol is communication-constrained and must preserve the required QoS of the scheduled UEs.
- *Voxel-sequence and candidate-distribution layer*: the CPU assigns each requested voxel a deterministic sequence index and distributes a bounded candidate set to the active DL APs.
- *Payload-attached sensing layer*: once a payload stream  $(j, n)$  is available, it may also carry the deterministic sequence of one requested voxel. In this way, one payload stream can simultaneously support communication and contribute illumination to a voxel. The probing fraction is controlled by a scalar coefficient and is activated only if the corresponding communication-feasibility condition remains satisfied.
- *Dedicated virtual sensing layer*: the same AP may use residual spatial and power resources to illuminate additional candidates that were not matched to payload streams. These beams are purely deterministic sensing branches and are admitted only if the communication constraints are still respected.

The CPU forms binary candidate-distribution variables

$$z_{jv}[t] \in \{0, 1\}, \quad j \in \mathcal{A}_d[t], \quad v \in \mathcal{V}_{\text{scan}}[t], \quad (5.2)$$

where  $z_{jv}[t] = 1$  means that voxel  $v$  is sent to DL AP  $j$  for local hybrid/dedicated processing. The local candidate set is

$$\mathcal{V}_j^{\text{CPU}}[t] \triangleq \{v \in \mathcal{V}_{\text{scan}}[t] : z_{jv}[t] = 1\}. \quad (5.3)$$

The final illumination of a voxel is then the aggregate of all admitted local contributions,

$$W_v[t] = \sum_{j \in \mathcal{A}_d[t]} W_{jv}[t]. \quad (5.4)$$

On the UL side, the receiving APs serve a dual role: they decode UE payloads and collect echo-bearing observations for target detection. For sensing, the CPU assigns each UL AP a bounded list of voxel hypotheses

$$\mathcal{V}_m^{\text{sens}}[t], \quad |\mathcal{V}_m^{\text{sens}}[t]| \leq B_u^{\text{max}}, \quad m \in \mathcal{A}_u[t], \quad (5.5)$$

where  $B_u^{\text{max}}$  is a fixed receiver-side post-processing cap.

The UL path supports two complementary processing modes. In the *distributed implementation*, each UL AP applies a local combiner to detect the communication payloads and one voxel-dependent receive combiner for each  $v \in \mathcal{V}_m^{\text{sens}}[t]$  to the same baseband block, then forwards the resulting low-dimensional communication soft symbols and scalar sensing evidence to the CPU. This mode preserves user-centric processing and remains scalable with the number of APs and UEs, in line with CF principles [12, 17]. In the *centralized implementation*, active UL APs upload raw antenna-domain blocks to the CPU, which performs joint communication combining and centralized detection via GLRT.

### 5.3 CF-ISAC SCALABILITY CRITERION

Scalability is a first-class design requirement of the proposed framework. In classical user-centric CF MMIMO, the key principle is that the online processing burden of each AP and the corresponding fronthaul exchange must remain bounded as the number of UEs grows [12, 17]. In the present setting, this idea must be extended because the network handles not only UEs but also a voxelized sensing task. A protocol may be

scalable from a communication viewpoint and still become non-scalable once dense UE-voxel or AP-voxel couplings are introduced.

Following the foundational scalability principles defined by Bjornson and Sanguinetti [12], the thesis adopts the following CF-ISAC scalability criterion.

**Definition 5.1.** *A CF-ISAC protocol is scalable if it satisfies the following conditions:*

1. *Per-AP online computation and fronthaul must remain bounded as the numbers of UEs and candidate voxels grow.*
2. *Per-requested-voxel online work must remain bounded, so that each voxel is handled through short local lists rather than through all UEs or all APs.*
3. *The fast-timescale protocol must avoid dense all-to-all online laws such as*

$$\mathcal{O}(KV), \quad \mathcal{O}(MV), \quad \mathcal{O}(MK), \quad (5.6)$$

*where  $K$  is the number of active UEs,  $V$  is the number of candidate voxels, and  $M$  is the number of APs.*

This criterion extends the spirit of scalable CF processing from communication-only systems to the hybrid communication-sensing protocol considered here. Slow-timescale updates of the grid descriptor, pilot tables, protection lists, and background nuisance descriptors are allowed, provided that they are not recomputed through dense all-to-all operations in every slot. The design choices adopted in the following chapters are all evaluated against this criterion.

## 5.4 COMMUNICATION-CONSTRAINED SENSING PROBLEM

The proposed framework is designed according to a communication-constrained sensing philosophy. Communication feasibility is imposed first through bounded user-centric service, pilot-aware AP capping, and protected DL design; sensing is then embedded into the remaining feasible spatial and power DoF. This viewpoint is

compatible with the broader ISAC literature, where sensing performance is commonly optimized subject to communication-quality constraints [10, 15, 18], but it is specialized here to the structural constraints of CF MMIMO, voxelized surveillance, and DTDD operation.

A suitable guiding objective for slot  $t$  is therefore

$$\max_{\substack{\mathcal{A}_u[t], \mathcal{A}_d[t], \mathcal{V}_{\text{scan}}[t], \{\mathcal{V}_m^{\text{sens}}[t]\}, \\ \{P_{d,jn}^{(c)}[t]\}, \{\rho_{jn}[t]\}, \{\beta_j^{\text{act}}[t]\}}} \sum_{v \in \mathcal{V}_{\text{scan}}[t]} I[v, t] P_{D,v}[t], \quad (5.7)$$

subject to

$$\text{SINR}_n^{\text{DL}}[t] \geq \Gamma_n^{\min}, \quad \forall n \in \mathcal{U}_d[t], \quad (5.8)$$

$$\text{SINR}_k^{\text{UL}}[t] \geq \Gamma_k^{\min}, \quad \forall k \in \mathcal{U}_u[t], \quad (5.9)$$

$$\sum_{n \in \mathcal{D}_j[t]} P_{d,jn}^{(c)}[t] + \sum_{q=1}^{Q_j^{(s)}[t]} P_{j,q}^{(s)}[t] \leq P_j^{\max}, \quad \forall j \in \mathcal{A}_d[t], \quad (5.10)$$

$$|\mathcal{V}_j^{\text{CPU}}[t]| \leq N, \quad \forall j \in \mathcal{A}_d[t], \quad (5.11)$$

$$|\mathcal{V}_m^{\text{sens}}[t]| \leq B_u^{\max}, \quad \forall m \in \mathcal{A}_u[t], \quad (5.12)$$

where  $\text{SINR}_n^{\text{DL}}[t]$  and  $\text{SINR}_k^{\text{UL}}[t]$  denote the effective DL and UL SINRs, respectively. Moreover,  $\mathcal{D}_j[t]$  is the final capped set of DL payload streams emitted by AP  $j$ ,  $\mathcal{V}_j^{\text{CPU}}[t]$  is the requested-voxel candidate set sent by the CPU to AP  $j$ ,  $\mathcal{V}_m^{\text{sens}}[t]$  is the voxel-processing list sent by the CPU to UL AP  $m$ ,  $P_{d,jn}^{(c)}[t]$  is the communication power allocated by AP  $j$  to UE stream  $n$ ,  $Q_j^{(s)}[t]$  is the number of local dedicated sensing beams emitted by AP  $j$  and  $P_{j,q}^{(s)}[t]$  is the power of the  $q$ -th such beam.

#### 5.4.1 Best-effort sensing

The absence of a lower bound on the sensing objective in (5.7)–(5.12) is the formal expression of the best-effort sensing design introduced in Section 1.3.2: the network maximizes the achievable detection probability using whatever spatial, power, and processing resources remain after the communication constraints are satisfied, but it makes no commitment about the minimum level of sensing that will be delivered.

The scan-and-track architecture of Chapter 6 ensures that voxels that receive no illumination in slot  $t$  remain in the search frontier and are re-queued in subsequent slots according to the revisit-budget logic of Section 6.2. Persistent surveillance is therefore achieved by maximizing  $P_{D,v}[t]$  in each slot where resources are available, and relying on the revisit scheduler to ensure that every region of interest is served over time.

Adding a hard sensing floor to (5.7)–(5.12) would prevent the network from satisfying its communication obligations whenever the sensing floor and the communication constraints compete for the same resources. The complementary problem of minimizing transmit energy subject to both communication constraints and a minimum sensing guarantee is a natural extension that is discussed in Section 10.3.

A further consequence of the best-effort formulation concerns how the resulting sensing performance should be benchmarked. The sensing task considered here does not admit a canonical reference. The performance of the sensing task is the per-voxel detection probability  $P_{D,v}[t]$  accumulated over a time-varying scan-and-track process, and the resources available for sensing are not a fixed budget but the residual left after the communication constraints are satisfied in each slot. A fundamental sensing bound would therefore be conditioned on the realized communication load and network geometry, so that the appropriate notion of optimality is a sensing–communication trade-off region rather than a single scalar limit. For this reason, the detection results reported in Chapter 9 are compared against operational references, such as the centralized receiver as an upper benchmark, and the opportunistic and false-alarm-floor baselines as lower references, rather than against a protocol-independent fundamental limit. The derivation of such a limit for best-effort multi-static CF-ISAC is discussed as a future direction in Section 10.3.

#### 5.4.2 *Solving philosophy and procedure*

The role of (5.7) is conceptual. It states what the network should achieve, not how the network should solve the problem online. In fact, a direct solution of (5.7) would

mix binary AP-mode decisions, graph-assignment variables, nonlinear beamforming variables, communication metrics, and statistical detection constraints in one slot-level program. This coupled design would become difficult to solve reliably in real time, especially in a distributed architecture with fronthaul limitations and large candidate voxel sets. The contribution of the thesis is therefore the claim that the objective can be realized through a carefully structured procedure whose constituent steps remain bounded, explicit, and compatible with the information that is actually available at the APs and at the CPU.

The master problem in (5.7) is used as a design compass and is then converted into a slot-level protocol with three guiding principles. The first principle is *feasibility before sensing utility*: the network first constructs a communication-feasible DTDD configuration and only afterwards spends the remaining spatial, power, and processing resources on sensing. The second principle is *locality before global optimality*: every online decision must be made on bounded UE clusters, short AP neighborhoods, or bounded voxel lists, even if this means sacrificing the globally optimal solution of the conceptual problem. The third principle is *operational observability*: every detector used later in the thesis must depend only on quantities that the corresponding processor can know or receive through the protocol, while unknown target strength, bistatic propagation uncertainty, and residual modeling errors are absorbed into unknown complex amplitudes in the detection model.

This philosophy leads to a sequential solving procedure, and the framework is therefore organized as a sequence of bounded subproblems. Figure 5.1 illustrates the protocol as a staged slot-level execution architecture. This decomposition makes the guiding objective in (5.7) implementable in a real CF network. Each block is designed to be solved by a short local search, an exact scalar rule, a sparse graph operation, or a small local linear-algebra step [26, 48, 49].

The decomposition adopted in the thesis consists of the following stages.

- *Control and feasibility*. The scheduling layer of Chapter 6 schedules the APs in UL or DL mode and maintains scan bookkeeping, producing  $\mathcal{A}_d[t]$ ,  $\mathcal{A}_u[t]$ ,

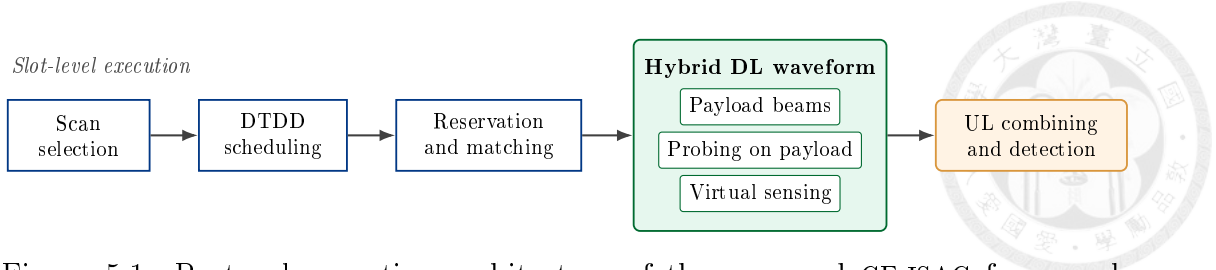


Figure 5.1: Protocol execution architecture of the proposed CF-ISAC framework. The figure shows the slot-level execution flow: requested-voxel handling, DTDD mode selection, bounded reservation and matching, hybrid DL waveform construction, and UL communication-plus-detection processing. The three internal components of the DL block emphasize the hybrid transmit structure based on payload communication beams, superimposed probing, and virtual sensing beams.

$\mathcal{V}_{\text{scan}}[t]$ , and  $\mathcal{V}_{\text{inf}}[t]$  through traffic-aware duplexing, feasibility correction, and track maintenance. This stage corresponds to the binary and set-valued variables of (5.7), but it handles them through the schedulers of Sections 6.2–6.4 rather than by exhaustive search.

- *Bounded candidate exposure.* The same control layer selects locally served UEs, assigns sequence indices to requested voxels via anchor-cell reuse, distributes them to DL APs with candidate sets  $\mathcal{V}_j^{\text{CPU}}[t]$ , and assigns UL voxel-processing lists  $\mathcal{V}_m^{\text{sens}}[t]$ . This realizes the scalability constraints (5.11)–(5.12) through the candidate-distribution and UL assignment rules of Sections 6.6 and 6.7.
- *Communication baseline and payload-attached probing.* The waveform-design stage of Chapter 7 shapes the DL signal through adaptive protected payload precoding and hybrid matching. This fixes the payload beams, payload-voxel associations, and minimum probing fractions  $\rho_{jn}[t]$ , thereby translating the DL communication constraint (5.8) into the explicit protected-precoding and probing-feasibility rules of Sections 7.2 and 7.3.
- *Dedicated sensing from residual resources.* The complementary sensing stage evaluates remaining local candidates at each DL AP, admits dedicated sensing beams, and redistributes leftover headroom while keeping beams in the complement of the served-UE subspace. This implements the power constraint (5.10) and the residual-sensing part of the objective through Section 7.4, while

the effective scalar reports of Section 7.4.1 tell the CPU which voxels were actually illuminated.

- *Receiver processing and detection.* The final stage performs coherent voxel-wise detection on UL observations while preserving UL communication. The distributed path uses local accumulation and CPU-side scalar fusion with CFAR thresholds, while the centralized raw-uplink variant moves receive processing to the CPU. These alternatives are developed in Sections 8.2–8.5.1 and Section 8.6; they provide the protocol-level realization of  $P_{D,v}[t]$  in the objective and of the UL communication constraint (5.9).

The remainder of the contribution part follows the same logic as the decomposition above. Chapter 6 presents the adaptive scheduling and control mechanisms that scale efficiently across the network, Chapter 7 addresses the transmission-side design and optimization, Chapter 8 combines UL signal reception with target detection capabilities, and Chapter 9 demonstrates the performance through simulations and analysis.



# 6

## SCHEDULING AND CONTROL ARCHITECTURE



This chapter turns the principles introduced in the previous chapter into an explicit control architecture. Its purpose is to explain how the network decides, at each slot, which APs operate in UL or DL mode and which voxels are requested for scan. This role of scheduling is particularly important in DTDD CF networks. Unlike a conventional static TDD baseline, a DTDD system must decide the direction of operation of each AP on the same time-frequency resource while coping with local traffic asymmetries and with the sensing requirement that a requested voxel should remain observable by at least one DL illuminator and one UL receiver in its neighborhood.

The chapter is organized as follows. Section 6.1 clarifies why scheduling is central in the proposed protocol. Section 6.2 defines requested and informative voxels together with the scan semantics adopted. Section 6.3 introduces the local neighborhoods that make the protocol scalable. Sections 6.4–6.6 describe slot-level decisions: DTDD AP scheduling, pilot coordination, deterministic sequences planning, and CPU-side candidate distribution. Section 6.7 concludes the control loop with UL sensing assignment, informative-bitmap updates, and the final slot outputs delivered to the physical-layer blocks. Finally, Section 6.8 summarizes the complexity and fronthaul implications of the entire control layer.

### 6.1 ROLE OF SCHEDULING IN THE PROPOSED PROTOCOL

In this framework, the network must first determine which entities are relevant for the current slot and which relations among them are feasible. This includes at least five coupled questions: which APs should transmit or receive, which voxels should be

requested for scan, which UEs should be retained in the local DL and UL service sets, which requested voxels should be exposed to each DL AP as local sensing candidates, and which UL APs should process each voxel in the sensing branch.

In scalable CF systems, the network first constructs bounded user-centric service sets and only then performs local precoding or combining [12, 17]. In DTDD CF-ISAC, the direction of operation of the APs must also reflect the local UL/DL traffic and the sensing geometry around the target region [15]. Figure 6.1 summarizes this logic as a two-level control chain. The upper band contains the slow-timescale descriptors reused across many slots. The lower band contains the slot-level steps that start from scan requests and end with the bounded control objects delivered to the DL and UL physical-layer processors.

## 6.2 SCAN SEMANTICS

At slot  $t$ , the CPU selects a set  $\mathcal{V}_{\text{scan}}[t]$  of voxels to be requested for scan. However, a requested voxel is not automatically informative. A voxel becomes informative only if the slot-level configuration effectively provides the minimum illumination and observation support required by the hybrid sensing stage. This distinction is encoded by the binary illumination flag

$$I[v, t] \in \{0, 1\}, \quad v \in \mathcal{V}_{\text{scan}}[t], \quad (6.1)$$

and by the informative scan set

$$\mathcal{V}_{\text{inf}}[t] \triangleq \{v \in \mathcal{V}_{\text{scan}}[t] : I[v, t] = 1\}. \quad (6.2)$$

Only voxels in  $\mathcal{V}_{\text{inf}}[t]$  are treated as truly scanned for revisit-time accounting, scan-pointer advancement, and track maintenance. The informative budget effectively delivered in slot  $t$  is therefore

$$Q_{\text{eff}}[t] \triangleq |\mathcal{V}_{\text{inf}}[t]| = \sum_{v \in \mathcal{V}_{\text{scan}}[t]} I[v, t]. \quad (6.3)$$

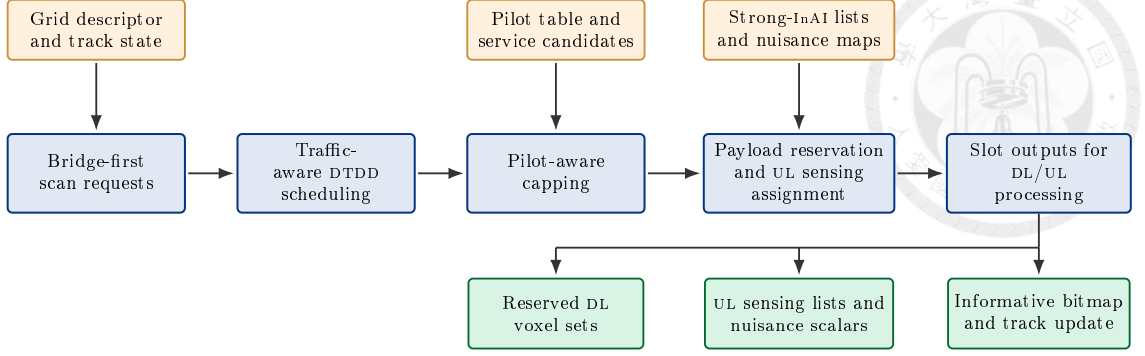


Figure 6.1: Control architecture of the scheduling layer. The upper band contains the slow-timescale control objects maintained by the CPU and reused across multiple slots. The lower chain shows the slot-level execution logic: bridge-aware scan requests, traffic-aware DTDD scheduling, pilot-aware service capping, sequence planning, coverage-balanced DL candidate distribution, UL sensing assignment, and the final control outputs consumed by DL and UL physical-layer processors.

The requested budget and the informative budget are thus distinct objects. The first is imposed by the scheduler, whereas the second is revealed only after the illumination bitmap is known.

### 6.2.1 Adaptive coarse/fine scan logic

The monitored region is modeled as a Cartesian box with bounds  $[x_{\min}, x_{\max}]$ ,  $[y_{\min}, y_{\max}]$ , and  $[z_{\min}, z_{\max}]$ . The scan controller maintains a coarse Cartesian grid over this box and attaches a local fine grid with refinement factor  $F$  to each coarse voxel. The coarse grid determines the long-term discovery and tracking cadence, whereas the fine children are used to localize a target after a coarse region has become informative. This coarse-to-fine logic is standard in scan-and-track radar architectures because it balances revisit time, coverage, and spatial resolution [23, 24]. The grid must also be consistent with the number of informative voxels that the DTDD CF-ISAC network can sustain in the presence of active tracks.

Let

$$W \triangleq \left\lceil \frac{T_{\text{rev}}^{\max}}{T_{\text{slot}}} \right\rceil \quad (6.4)$$

be the coarse revisit window expressed in slots, and let  $D_{\text{trk}}$  denote the dwell time, also in slots, after which an unrefreshed track is declared dead. The two quantities are tied in the control layer: the grid is not allowed to grow so large that the expected

coarse revisit period approaches the track-death horizon. This is enforced through the effective grid-sizing window

$$W_{\text{eff}} \triangleq \min\{W, \xi_g D_{\text{trk}}\}, \quad (6.5)$$

where  $\xi_g \in (0, \infty]$  is a protocol parameter. The value  $\xi_g = \infty$  removes the dwell constraint, whereas  $\xi_g < 1$  forces more than one scan opportunity per dwell window for marginal targets.

The CPU observes two slot-level quantities. The first is the informative budget  $Q_{\text{eff}}[t]$  in (6.3); The second is the number of requested voxels consumed by active tracks,  $Q_{\text{trk}}[t]$ . This quantity is a voxel count, since a single track can require several candidate voxels when its support spans fine children and adjacent coarse voxels, and this effect becomes especially important after grid refinement. Over the same slow window used for the informative budget, define

$$\bar{Q}_{\text{eff}}[t] \triangleq \frac{1}{W} \sum_{\tau=t-W+1}^t Q_{\text{eff}}[\tau], \quad \bar{Q}_{\text{trk}}[t] \triangleq \frac{1}{W} \sum_{\tau=t-W+1}^t Q_{\text{trk}}[\tau]. \quad (6.6)$$

Only the difference between these two averages is available for coarse discovery. With a safety factor  $\kappa_g \geq 1$ , the coarse-discovery budget is

$$B_c[t] \triangleq \left\lfloor \frac{\max\{0, \bar{Q}_{\text{eff}}[t] - \bar{Q}_{\text{trk}}[t]\} W_{\text{eff}}}{\kappa_g} \right\rfloor. \quad (6.7)$$

The factor  $\kappa_g$  reserves headroom for slot-to-slot variability and prevents premature refinement when the track population is already consuming a large part of the sensing budget.

Let

$$N_c[t] = N_x[t] N_y[t] N_z[t] \quad (6.8)$$

be the current number of coarse voxels. The grid is coarsened, kept, or refined according to the rule

$$\begin{cases} \text{coarsen,} & N_c[t] > B_c[t], \\ \text{stable,} & B_c[t]/2 < N_c[t] \leq B_c[t], \\ \text{refine,} & 2N_c[t] \leq B_c[t]. \end{cases} \quad (6.9)$$

Thus a grid initialized too coarsely refines upward into the band, while a grid initialized too finely coarsens downward into the same band.

A resize event changes only one axis. Let

$$\Delta_x[t] = \frac{x_{\max} - x_{\min}}{N_x[t]}, \quad \Delta_y[t] = \frac{y_{\max} - y_{\min}}{N_y[t]}, \quad \Delta_z[t] = \frac{z_{\max} - z_{\min}}{N_z[t]} \quad (6.10)$$

be the physical widths of the current coarse voxels. If (6.9) requests refinement, the controller doubles the least resolved axis,

$$a^+[t] = \arg \max_{a \in \{x,y,z\}} \Delta_a[t], \quad N_{a^+}[t] \leftarrow 2N_{a^+}[t]. \quad (6.11)$$

If coarsening is requested, it halves the most over-resolved axis among those with more than one cell,

$$a^-[t] = \arg \min_{a \in \{x,y,z\}: N_a[t] > 1} \Delta_a[t], \quad N_{a^-}[t] \leftarrow \left\lceil \frac{N_{a^-}[t]}{2} \right\rceil. \quad (6.12)$$

This one-axis rule keeps the voxel aspect ratio balanced: refinement is spent on the longest physical segment, while coarsening removes resolution from the shortest one.

Algorithm 6.1 summarizes the adaptive grid rule.

### 6.2.2 Track state, support-preserving remap, and bridge requests

The CPU keeps a coarse round-robin pointer for the global search and a set of active tracks. Each track  $\ell$  stores a current coarse owner voxel  $c_\ell$ , a fine-local state  $f_\ell$ , a cyclic ordinary scan list, an expiration time, and the physical support box of its last informative observation,

$$B_\ell = [x_\ell^-, x_\ell^+] \times [y_\ell^-, y_\ell^+] \times [z_\ell^-, z_\ell^+]. \quad (6.13)$$

The state  $f_\ell$  records the last informative fine voxel whenever such a localization exists. If the last informative refresh is only coarse, the protocol stores the coarse owner itself,  $f_\ell = c_\ell$ .

The ordinary track list is written as

$$\mathcal{L} \leftarrow [f_\ell, \mathcal{F}(c_\ell) \setminus \{f_\ell\}, \mathcal{N}_c^{\text{adj}}(f_\ell)], \quad (6.14)$$

**Algorithm 6.1:** Adaptive coarse/fine grid update

---

**Input:** current grid  $(N_x, N_y, N_z)$ ; recent histories  $Q_{\text{eff}}[\cdot]$  and  $Q_{\text{trk}}[\cdot]$ ;  $W, D_{\text{trk}}, \xi_g, \kappa_g$

**State:** last resize slot  $t_{\text{rsz}}$ ; active tracks and bridge flags

**Output:** updated grid descriptor  $\mathcal{G}[t]$

---

```

1  $W_{\text{eff}} \leftarrow \min\{W, \xi_g D_{\text{trk}}\};$ 
2 compute  $\bar{Q}_{\text{eff}}[t]$  and  $\bar{Q}_{\text{trk}}[t]$  from (6.6);
3 compute  $B_c[t]$  from (6.7) and  $N_c[t] \leftarrow N_x N_y N_z$ ;
4 if one or more tracks are in bridge state then
5   | close the active bridge states before resizing, extend their dwell timers
   | consistently, and rebuild ordinary scan lists;
6 end
7 if  $N_c[t] > B_c[t]$  then
8   | compute  $a^-[t]$  from (6.12) and halve only that axis;
9   |  $t_{\text{rsz}} \leftarrow t$ ;
10 else
11   | if  $2N_c[t] \leq B_c[t]$  then
12   |   | compute  $a^+[t]$  from (6.11) and double only that axis;
13   |   |  $t_{\text{rsz}} \leftarrow t$ ;
14   | else
15   |   | keep the current grid level;
16   | end
17 end
18 rebuild the coarse list  $\mathcal{L}_c$ , the coarse-neighborhood maps, and the fine-child maps
   using deterministic rules;

```

---

where  $\mathcal{F}(c_\ell)$  is the set of fine children of  $c_\ell$  and  $\mathcal{N}_c^{\text{adj}}(f_\ell)$  denotes the adjacent coarse voxels whose boundary physically touches the last confirmed fine voxel. If  $f_\ell$  is only a coarse identifier, the list is built from the coarse owner and its coarse neighbors, with the dwell timer inflated so that the longer uncertainty list does not immediately kill the track. The list is rebuilt whenever an informative detection updates the fine-local state or recenters the track to a new coarse owner.

Suppose now that a grid update changes the coarse grid. Because a resize modifies only one axis by a factor two, the previous support box intersects a small set of updated coarse voxels. The support-preserving coarse set is

$$\mathcal{S}_\ell^+ \triangleq \{u \in \mathcal{V}_c^+ : \text{vol}(B_c^+(u) \cap B_\ell) > 0\}, \quad (6.15)$$

where  $B_c^+(u)$  is the Cartesian box of coarse voxel  $u$  in the updated grid. If  $|\mathcal{S}_\ell^+| = 1$ , the track is immediately assigned to the unique compatible coarse voxel. If the track had a fine-local state, the corresponding Cartesian center is remapped together with

the coarse owner and retained only if it remains a valid fine child of the selected coarse voxel.

If  $|\mathcal{S}_\ell^+| > 1$ , the track enters a temporary bridge state. The bridge list is the support-preserving set sorted by proximity to the remapped fine-local position,

$$\mathcal{L}_\ell^{\text{br}} \leftarrow \text{sort}_{u \in \mathcal{S}_\ell^+} \|\mathbf{p}_u - \mathbf{p}_{\tilde{f}_\ell}\|^2, \quad (6.16)$$

where  $\tilde{f}_\ell$  is the remapped fine-local reference when available, and otherwise the remapped coarse center is used. The bridge state also stores a bridge pointer and a cycle counter  $C_\ell^{\text{br}}$ , initialized to zero. The bridge collapses as soon as one bridge voxel becomes informative and yields a valid detection. If the bridge list has been scanned once without finding the target, the track is declared dead; this prevents unresolved bridge supports from permanently consuming scan budget.

### 6.2.3 Bridge-capped requested-voxel scheduler

With  $\mathcal{D}_j[t]$  denoting the final capped DL UE set of AP  $j$ , the request budget is

$$Q_{\text{req}}[t] \triangleq \sum_{j \in \mathcal{A}_d[t]} |\mathcal{D}_j[t]|. \quad (6.17)$$

This rule bounds how many scan requests the control layer injects into the slot. Bridge requests have priority because they resolve the ambiguity created by a grid resize, but they are not allowed to exhaust the whole request budget. Given a bridge-budget fraction  $\phi_{\text{br}} \in [0, 1]$ , define

$$Q_{\text{br}}^{\text{max}}[t] \triangleq \lfloor \phi_{\text{br}} Q_{\text{req}}[t] \rfloor. \quad (6.18)$$

Only the first  $Q_{\text{br}}^{\text{max}}[t]$  requests of the slot can be spent on bridge tracks. The remaining budget is reserved for ordinary track maintenance and coarse round-robin discovery.

The online request scheduler therefore fills the budget in three passes. First, it serves bridge requests up to the cap in (6.18). Second, it serves ordinary track requests taken cyclically from the lists  $\mathcal{L}_\ell$  of the non-bridge tracks. Third, it serves global coarse requests from the stride-ordered coarse round-robin list. The ordinary

track pointers and the global coarse pointer are advanced only after the illumination bitmap is known and only for requests that became informative. The temporary bridge pointer is advanced when the bridge request is issued. Algorithm 6.2 summarizes the request rule.

### 6.3 UE AND VOXEL NEIGHBORHOODS

UEs are handled through bounded user-centric service sets, and voxels are handled through short neighborhood queries. The UE neighborhood describes local communication traffic around the APs, whereas the voxel neighborhood describes which APs can provide useful sensing exposure to it. The neighborhood radii are inherited from the DTDD scheduling rule of Chowdhury et al. [15]:

$$r_{\text{UE}} \triangleq \max \left\{ d_{\text{SNR}_0}^{(\text{UE})}, \max_{k \in \mathcal{U}} \min_{m \in \mathcal{A}} d_{mk} \right\}, \quad r_v \triangleq \max \left\{ d_{\text{SNR}_0}^{(v)} \right\}. \quad (6.19)$$

Here  $\text{SNR}_0$  is a protocol parameter, a fixed minimum SNR below which an AP is considered to provide no useful link to a given UE or voxel. The associated distances  $d_{\text{SNR}_0}^{(\text{UE})}$  and  $d_{\text{SNR}_0}^{(v)}$  are the *cutoff ranges* at which the large-scale received SNR decays exactly to  $\text{SNR}_0$ : beyond this distance an AP is excluded from the neighborhood. The first radius is used for the communication traffic neighborhoods and includes a closest-AP guard, namely the term  $\max_{k \in \mathcal{U}} \min_{m \in \mathcal{A}} d_{mk}$ , which is the largest UE-to-nearest-AP distance in the network; this ensures that even a UE whose closest AP lies beyond  $d_{\text{SNR}_0}^{(\text{UE})}$  still retains at least one local traffic reference. The second radius is used for voxel exposure and is based on a voxel-specific sensing threshold, so that only APs whose illumination of the voxel exceeds  $\text{SNR}_0$  are counted.

The traffic radius is used to quantify local UL/DL demand around each AP. For every AP  $m$ , define the UL and DL traffic neighborhood loads as

$$N_{m,u}[t] \triangleq \sum_{k \in \mathcal{U}_u[t]} \mathbf{1}\{m \in \mathcal{C}_k\}, \quad N_{m,d}[t] \triangleq \sum_{n \in \mathcal{U}_d[t]} \mathbf{1}\{m \in \mathcal{C}_n\}, \quad (6.20)$$

where  $\mathcal{C}_k$  denotes the bounded user-centric candidate cluster of UE  $k$ . These quantities drive the first stage of the DTDD scheduling rule.

**Algorithm 6.2:** Scan request scheduler

---

**Input:**  $\mathcal{A}_d[t]$ ,  $\{\mathcal{D}_j[t]\}$ ; coarse list  $\mathcal{L}_c$  and pointer  $p_c[t-1]$ ; ordinary track lists  $\{\mathcal{L}_\ell\}$  and pointers  $\{p_\ell\}$ ; bridge flags, bridge lists  $\{\mathcal{L}_\ell^{\text{br}}\}$ , bridge pointers  $\{q_\ell\}$ , and bridge counters  $\{C_\ell^{\text{br}}\}$

**Output:** requested scan set  $\mathcal{V}_{\text{scan}}[t]$  and track-request count  $Q_{\text{trk}}[t]$

---

- 1  $\mathcal{V}_{\text{scan}}[t] \leftarrow \emptyset$ ,  $Q_{\text{trk}}[t] \leftarrow 0$ ;
- 2 compute  $Q_{\text{req}}[t]$  from (6.17) and  $Q_{\text{br}}^{\text{max}}[t]$  from (6.18);
- 3 **Pass 1: capped bridge requests;**
- 4 while bridge tracks exist,  $Q_{\text{trk}}[t] < Q_{\text{br}}^{\text{max}}[t]$ , and  $|\mathcal{V}_{\text{scan}}[t]| < Q_{\text{req}}[t]$ , pick the next bridge track in round-robin, request the next voxel in  $\mathcal{L}_\ell^{\text{br}}$ , advance  $q_\ell$ , increment  $Q_{\text{trk}}[t]$ , and increase  $C_\ell^{\text{br}}$  if the pointer wraps;
- 5 **Pass 2: ordinary track requests;**
- 6 while non-bridge tracks exist and  $|\mathcal{V}_{\text{scan}}[t]| < Q_{\text{req}}[t]$ , pick the next ordinary track in round-robin, request the voxel indexed by its ordinary pointer  $p_\ell$ , and increment  $Q_{\text{trk}}[t]$ ;
- 7 **Pass 3: coarse fallback;**
- 8 while  $|\mathcal{V}_{\text{scan}}[t]| < Q_{\text{req}}[t]$ , request the next voxel in the stride-ordered coarse list  $\mathcal{L}_c$  starting from  $p_c[t-1]$ ;
- 9 remove duplicates if the same voxel is selected twice and return  $\mathcal{V}_{\text{scan}}[t]$ ;

---

Voxel handling follows the same bounded philosophy, but uses the voxel radius  $r_v$ . For any voxel  $v$ , the CPU reconstructs the voxel center  $\mathbf{p}_v$  from the current grid  $\mathcal{G}[t]$  and forms the query

$$v \mapsto (\mathcal{A}(v), i^{(1)}(v), \text{NN}_v), \quad (6.21)$$

where the set

$$\mathcal{A}(v) \triangleq \{m \in \mathcal{A} : d_{mv} \leq r_v\} \quad (6.22)$$

contains the APs that are close enough to voxel  $v$  to be considered eligible sensing neighbors. If this set is empty, the query falls back to the nearest AP,

$$\mathcal{A}(v) \leftarrow \{i^{(1)}(v)\}, \quad i^{(1)}(v) \triangleq \arg \min_{m \in \mathcal{A}} d_{mv}. \quad (6.23)$$

The ordered list

$$\text{NN}_v = [i^{(1)}(v), i^{(2)}(v), \dots, i^{(L_v)}(v)] \quad (6.24)$$

stores the  $L_v$  nearest APs to voxel  $v$ , with small list length  $L_v$ . The set  $\mathcal{A}(v)$  contains the eligible APs for voxel  $v$ .

## 6.4 DTDD AP SCHEDULING

The DTDD scheduler starts from local traffic and then corrects the result so that every requested voxel has at least one eligible UL AP and at least one eligible DL AP in its voxel neighborhood. This rule retains the traffic-based scheduling logic of Chowdhury et al. [15] and augments it with a requested-voxel feasibility correction adapted to the voxelized hybrid sensing architecture.

For each requested voxel  $v \in \mathcal{V}_{\text{scan}}[t]$ , define the neighborhood counts

$$N_{v,u}[t] \triangleq |\mathcal{A}_u[t] \cap \mathcal{A}(v)|, \quad N_{v,d}[t] \triangleq |\mathcal{A}_d[t] \cap \mathcal{A}(v)|. \quad (6.25)$$

The correction stage enforces

$$N_{v,u}[t] \geq 1, \quad N_{v,d}[t] \geq 1, \quad \forall v \in \mathcal{V}_{\text{scan}}[t]. \quad (6.26)$$

The scheduler proceeds in two conceptual stages. First, each AP is assigned to the direction with the larger local traffic load, using the comparison of  $N_{m,u}[t]$  and  $N_{m,d}[t]$ . Second, the requested voxels are scanned one by one. Whenever a voxel lacks a nearby UL AP, the nearest currently DL AP in  $\text{NN}_v$  is flipped to UL mode, and vice versa if the voxel lacks a nearby DL AP. Because AP neighborhoods overlap, a single repair may damage the feasibility of another voxel. For this reason, the correction pass is optionally repeated once. Algorithm 6.3 collects the resulting rule.

The computational cost of this stage is very low. The traffic decision costs  $\mathcal{O}(M)$  and the requested-voxel repairs give a slot-level cost proportional to the number of requested voxels and their short local AP lists.

## 6.5 PILOT COORDINATION

Pilot contamination is one of the main scalability bottlenecks in CF systems, as explained in Section 2.2, and one-UE-per-pilot capping is a simple yet effective way to control it. The proposed protocol implements a pilot-aware capping rule inspired by Bjornson and Sanguinetti [12].

**Algorithm 6.3:** Traffic-aware DTDD AP scheduling

---

**Input:** active UL users  $\mathcal{U}_u[t]$ , active DL users  $\mathcal{U}_d[t]$ , user-centric clusters  $\{\mathcal{C}_k\}$ , requested scan set  $\mathcal{V}_{\text{scan}}[t]$ , on-demand query  $v \mapsto (\mathcal{A}(v), \text{NN}_v)$

**Output:** scheduled UL and DL AP sets  $\mathcal{A}_u[t]$ ,  $\mathcal{A}_d[t]$

```

1 for  $m = 1$  to  $M$  do
2    $N_{m,u}[t] \leftarrow \sum_{k \in \mathcal{U}_u[t]} \mathbf{1}\{m \in \mathcal{C}_k\}$ ,  $N_{m,d}[t] \leftarrow \sum_{k \in \mathcal{U}_d[t]} \mathbf{1}\{m \in \mathcal{C}_k\}$ ;
3 end
4 initialize  $\mathcal{A}_u[t]$  and  $\mathcal{A}_d[t]$  by assigning each AP to the larger of its local UL/DL loads;
5 foreach  $v \in \mathcal{V}_{\text{scan}}[t]$  do
6   query  $\mathcal{A}(v)$  and  $\text{NN}_v$ ;
7   if  $|\mathcal{A}_u[t] \cap \mathcal{A}(v)| = 0$  then
8     flip the first DL AP in  $\text{NN}_v$  to UL mode;
9   end
10  if  $|\mathcal{A}_d[t] \cap \mathcal{A}(v)| = 0$  then
11    flip the first UL AP in  $\text{NN}_v$  to DL mode;
12  end
13 end
14 optionally repeat the voxel-repair pass once;
15 return  $\mathcal{A}_u[t]$  and  $\mathcal{A}_d[t]$ ;

```

---

Since a slot contains a common pilot phase for all active UEs, this yields the deterministic per-AP bounds

$$|\mathcal{D}_j[t]| \leq \tau_p, \quad |\mathcal{U}_m[t]| \leq \tau_p, \quad (6.27)$$

where  $\mathcal{D}_j[t]$  and  $\mathcal{U}_m[t]$  are the final capped DL and UL UE sets handled by DL AP  $j$  and UL AP  $m$ , respectively, and  $\tau_p$  is the number of orthogonal pilots.

The single pilot phase is reused by both communication directions. Let

$$\mathcal{U}_{\text{act}}[t] \triangleq \mathcal{U}_u[t] \cup \mathcal{U}_d[t] \quad (6.28)$$

be the set of all active UEs in slot  $t$ . Every UE  $k \in \mathcal{U}_{\text{act}}[t]$  transmits its assigned pilot  $\varphi_{\pi_k}$  during the common pilot phase. The same measurements are then reused by UL processors for local combining and by DL processors for reciprocity-based precoding.

Pilot assignment is executed only on the slow timescale. For UE  $k$ , define its master AP as

$$m_k^* \triangleq \arg \max_m \beta_{mk}, \quad (6.29)$$

where  $\beta_{mk}$  is the large-scale fading coefficient between AP  $m$  and UE  $k$ . The master AP is the most critical reference for estimation quality and is the natural vantage

point for pilot contamination evaluation [12]. The contamination score of pilot  $q$  measured at the master AP is

$$J_k(q) \triangleq \sum_{i: \pi_i=q} \beta_{m_k^* i}, \quad (6.30)$$

and the pilot is selected according to

$$\pi_k \leftarrow \arg \min_{q \in \{1, \dots, \tau_p\}} J_k(q). \quad (6.31)$$

Once pilots are fixed, the slot-level capping must preserve two properties at the same time. First, each active UE should be allowed to keep a user-centric coherent support made of several nearby APs. Second, no AP should serve two UEs that reuse the same pilot, because this would create pilot-contaminated channel estimates. The CPU therefore scans the candidate service lists of the active DL UEs over the currently scheduled DL APs, ordered by decreasing large-scale channel strength, and assigns each UE to every feasible AP whose local pilot table does not yet contain that UE's pilot. The same rule is applied independently on the UL side. This procedure produces the bounded local service sets  $\{\mathcal{D}_j[t]\}$  and  $\{\mathcal{U}_m[t]\}$ .

## 6.6 VOXEL-SEQUENCE PLANNING

The next control task is assigning a deterministic sequence index to every requested voxel and distributing requested voxels across the active DL APs, so that each AP receives only a small candidate set for local waveform admission.

### 6.6.1 Deterministic sensing sequence

Let

$$T \triangleq T_{\text{slot}} - \tau_p \quad (6.32)$$

be the number of payload samples available for deterministic probing. The protocol uses the length- $T$  DFT family

$$d_q[\tau] = \exp\left(-j \frac{2\pi}{T} (\tau - 1)(q - 1)\right), \quad q = 1, \dots, T, \quad (6.33)$$

which is pairwise orthogonal over the payload interval. The sequence-planning problem is therefore to assign an index  $s_v[t] \in \{1, \dots, T\}$  to each requested voxel  $v$ .

The assignment is built on the current coarse anchor lattice. Let

$$a(v) = (i_v^{(0)}, j_v^{(0)}, k_v^{(0)}) \quad (6.34)$$

be the coarse anchor cell containing requested voxel  $v$ . This definition applies both to unrefined coarse voxels and to fine voxels produced by local refinement inside that coarse cell. For every occupied anchor cell  $a$ , define

$$n_a[t] \triangleq |\{v \in \mathcal{V}_{\text{scan}}[t] : a(v) = a\}|, \quad n_{\max}[t] \triangleq \max_a n_a[t]. \quad (6.35)$$

The quantity  $n_{\max}[t]$  is the number of local sequence labels that must be reserved to separate the most crowded anchor cell. Hence exact sequence feasibility requires

$$n_{\max}[t] \leq T. \quad (6.36)$$

If this condition fails, a single anchor cell already contains more simultaneously requested voxels than the number of available orthogonal deterministic waveforms; the CPU therefore defers the lowest-priority excess requests in the overcrowded anchor cells until (6.36) holds.

Once the request set is sequence-feasible, the remaining sequence budget is converted into a periodic spatial reuse pattern. Define

$$B[t] \triangleq \left\lfloor \frac{T}{n_{\max}[t]} \right\rfloor, \quad (6.37)$$

and choose positive integers  $K_x[t]$ ,  $K_y[t]$ , and  $K_z[t]$  such that

$$K_x[t]K_y[t]K_z[t] \leq B[t]. \quad (6.38)$$

For an anchor grid spanning  $G_x$ ,  $G_y$ , and  $G_z$  coarse cells, a balanced choice is obtained from

$$\alpha[t] = \left( \frac{B[t]}{G_x G_y G_z} \right)^{1/3}, \quad K_x[t] = \max\{1, \lfloor \alpha[t] G_x \rfloor\}, \quad (6.39)$$

with analogous definitions for  $K_y[t]$  and  $K_z[t]$ , followed by decrements if needed to satisfy (6.38). For low-altitude surveillance, a simpler horizontal reuse rule can set  $K_z[t] = 1$  and choose only  $K_x[t]K_y[t] \leq B[t]$ .

The base hash class of requested voxel  $v$  is

$$b_v[t] \triangleq (i_v^{(0)} \bmod K_x[t]) + K_x[t](j_v^{(0)} \bmod K_y[t]) + K_x[t]K_y[t](k_v^{(0)} \bmod K_z[t]). \quad (6.40)$$

Inside each anchor cell, the requested voxels receive distinct local labels

$$\ell_v[t] \in \{0, 1, \dots, n_a[t] - 1\}, \quad a(v) = a. \quad (6.41)$$

The final sequence index is

$$s_v[t] \triangleq 1 + b_v[t]n_{\max}[t] + \ell_v[t]. \quad (6.42)$$

By construction,

$$K_x[t]K_y[t]K_z[t]n_{\max}[t] \leq B[t]n_{\max}[t] \leq T, \quad (6.43)$$

so the number of used indices never exceeds the available DFT family. All APs that later illuminate voxel  $v$  use the same waveform  $d_{s_v[t]}[\tau]$ . Distinct fine voxels inside the same anchor cell are always separated by different local labels, while sufficiently separated anchor cells may reuse the same sequence index.

### 6.6.2 Coverage-balanced candidate distribution

For every requested voxel, define the eligible active DL neighborhood

$$\mathcal{N}_v^{(d)}[t] \triangleq \mathcal{A}_d[t] \cap \mathcal{A}(v) = \{j \in \mathcal{A}_d[t] : d(j, v) \leq r_v\}. \quad (6.44)$$

Each active DL AP receives at most  $N$  candidate voxels,

$$C_j[t] \triangleq N, \quad j \in \mathcal{A}_d[t], \quad (6.45)$$

because an AP with  $N$  antennas cannot finally support more than  $N$  simultaneous DL branches.

Let

$$C_{\text{tot}}[t] \triangleq \sum_{j \in \mathcal{A}_d[t]} C_j[t] = N|\mathcal{A}_d[t]|. \quad (6.46)$$

Given a desired minimum replication parameter  $L_{\min}$ , the slot-adaptive replication level is

$$L_{\min}^{\text{act}}[t] \triangleq \max \left\{ 1, \min \left( L_{\min}, \left\lfloor \frac{C_{\text{tot}}[t]}{|\mathcal{V}_{\text{scan}}[t]|} \right\rfloor \right) \right\}, \quad (6.47)$$

and the target replication count for voxel  $v$  is

$$L_v^{\text{tar}}[t] \triangleq \min\{L_{\min}^{\text{act}}[t], |\mathcal{N}_v^{(d)}[t]|\}. \quad (6.48)$$

Thus the CPU first tries to make coverage broad, and only then spends leftover candidate capacity on redundancy.

Introduce binary distribution variables

$$z_{jv}[t] \in \{0, 1\}, \quad (6.49)$$

with  $z_{jv}[t] = 0$  for  $j \notin \mathcal{N}_v^{(d)}[t]$  and

$$\sum_{v \in \mathcal{V}_{\text{scan}}[t]} z_{jv}[t] \leq N, \quad c_v[t] \triangleq \sum_{j \in \mathcal{A}_d[t]} z_{jv}[t]. \quad (6.50)$$

The greedy round-fill rule initializes residual capacities  $b_j[t] \leftarrow N$  and counts  $c_v[t] \leftarrow 0$ . For every feasible edge, define the proximity score

$$q_{jv}[t] \triangleq \left( \frac{r_v - d_{jv}}{r_v} \right)^+ \in [0, 1]. \quad (6.51)$$

For rounds  $\ell = 1, \dots, L_{\min}^{\text{act}}[t]$ , the CPU processes the still-under-covered voxels in ascending order of the number of currently available eligible APs. For a voxel  $v$ , it chooses

$$j^*(v) = \arg \max_{j \in \mathcal{N}_v^{(d)}[t]: b_j[t] > 0, z_{jv}[t] = 0} \left[ \lambda_q q_{jv}[t] + (1 - \lambda_q) \frac{b_j[t]}{N} \right], \quad (6.52)$$

where  $\lambda_q \in [0, 1]$  balances geometric proximity and remaining AP capacity. It then sets  $z_{j^*(v),v}[t] = 1$ , decreases  $b_{j^*(v)}[t]$ , and increases  $c_v[t]$ . A single swap-repair pass may move an assignment from an over-replicated donor voxel to a deficit voxel, and a final saturation pass may spend any leftover AP capacity on additional feasible edges.

The resulting local candidate set sent to DL AP  $j$  is

$$\mathcal{V}_j^{\text{CPU}}[t] \triangleq \{v \in \mathcal{V}_{\text{scan}}[t] : z_{jv}[t] = 1\}, \quad |\mathcal{V}_j^{\text{CPU}}[t]| \leq N. \quad (6.53)$$

Together with each voxel index, the CPU sends the sequence index  $s_v[t]$  from (6.42). The sensing admission decisions are deferred to Chapter 7, because they depend on the final payload beams, probing ceilings, and residual power at each AP. The scheduling layer therefore guarantees bounded candidate exposure, while the physical layer decides which of those candidates become actual illumination.

## 6.7 UL SENSING ASSIGNMENT

The control layer closes by deciding which voxel hypotheses each UL AP will process in the sensing branch. The UL sensing assignment is built from the CPU-selected candidate set

$$\mathcal{V}^{\text{CPU}}[t] \triangleq \bigcup_{j \in \mathcal{A}_d[t]} \mathcal{V}_j^{\text{CPU}}[t]. \quad (6.54)$$

This set contains the voxel hypotheses that may become illuminated in the slot.

For each UL AP  $m \in \mathcal{A}_u[t]$ , the CPU forms the local sensing candidate list

$$\mathcal{C}_m^{\text{sens}}[t] \triangleq \{v \in \mathcal{V}^{\text{CPU}}[t] : d_{mv} \leq r_v\}. \quad (6.55)$$

Thus, an UL AP is asked to process only the voxel hypotheses for which it belongs to the voxel neighborhood already defined in (6.22). The number of processed voxel hypotheses is controlled by the post-processing cap  $B_u^{\text{max}}$ , which denotes the maximum number of voxel matched filters and local detection statistics evaluated by one UL AP in one slot. The local sensing list satisfies

$$\mathcal{V}_m^{\text{sens}}[t] \subseteq \mathcal{C}_m^{\text{sens}}[t], \quad |\mathcal{V}_m^{\text{sens}}[t]| \leq B_u^{\text{max}}, \quad (6.56)$$

and the receiver set associated with voxel  $v$  is

$$\mathcal{M}_v[t] \triangleq \{m \in \mathcal{A}_u[t] : v \in \mathcal{V}_m^{\text{sens}}[t]\}. \quad (6.57)$$

The UL AP receives one  $N$ -dimensional baseband block over the payload interval and can apply several voxel-dependent receive combiners to that same block, as detailed in Chapter 8.

If the local candidate list is already short, the CPU assigns all its entries,

$$|\mathcal{C}_m^{\text{sens}}[t]| \leq B_u^{\text{max}} \implies \mathcal{V}_m^{\text{sens}}[t] = \mathcal{C}_m^{\text{sens}}[t]. \quad (6.58)$$

If the candidate list is longer than  $B_u^{\text{max}}$ , a coverage-first logic is applied. For every feasible pair  $(m, v)$ , define a large-scale receive-quality score

$$q_{mv}^{(u)}[t] \triangleq \left( \frac{r_v - d_{mv}}{r_v} \right)^+. \quad (6.59)$$

The primary pass scans the selected voxel set  $\mathcal{V}^{\text{CPU}}[t]$  and attempts to give each voxel at least one UL processor. For voxel  $v$ , the set of non-full candidate receivers is

$$\mathcal{A}_{u,v}^{\text{free}}[t] \triangleq \{m \in \mathcal{A}_u[t] : v \in \mathcal{C}_m^{\text{sens}}[t], |\mathcal{V}_m^{\text{sens}}[t]| < B_u^{\text{max}}\}. \quad (6.60)$$

If  $\mathcal{M}_v[t] = \emptyset$  and  $\mathcal{A}_{u,v}^{\text{free}}[t] \neq \emptyset$ , the CPU chooses

$$m^*(v) = \arg \max_{m \in \mathcal{A}_{u,v}^{\text{free}}[t]} q_{mv}^{(u)}[t] \quad (6.61)$$

and inserts  $v$  in  $\mathcal{V}_{m^*(v)}^{\text{sens}}[t]$ .

The diversity pass uses any remaining UL processing capacity to add further multi-static observations. Let

$$L_v^{(u)}[t] \triangleq |\mathcal{M}_v[t]| \quad (6.62)$$

be the current number of assigned UL processors for voxel  $v$ . Every admissible pair  $(m, v)$  with  $v \in \mathcal{C}_m^{\text{sens}}[t]$ ,  $v \notin \mathcal{V}_m^{\text{sens}}[t]$ , and  $|\mathcal{V}_m^{\text{sens}}[t]| < B_u^{\text{max}}$  receives the score

$$\psi_{mv}^{(u)}[t] \triangleq \frac{q_{mv}^{(u)}[t]}{1 + L_v^{(u)}[t]}. \quad (6.63)$$

The CPU greedily adds the largest-score admissible pairs until no pair remains or all active UL APs have reached the cap in (6.56). The denominator in (6.63) prevents a single well-positioned voxel from absorbing all remaining receiver-side processing budget.

After the slot is executed, the actual informative set is determined from the DL admission reports and the voxel-wise detection statistics. For each candidate voxel, the DL physical layer reveals whether at least one deterministic payload-attached or dedicated branch was admitted. A voxel can be informative only if it has both an admitted DL deterministic branch and at least one reporting UL receiver from  $\mathcal{M}_v[t]$ . The CPU then forms the illumination bitmap  $\{I[v, t]\}$  and updates  $\mathcal{V}_{\text{inf}}[t]$  through (6.2).

## 6.8 FRONTHAUL AND COMPUTATIONAL SCALABILITY

The scheduling layer is scalable by design: it operates on bounded user-centric sets, requested voxels, or short AP neighborhoods, and therefore avoids dense online AP-voxel tables, full-grid steering libraries, or all-to-all fast-timescale searches. Table 6.1 summarizes the dominant control costs.

Sequence planning limits voxel exposure before AP-local admission, and each DL AP receives at most  $N$  candidates. UL sensing is local and budgeted, while sensing-specific signaling is lower order. Thus the main fronthaul load remains the ordinary payload and UL soft-symbol return, and the architecture meets the CF-ISAC scalability criterion from Chapter 5.

| EXCHANGED OBJECT                                                       | DOMINANT ONLINE COST                                                                                                                            |
|------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|
| Local UE-cluster indicators                                            | Traffic-load counting and pilot-aware capping:<br>$\mathcal{O}\left(\sum_{k \in \mathcal{U}_{\text{act}}[t]}  \mathcal{C}_k \right)$            |
| Queried voxel neighborhoods                                            | DTDD scheduling and requested-voxel correction: $\mathcal{O}(M +  \mathcal{V}_{\text{scan}}[t] (\log M + L_v))$                                 |
| Track metadata and informative bitmap                                  | Bridge-capped scan maintenance:<br>$\mathcal{O}( \mathcal{V}_{\text{scan}}[t]  + T_{\text{br}}[t])$                                             |
| Bounded candidate lists and sequence indices to active DL APs          | Sequence planning and DL candidate distribution: $\mathcal{O}(V[t] \log V[t] + V[t]L_v)$                                                        |
| Sensing lists with at most $B_u^{\max}$ voxel indices to active UL APs | UL sensing assignment:<br>$\mathcal{O}\left( \mathcal{V}^{\text{CPU}}[t] L_v + \sum_{m \in \mathcal{A}_u[t]} B_u^{\max} \log B_u^{\max}\right)$ |
| Nuisance scalars, final illumination reports                           | Sensing-specific control signaling: lower order                                                                                                 |

Table 6.1: Dominant online cost and control signaling of the scheduling layer. Here  $V[t] = |\mathcal{V}_{\text{scan}}[t]|$ ,  $\mathcal{V}^{\text{CPU}}[t] = \bigcup_{j \in \mathcal{A}_d[t]} \mathcal{V}_j^{\text{CPU}}[t]$ , and  $S_d[t] = \sum_{j \in \mathcal{A}_d[t]} |\mathcal{D}_j[t]|$ . The control-layer fronthaul remains lower order with respect to the main payload distribution and UL soft-symbol or raw samples return flow.



# 7

## HYBRID DOWNLINK SIGNAL DESIGN

This chapter shows how the proposed DTDD CF-ISAC network builds a hybrid DL signal that preserves user service, adds deterministic probing when possible, and uses residual spatial and power resources for dedicated sensing beams.

The chapter follows the protocol execution order. First, after the common pilot phase, each active DL AP constructs unit-norm protected payload directions for its served-UE set. Second, the AP reports one effective-gain scalar per served stream to the CPU, which uses it to update the payload-gain state and assign payload powers. Third, each AP evaluates the requested-voxel candidate set received from the CPU and admits only payload-attached pairs with sufficient sensing utility. Fourth, a minimum probing fraction is assigned to each admitted payload stream under the communication-feasibility rule. Fifth, remaining local candidate voxels are considered for dedicated sensing. Once these building blocks are defined, the complete hybrid DL signal is written explicitly. The communication-side reception model, the resulting DL SINR/SE expressions, and the DL-layer scalability recap then follow.

In the literature, ISAC waveform design is often described either as communication centric reuse of existing communication beams or as a joint design with additional sensing branches [10, 13, 14]. The protocol developed here lies on the communication-constrained side of that spectrum, but it is not payload-only. It is hybrid in a precise sense: the DL first reuses UE-directed payload beams through probing, then invokes dedicated sensing directions for remaining local candidates supported by residual spatial and power resources. The resulting architecture is summarized in Table 7.1.

| DL BRANCH                | MAIN CARRIER                                                              | FUNCTION                                                                                             |
|--------------------------|---------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------|
| Payload                  | UE data streams on bounded local service sets                             | Preserves communication feasibility and provides the communication baseline                          |
| Payload-attached probing | Voxel-indexed deterministic sequences superimposed on fixed payload beams | Illuminates admitted local voxel candidates by reusing active communication streams                  |
| Dedicated sensing        | Dedicated deterministic beams on remaining local voxel candidates         | Exploits residual spatial and power resources, while remaining separated from the served-UE subspace |

Table 7.1: Hybrid layers of the proposed downlink design. The communication branch is primary; probing reuses that branch when feasible; dedicated sensing is invoked only for remaining local candidates that pass the dedicated utility and power tests.

## 7.1 DOWNLINK DESIGN OBJECTIVES

The downlink task is to construct a slot-level transmit architecture satisfying three simultaneous requirements.

First, the payload streams must remain communication-feasible. The communication baseline is therefore defined before any sensing branch is attached to it, and the later sensing decisions are allowed to modify only scalar power splits or to use residual subspaces, but not to invalidate the service guarantee already established for the DL UEs.

Second, the waveform should exploit already active payload streams for sensing whenever possible. This is the most resource-efficient way to illuminate a requested voxel, because a payload stream already consumes one spatial DoF and already receives UE data from the fronthaul. If such a stream can also carry a known deterministic probing component while still satisfying the communication constraint, then sensing is obtained without introducing an additional beamformer bank [13, 14].

Third, if payload reuse is insufficient, the AP may exploit leftover spatial resources for additional sensing. This is the role of the dedicated sensing branch. However,

it must remain compatible with the strong-InAI protection logic inherited from the DTDD setting and should avoid spending power inside the subspace already occupied by the served-UE payload streams.

As in the previous chapters, a slot corresponds to one coherence block of  $T_{\text{slot}}$  channel uses, indexed by  $\tau = 1, \dots, T_{\text{slot}}$ . The common pilot phase is followed by a short scalar-control exchange before the payload/sensing symbols of the same slot. All scheduling and power-allocation variables are fixed over the slot, while the deterministic sequences and the UE data symbols vary with  $\tau$ . Coherent multi-AP support is assumed throughout the chapter: phase calibration/synchronization among distributed APs is not treated as a bottleneck. Larsson [50] demonstrates that reciprocity-based coherent DL beamforming can be supported in distributed antenna systems via scalable Over-The-Air (OTA) calibration protocols.

## 7.2 PROTECTED LOCAL PAYLOAD PRECODING

Let  $\hat{\mathbf{h}}_{jn}[t] \in \mathbb{C}^N$  denote the UL-pilot-based channel estimate at DL AP  $j$  for UE  $n \in \mathcal{D}_j[t]$ . The local payload beam is designed to remain communication-oriented, but it is also shaped to be safer for the simultaneous UL side of the DTDD network. This is achieved through a protected regularization matrix that augments the usual local communication Gram matrix with a bounded penalty toward the most vulnerable UL APs neighbors.

Each DL AP  $j$  stores a slow-timescale strong-InAI protection list

$$\mathcal{N}_j^{\text{IA}} \triangleq \text{Top}_{L_{\text{IA}}} \{m : \zeta_{mj}\}, \quad (7.1)$$

where  $L_{\text{IA}}$  is a small protocol constant and  $\zeta_{mj}$  is the nuisance-priority coefficient received from the scheduling layer. In slot  $t$ , only the active UL neighbors matter, hence

$$\mathcal{N}_j^{\text{IA,act}}[t] \triangleq \mathcal{N}_j^{\text{IA}} \cap \mathcal{A}_u[t]. \quad (7.2)$$

For each protected neighbor  $m \in \mathcal{N}_j^{\text{IA}}$ , AP  $j$  stores a unit-norm steering proxy  $\mathbf{d}_{jm} \in \mathbb{C}^N$  pointing toward the physical location of that UL AP.

The protected precoder is first constructed as a unit-norm direction, using the local channel geometry. Define the unpowered local communication Gram matrix

$$\mathbf{S}_j[t] \triangleq \sum_{i \in \mathcal{D}_j[t]} \hat{\mathbf{h}}_{ji}[t] \hat{\mathbf{h}}_{ji}^H[t], \quad (7.3)$$

and its average per-antenna scale

$$s_j[t] \triangleq \frac{1}{N} \text{tr}(\mathbf{S}_j[t]), \quad \bar{s}_j[t] \triangleq \max\{s_j[t], s_{\min}\}, \quad (7.4)$$

with a small numerical floor  $s_{\min} > 0$ . The stored-neighbor normalization constant is

$$Z_j^{\text{store}} \triangleq \sum_{r \in \mathcal{N}_j^{\text{IA}}} \zeta_{rj}^{\text{res}}, \quad (7.5)$$

and the stored-neighbor normalized weights are

$$\mu_{mj}^{\text{store}} \triangleq \frac{\zeta_{mj}^{\text{res}}}{\max\{Z_j^{\text{store}}, \epsilon_\zeta\}}, \quad m \in \mathcal{N}_j^{\text{IA}}. \quad (7.6)$$

The active penalty shape is

$$\mathbf{P}_j^{\text{IA}}[t] \triangleq \sum_{m \in \mathcal{N}_j^{\text{IA,act}}[t]} \mu_{mj}^{\text{store}} \mathbf{d}_{jm} \mathbf{d}_{jm}^H. \quad (7.7)$$

The active mass of this penalty is not forced to one:

$$\sum_{m \in \mathcal{N}_j^{\text{IA,act}}[t]} \mu_{mj}^{\text{store}} = \frac{\sum_{m \in \mathcal{N}_j^{\text{IA,act}}[t]} \zeta_{mj}^{\text{res}}}{\sum_{r \in \mathcal{N}_j^{\text{IA}}} \zeta_{rj}^{\text{res}}} \leq 1, \quad (7.8)$$

thus the spatial penalty becomes weaker when only a small part of the stored vulnerable neighborhood is active in UL mode.

The scalar strength of the protection is also adaptive. Define the active residual InAI-to-noise threat of DL AP  $j$  as

$$\Gamma_{\text{IA},j}[t] \triangleq \frac{P_{j,\text{DL}}^{\text{ref}}[t] \sum_{m \in \mathcal{N}_j^{\text{IA,act}}[t]} \zeta_{mj}^{\text{res}}}{N_0}, \quad (7.9)$$

where  $P_{j,\text{DL}}^{\text{ref}}[t]$  is a reference downlink power, for instance the planned total radiated power of AP  $j$ . The effective protection coefficient is

$$\lambda_{\text{IA},j}^{\text{eff}}[t] = \bar{\lambda}_{\text{IA}} \min\left\{\frac{\Gamma_{\text{IA},j}[t]}{\Gamma_{\text{IA}}^{\text{ref}}}, 1\right\}, \quad (7.10)$$

with maximum strength  $\bar{\lambda}_{\text{IA}}$  and design reference  $\Gamma_{\text{IA}}^{\text{ref}} > 0$ . When the active residual InAI threat is weak, the beam remains close to ordinary local regularized beamforming; when the threat reaches the reference level, the full protection strength is recovered.

The protected regularization matrix is

$$\mathbf{R}_j[t] \triangleq \mathbf{S}_j[t] + \lambda_{\text{IA},j}^{\text{eff}}[t] \bar{s}_j[t] \mathbf{P}_j^{\text{IA}}[t] + \bar{\nu} \bar{s}_j[t] \mathbf{I}_N, \quad (7.11)$$

where the first term is the local communication Gram matrix; the second is the adaptive InAI-aware spatial penalty; and the third is a scale-normalized diagonal loading. For each served UE  $n \in \mathcal{D}_j[t]$ , AP  $j$  computes

$$\tilde{\mathbf{w}}_{jn}^{\text{P}}[t] \triangleq \mathbf{R}_j^{-1}[t] \hat{\mathbf{h}}_{jn}[t], \quad \mathbf{w}_{jn}^{\text{P}}[t] \triangleq \frac{\tilde{\mathbf{w}}_{jn}^{\text{P}}[t]}{\|\tilde{\mathbf{w}}_{jn}^{\text{P}}[t]\|}. \quad (7.12)$$

The payload beam is therefore protected through the soft regularization in (7.11), while remaining fully communication-oriented in the transmit space.

### 7.2.1 Scalar feedback and CPU-tracked payload power loading

After constructing  $\mathbf{w}_{jn}^{\text{P}}[t]$ , AP  $j$  computes the effective desired-link gain

$$\Psi_{jn}[t] \triangleq |\hat{\mathbf{h}}_{jn}^H[t] \mathbf{w}_{jn}^{\text{P}}[t]|, \quad n \in \mathcal{D}_j[t]. \quad (7.13)$$

This scalar is reported to the CPU after the pilot phase: since  $|\mathcal{D}_j[t]| \leq \tau_p$ , the per-AP scalar feedback remains bounded. The instantaneous scalar used for power loading is

$$\xi_{jn}[t] \triangleq \Psi_{jn}^2[t], \quad (7.14)$$

and the CPU tracks it by exponential smoothing,

$$\hat{\xi}_{jn}[t] = (1 - \lambda_\xi) \hat{\xi}_{jn}[t-1] + \lambda_\xi \xi_{jn}[t], \quad 0 < \lambda_\xi \leq 1. \quad (7.15)$$

The case  $\lambda_\xi = 1$  corresponds to using the same-slot reported scalar without temporal smoothing. For compactness, define  $\hat{\Psi}_{jn}[t] \triangleq \sqrt{\hat{\xi}_{jn}[t]}$ . The pre-probing payload budget assigned by the CPU to pair  $(j, n)$  is then

$$P_{d,jn}[t] = P_j^{\max} \frac{\hat{\xi}_{jn}[t]}{\sum_{k \in \mathcal{D}_j[t]} \hat{\xi}_{jk}[t] + \epsilon_P}, \quad n \in \mathcal{D}_j[t]. \quad (7.16)$$

### 7.2.2 Protection trade-offs and applicability

The natural alternative to the protected precoder is to construct a classical LP-MMSE precoder and rely on the UL receiver to model and suppress the resulting leakage. The protected design takes a fundamentally different approach: the penalty matrix  $\mathbf{P}_j^{\text{IA}}[t]$  in (7.7) penalizes energy in the directions of the most vulnerable UL, and the adaptive scalar  $\lambda_{\text{IA},j}^{\text{eff}}[t]$  in (7.10) modulates the penalty strength in proportion to the actual residual threat. The key distinction from treating InAI as noise at the receiver is that the interference power injected into neighboring UL APs is reduced before any combiner ever sees it, rather than being accepted as given and then processed away. Compared with a noise-tolerant approach, the transmit-side strategy also lightens the subsequent receive-side burden: a receiver confronted with less incident interference needs a weaker directional penalty, preserving more combining gain for the desired UL signal. The protection carries four practical costs that must be weighed against its benefit.

- *Beamforming gain loss.* Adding the leakage-suppression term to the regularization matrix  $\mathbf{R}_j[t]$  effectively pulls the resulting precoder  $\mathbf{w}_{jn}^{\text{P}}[t]$  away from the pure channel-matched direction and provides less coherent array gain toward the served DL UE. The magnitude of this loss grows with the protection strength  $\lambda_{\text{IA},j}^{\text{eff}}[t]$  and with the angular proximity between the UL neighbor direction and the served-UE direction. Section 9.4 quantifies this loss: the protected precoder sacrifices a portion of DL SE, while recovering a substantially larger amount of UL SE in return.
- *Required inter-AP knowledge.* The protected precoder relies on slow-timescale inter-AP statistics that must be distributed through the control plane, such as the coefficient and the steering vector of each stored neighbor. The quality of the protection therefore depends on the accuracy of the available coupling model: poorly calibrated inter-AP path-loss estimates lead to misaligned penalty directions, reducing the actual advantage.

- *Storage per AP.* Each DL AP  $j$  must store the protection list  $\mathcal{N}_j^{\text{IA}}$  of up to  $L_{\text{IA}}$  entries, the associated weights, and the spatial direction vectors. The total memory cost per AP is  $\mathcal{O}(L_{\text{IA}}N)$  complex scalars. However, since  $L_{\text{IA}}$  is a small protocol constant, this overhead is marginal relative to the channel estimate buffer, which already requires  $\mathcal{O}(|\mathcal{D}_j[t]|N)$  complex entries and is rebuilt every slot.
- *Computational overhead.* The main online cost is forming the active penalty shape  $\mathbf{P}_j^{\text{IA}}[t]$  from the stored direction vectors and evaluating the scalar  $\lambda_{\text{IA},j}^{\text{eff}}[t]$ . Both steps involve only a summation over at most  $L_{\text{IA}}$  entries and one scalar comparison. As reflected in (7.68), the overhead appears only in the  $\mathcal{O}((|\mathcal{D}_j[t]| + L_{\text{IA}})N^2)$  term, which remains bounded independently of the network size.

**MOST BENEFICIAL SCENARIOS.** The adaptive scalar  $\lambda_{\text{IA},j}^{\text{eff}}[t]$  already implements a continuous version of activation logic, but it is instructive to identify the conditions under which the gain from the protection is most pronounced.

- *Strong residual InAI.* The benefit grows proportionally to the residual coupling  $\zeta_{mj}^{\text{res}}$ . When the inter-AP isolation is limited, due to short physical separation, limited antenna pattern isolation, or the absence of hardware cancellation between nodes, the unprotected precoder radiates substantial power toward receiving neighbors.
- *High DL transmit power.* The active threat measure  $\Gamma_{\text{IA},j}[t]$  in (7.9) is proportional to  $P_{j,\text{DL}}^{\text{ref}}[t]$ . When DL APs operate near their power ceiling the absolute InAI power injected into neighboring receivers increases even if the coupling ratio remains constant.
- *Dense deployments.* In a dense CF deployment, inter-AP distances are shorter, and the probability that a DL AP and a UL AP are in close proximity is higher. In such configurations the directional penalty can provide its greatest benefit

because rotating the beam slightly away from the victim direction substantially reduces the leaked power.

- *Limited hardware interference cancellation.* Unlike self-interference in full-duplex systems, the InAI in DTDD arises between different physical APs. No local analog or digital canceller at the transmitter can reduce this leakage, because the transmitting and receiving hardware are not co-located. The protected precoder and the complementary combiner of Chapter 8 are therefore the primary network-level tools available to suppress this interference in the absence of dedicated hardware cancellation paths between APs.

### 7.3 HYBRID VOXEL ADMISSION AND PROBING SPLIT

Once the communication baseline  $\mathbf{w}_{jn}^P[t]$  is fixed, AP  $j$  evaluates the candidate set  $\mathcal{V}_j^{\text{CPU}}[t]$  delivered by the CPU. For every candidate voxel, it reconstructs the unit-norm departure steering vector  $\bar{\mathbf{a}}_{jv}[t] \in \mathbb{C}^N$  and computes the fixed-beam directional score

$$\phi_{jnv}[t] \triangleq |\bar{\mathbf{a}}_{jv}^H[t] \mathbf{w}_{jn}^P[t]|^2, \quad 0 \leq \phi_{jnv}[t] \leq 1. \quad (7.17)$$

This score measures what the already-fixed communication beam can contribute toward voxel  $v$ .

To determine how much deterministic probing can be superimposed on a fixed payload beam, the AP needs a model of the communication support already provided to the corresponding UE by its other serving APs. The exact instantaneous DL SINR is a network-wide quantity and would require all cross-effective channels. The protocol therefore uses the pilot-capped serving sets from Chapter 6 and forms a conservative scalar support law.

For a DL UE  $n$ , define its actual slot-level DL serving set as

$$\mathcal{S}_n^{(d)}[t] \triangleq \{i \in \mathcal{A}_d[t] : n \in \mathcal{D}_i[t]\}. \quad (7.18)$$

The coherent support available to stream  $(j, n)$  from the other serving APs is approximated as

$$\Omega_{jn}^{\text{safe}}[t] \triangleq \kappa_{\Omega} \sum_{i \in \mathcal{S}_n^{(d)}[t] \setminus \{j\}} \sqrt{P_{d,in}[t]} \hat{\Psi}_{in}[t], \quad 0 < \kappa_{\Omega} \leq 1. \quad (7.19)$$

The factor  $\kappa_{\Omega} \leq 1$  deliberately deflates the pre-probing coherent support. This is necessary because the CPU computes  $\Omega_{jn}^{\text{safe}}[t]$  before knowing the final probing fractions chosen at the other APs; those APs may also spend part of their payload streams on deterministic probing. Thus  $\Omega_{jn}^{\text{safe}}[t]$  is not an exact final coherent amplitude, but a conservative pre-probing support proxy.

The first-hop co-served user set is

$$\mathcal{K}_n^{(1)}[t] \triangleq \left( \bigcup_{i \in \mathcal{S}_n^{(d)}[t]} \mathcal{D}_i[t] \right) \setminus \{n\}. \quad (7.20)$$

For any DL UE  $k$ , let

$$\mathcal{S}_k^{(d)}[t] \triangleq \{i \in \mathcal{A}_d[t] : k \in \mathcal{D}_i[t]\}. \quad (7.21)$$

The corresponding overlap-local AP set is

$$\mathcal{A}_n^{(2)}[t] \triangleq \bigcup_{k \in \mathcal{K}_n^{(1)}[t] \cup \{n\}} \mathcal{S}_k^{(d)}[t], \quad (7.22)$$

which contains the active DL APs serving either UE  $n$  or any of its first-hop co-served UEs. Under the pilot-aware capping of Chapter 6, the sizes of these sets remain bounded. With

$$P_{d,i}^{\text{tot}}[t] \triangleq \sum_{k \in \mathcal{D}_i[t]} P_{d,ik}[t], \quad (7.23)$$

the local nuisance approximation is

$$\Upsilon_n^{\text{safe}}[t] \triangleq \eta \left[ N_0 + \sum_{i \in \mathcal{A}_n^{(2)}[t]} \beta_{in} P_{d,i}^{\text{tot}}[t] - \sum_{i \in \mathcal{S}_n^{(d)}[t]} \beta_{in} P_{d,in}[t] \right]_+, \quad (7.24)$$

where  $\beta_{in}$  is the slow large-scale channel coefficient from AP  $i$  to UE  $n$ ,  $\eta > 1$  is a nuisance-inflation factor, and  $[\cdot]_+ \triangleq \max\{\cdot, 0\}$ .

Using the locally available current effective gain  $\Psi_{jn}[t]$  from (7.13), the maximum communication-feasible probing fraction is

$$\bar{\rho}_{jn}[t] = \left[ 1 - \frac{(\sqrt{\Gamma_n^{\min} \Upsilon_n^{\text{safe}}[t]} - \Omega_{jn}^{\text{safe}}[t])_+^2}{P_{d,jn}[t] \Psi_{jn}^2[t] + \epsilon_P} \right]_{[0,1]}, \quad (7.25)$$

where  $[x]_{[0,1]} \triangleq \min\{1, \max\{0, x\}\}$ .

The hybrid branch is admitted through a normalized utility test. Define the local reference probing power

$$P_j^{\text{eq}}[t] \triangleq \frac{\sum_{n \in \mathcal{D}_j[t]} \bar{\rho}_{jn}[t] P_{d,jn}[t]}{\max\{|\{n \in \mathcal{D}_j[t] : \bar{\rho}_{jn}[t] > 0\}|, 1\}} \quad (7.26)$$

as the average probing power per active stream if all streams were probed at the maximum communication-feasible fraction. For each triple  $(j, n, v)$ , the maximum normalized achievable hybrid utility is

$$\eta_{jnv}^{\text{hyb,max}}[t] \triangleq \min\left\{1, \frac{\bar{\rho}_{jn}[t] P_{d,jn}[t] \phi_{jnv}[t]}{P_j^{\text{eq}}[t] + \epsilon_P}\right\}. \quad (7.27)$$

This quantity measures the maximum sensing utility that stream  $(j, n)$  can achieve on voxel  $v$  if probed at the maximum communication-feasible fraction, normalized by the reference probing power. The hybrid branch is admitted only if this quantity exceeds a threshold  $\alpha_{\text{hyb}} \in (0, 1)$ . Hence, we can define the local hybrid utility graph with edge set

$$\mathcal{E}_j^{\text{hyb}}[t] \triangleq \{(n, v) : n \in \mathcal{D}_j[t], v \in \mathcal{V}_j^{\text{CPU}}[t], \eta_{jnv}^{\text{hyb,max}}[t] \geq \alpha_{\text{hyb}}\}. \quad (7.28)$$

The AP then solves the small local matching

$$\max_{\{x_{jnv}[t]\}} \sum_{(n,v) \in \mathcal{E}_j^{\text{hyb}}[t]} \eta_{jnv}^{\text{hyb,max}}[t] x_{jnv}[t] \quad (7.29)$$

subject to

$$\sum_{v:(n,v) \in \mathcal{E}_j^{\text{hyb}}[t]} x_{jnv}[t] \leq 1, \quad \sum_{n:(n,v) \in \mathcal{E}_j^{\text{hyb}}[t]} x_{jnv}[t] \leq 1, \quad x_{jnv}[t] \in \{0, 1\}. \quad (7.30)$$

Let  $v_{j,n}^{\text{hyb}}[t]$  denote the voxel matched to stream  $(j, n)$ , if any, and define

$$\mathcal{V}_j^{\text{hyb}}[t] \triangleq \{v \in \mathcal{V}_j^{\text{CPU}}[t] : \exists n \in \mathcal{D}_j[t] \text{ with } x_{jnv}[t] = 1\}. \quad (7.31)$$

For an active matched pair, the minimum probing fraction that reaches the threshold is

$$\rho_{jn}^{\text{req}}[t] \triangleq \frac{\alpha_{\text{hyb}}(P_j^{\text{eq}}[t] + \epsilon_P)}{P_{d,jn}[t] \phi_{jnv_{j,n}^{\text{hyb}}}[t] + \epsilon_P}, \quad (7.32)$$

and the final probing fraction is

$$\rho_{jn}[t] \triangleq \min\{\bar{\rho}_{jn}[t], \rho_{jn}^{\text{req}}[t]\}. \quad (7.33)$$

If stream  $(j, n)$  is unmatched,  $\rho_{jn}[t] = 0$ . Thus the hybrid branch uses the minimum probing power required to make an admitted payload-voxel association useful.

The actual hybrid utility of an active payload-probing stream is

$$u_{jn}^{\text{hyb}}[t] \triangleq \rho_{jn}[t] P_{d,jn}[t] \phi_{jnv_{j,n}^{\text{hyb}}}[t], \quad (7.34)$$

and the number of active hybrid streams is

$$N_j^{\text{hyb}}[t] \triangleq |\{n \in \mathcal{D}_j[t] : \rho_{jn}[t] > 0\}|. \quad (7.35)$$

The reference utility used by the dedicated branch is

$$U_j^{\text{ref}}[t] \triangleq \begin{cases} \frac{\sum_{n \in \mathcal{D}_j[t]} u_{jn}^{\text{hyb}}[t]}{N_j^{\text{hyb}}[t]}, & N_j^{\text{hyb}}[t] \geq 1, \\ P_j^{\text{eq}}[t], & N_j^{\text{hyb}}[t] = 0. \end{cases} \quad (7.36)$$

## 7.4 DEDICATED SENSING ON REMAINING VOXELS

After the hybrid matching stage, the remaining local candidates at DL AP  $j$  are

$$\mathcal{V}_j^{\text{rem}}[t] \triangleq \mathcal{V}_j^{\text{CPU}}[t] \setminus \mathcal{V}_j^{\text{hyb}}[t]. \quad (7.37)$$

The dedicated sensing branch considers only this set. Its purpose is to add deterministic sensing beams when the AP has residual spatial and power resources, without spending dedicated sensing power in the served-UE subspace.

Define the local served-UE matrix

$$\mathbf{H}_j[t] = [\hat{\mathbf{h}}_{jn_1}[t], \dots, \hat{\mathbf{h}}_{jn_{|\mathcal{D}_j[t]|}}[t]], \quad (7.38)$$

and the projector onto the complement of its column space,

$$\mathbf{\Pi}_j^\perp[t] = \mathbf{I}_N - \mathbf{H}_j[t](\mathbf{H}_j^H[t]\mathbf{H}_j[t])^\dagger \mathbf{H}_j^H[t]. \quad (7.39)$$

The structural spatial budget for dedicated beams is

$$B_{j,\text{sp}}^{\text{ded}}[t] \triangleq (N - |\mathcal{D}_j[t]|)^+. \quad (7.40)$$

For a remaining candidate  $v \in \mathcal{V}_j^{\text{rem}}[t]$ , the adaptive protected seed is

$$\tilde{\mathbf{f}}_{jv}^{\text{IA}}[t] \triangleq \mathbf{R}_j^{-1}[t] \bar{\mathbf{a}}_{jv}[t], \quad (7.41)$$

and the dedicated candidate beam is formed by the served-UE-subspace projection

$$\bar{\mathbf{f}}_{jv}[t] \triangleq \mathbf{\Pi}_j^\perp[t] \tilde{\mathbf{f}}_{jv}^{\text{IA}}[t] = \mathbf{\Pi}_j^\perp[t] \mathbf{R}_j^{-1}[t] \bar{\mathbf{a}}_{jv}[t]. \quad (7.42)$$

If  $\|\bar{\mathbf{f}}_{jv}[t]\| \leq \epsilon_{\text{proj}}$ , the candidate is discarded. Otherwise,

$$\mathbf{f}_{jv}^{(s)}[t] \triangleq \frac{\bar{\mathbf{f}}_{jv}[t]}{\|\bar{\mathbf{f}}_{jv}[t]\|} \quad (7.43)$$

is the unit-norm dedicated sensing direction, and its directional score is

$$\psi_{jv}[t] \triangleq |\bar{\mathbf{a}}_{jv}^H[t] \mathbf{f}_{jv}^{(s)}[t]|^2. \quad (7.44)$$

The minimum power needed for voxel  $v$  to reach the dedicated utility threshold  $\alpha_{\text{ded}} \in (0, 1)$  is

$$P_{jv}^{\text{req}}[t] \triangleq \frac{\alpha_{\text{ded}}(U_j^{\text{ref}}[t] + \epsilon_U)}{\psi_{jv}[t] + \epsilon_\psi}. \quad (7.45)$$

For each served stream, define the minimum payload-branch power required after the chosen probing fraction as

$$P_{d,jn}^{\text{min}}[t] \triangleq \frac{(\sqrt{\Gamma_n^{\text{min}} \Upsilon_n^{\text{safe}}[t]} - \Omega_{jn}^{\text{safe}}[t])_+^2}{(1 - \rho_{jn}[t]) \Psi_{jn}^2[t] + \epsilon_P}, \quad n \in \mathcal{D}_j[t]. \quad (7.46)$$

Because  $\rho_{jn}[t] \leq \bar{\rho}_{jn}[t]$ , this quantity does not exceed the pre-probing allocation  $P_{d,jn}[t]$ , up to the numerical floor. The communication and payload-attached probing portions of this minimum protected branch are

$$P_{j,\text{comm}}^{\text{min}}[t] \triangleq \sum_{n \in \mathcal{D}_j[t]} (1 - \rho_{jn}[t]) P_{d,jn}^{\text{min}}[t], \quad P_{j,\text{hyb}}^{\text{min}}[t] \triangleq \sum_{n \in \mathcal{D}_j[t]} \rho_{jn}[t] P_{d,jn}^{\text{min}}[t]. \quad (7.47)$$

Their sum is the minimum total payload-branch power that must remain at AP  $j$  after the chosen probing fractions. The residual dedicated ceiling is therefore

$$P_{j,\text{max}}^{\text{ded}}[t] \triangleq [P_j^{\text{max}} - P_{j,\text{comm}}^{\text{min}}[t] - P_{j,\text{hyb}}^{\text{min}}[t]]_+, \quad \bar{\beta}_j[t] \triangleq \frac{P_{j,\text{max}}^{\text{ded}}[t]}{P_j^{\text{max}}}. \quad (7.48)$$

This definition makes the dedicated sensing budget an explicit residual-power quantity: the minimum data-bearing and payload-attached probing powers are reserved first, and dedicated sensing uses only the remaining slack between the nominal pre-probing allocation and the minimum protected payload branch.

The individually admissible dedicated candidate set is

$$\mathcal{C}_j^{\text{ded}}[t] \triangleq \{v \in \mathcal{V}_j^{\text{rem}}[t] : P_{jv}^{\text{req}}[t] \leq P_{j,\text{max}}^{\text{ded}}[t]\}. \quad (7.49)$$

The AP sorts these candidates by increasing  $P_{jv}^{\text{req}}[t]$  and selects the largest prefix such that

$$Q_j^{(s)}[t] \leq \min\{B_{j,\text{sp}}^{\text{ded}}[t], |\mathcal{C}_j^{\text{ded}}[t]|\}, \quad \sum_{q=1}^{Q_j^{(s)}[t]} P_{j,(q)}^{\text{req}}[t] \leq P_{j,\text{max}}^{\text{ded}}[t]. \quad (7.50)$$

The selected dedicated beams are the lowest-required-power candidates that fit the residual power and spatial budgets. Let their voxel indices be  $v_{j,1}^{(s)}, \dots, v_{j,Q_j^{(s)}[t]}^{(s)}$ .

Each selected candidate first receives its minimum required power,

$$P_{j,q}^{(s),\text{min}}[t] \triangleq P_{jv_{j,q}^{(s)}}^{\text{req}}[t]. \quad (7.51)$$

The remaining dedicated headroom is

$$P_{j,\text{ded}}^{\text{left}}[t] \triangleq P_{j,\text{max}}^{\text{ded}}[t] - \sum_{q=1}^{Q_j^{(s)}[t]} P_{j,q}^{(s),\text{min}}[t]. \quad (7.52)$$

A fixed coefficient  $\lambda_{\text{left}} \in [0, 1]$  splits this headroom into a dedicated part and a communication-return part,

$$P_{j,\text{left}}^{\text{ded}}[t] \triangleq \lambda_{\text{left}} P_{j,\text{ded}}^{\text{left}}[t], \quad P_{j,\text{left}}^{\text{com}}[t] \triangleq (1 - \lambda_{\text{left}}) P_{j,\text{ded}}^{\text{left}}[t]. \quad (7.53)$$

If no dedicated beam is active,  $\lambda_{\text{left}}$  is set to zero so that all leftover power returns to communication.

The dedicated part is distributed by equal utility increments. With

$$S_j^{\text{ded}}[t] \triangleq \sum_{q=1}^{Q_j^{(s)}[t]} \frac{1}{\psi_{jv_{j,q}^{(s)}}[t] + \epsilon_\psi}, \quad \Delta U_j^{\text{ded}}[t] \triangleq \frac{P_{j,\text{left}}^{\text{ded}}[t]}{S_j^{\text{ded}}[t] + \epsilon_S}, \quad (7.54)$$

the final dedicated powers are

$$P_{j,q}^{(s)}[t] \triangleq P_{j,q}^{(s),\text{min}}[t] + \frac{\Delta U_j^{\text{ded}}[t]}{\psi_{jv_{j,q}^{(s)}}[t] + \epsilon_\psi}, \quad q = 1, \dots, Q_j^{(s)}[t]. \quad (7.55)$$

The actually used dedicated power and split are

$$P_{j,\text{ded}}^{\text{used}}[t] \triangleq \sum_{q=1}^{Q_j^{(s)}[t]} P_{j,q}^{(s)}[t], \quad \beta_j^{\text{act}}[t] \triangleq \frac{P_{j,\text{ded}}^{\text{used}}[t]}{P_j^{\text{max}}}, \quad (7.56)$$

with  $0 \leq \beta_j^{\text{act}}[t] \leq \bar{\beta}_j[t]$ . The remaining communication-return power is distributed over the stream slacks

$$\Delta_{d,jn}[t] \triangleq [P_{d,jn}[t] - P_{d,jn}^{\text{min}}[t]]_+, \quad \Delta_{d,j}^{\text{tot}}[t] \triangleq \sum_{k \in \mathcal{D}_j[t]} \Delta_{d,jk}[t]. \quad (7.57)$$

The final payload-branch power of stream  $(j, n)$  is

$$P_{d,jn}^{(c)}[t] \triangleq P_{d,jn}^{\text{min}}[t] + P_{j,\text{left}}^{\text{com}}[t] \frac{\Delta_{d,jn}[t]}{\Delta_{d,j}^{\text{tot}}[t] + \epsilon_P}. \quad (7.58)$$

Thus each stream keeps at least the minimum power required by the probing-feasibility test, while any leftover headroom not spent on dedicated sensing is returned to the communication payloads rather than discarded. The actual data-bearing power on stream  $(j, n)$  is  $(1 - \rho_{jn}[t])P_{d,jn}^{(c)}[t]$ , and the actual payload-attached probing power is  $\rho_{jn}[t]P_{d,jn}^{(c)}[t]$ .

#### 7.4.1 Effective illumination scalar returned to the CPU

After the final probing fractions and dedicated beams have been selected, each DL AP returns to the CPU the effective transmit-side scalar induced by its locally transmitted deterministic branches. For a payload-attached probing branch on stream  $(j, n)$  matched to voxel  $v$ , define

$$\chi_{jn,v}^{\text{hyb}}[t] \triangleq \sqrt{\rho_{jn}[t]P_{d,jn}^{(c)}[t]} \bar{\mathbf{a}}_{jv}^H[t] \mathbf{w}_{jn}^P[t], \quad (7.59)$$

active only when  $\rho_{jn}[t] > 0$ . For a dedicated sensing branch  $q$  aimed at voxel  $v = v_{j,q}^{(s)}[t]$ , define

$$\chi_{jq,v}^{\text{ded}}[t] \triangleq \sqrt{P_{j,q}^{(s)}[t]} \bar{\mathbf{a}}_{jv}^H[t] \mathbf{f}_{jv}^{(s)}[t]. \quad (7.60)$$

The effective illumination scalar for voxel  $v$  at AP  $j$  is

$$\chi_{j,v}[t] \triangleq \begin{cases} \chi_{jn,v}^{\text{hyb}}[t], & \exists n \in \mathcal{D}_j[t] : v_{j,n}^{\text{hyb}}[t] = v, \\ \chi_{jq,v}^{\text{ded}}[t], & \exists q \in \{1, \dots, Q_j^{(s)}[t]\} : v_{j,q}^{(s)}[t] = v, \\ 0, & \text{otherwise.} \end{cases} \quad (7.61)$$

Let

$$\mathcal{V}_j^{\text{ill}}[t] \triangleq \{v : \chi_{j,v}[t] \neq 0\} \quad (7.62)$$

be the final set of voxels illuminated by AP  $j$ . The return report is

$$\mathcal{R}_j^{\text{DL}}[t] = \{(v, s_v[t], \chi_{j,v}[t]) : v \in \mathcal{V}_j^{\text{ill}}[t]\}. \quad (7.63)$$

This is a scalar quantity: the DL AP does not expose the full beamformer, and the CPU receives only the deterministic illumination support and the complex scalar needed to build the sensing signatures in Chapter 8.

## 7.5 HYBRID DOWNLINK TRANSMIT SIGNAL

After the protected payload beams, the local hybrid matching decisions, the minimum probing fractions, and the dedicated sensing branch are all fixed, the downlink waveform of AP  $j \in \mathcal{A}_d[t]$  can finally be written explicitly. For every matched payload stream, let  $v_{j,n}^{\text{hyb}}[t]$  denote its matched voxel; if the stream is unmatched,  $\rho_{jn}[t] = 0$  and the deterministic term is inactive. For every active dedicated beam  $q$ , let  $v_{j,q}^{(s)}[t]$  denote its selected voxel. The transmitted signal at payload-symbol time  $\tau$  is

$$\begin{aligned} \mathbf{x}_{d,j}[\tau] = \sum_{n \in \mathcal{D}_j[t]} \sqrt{P_{d,jn}^{(c)}[t]} \mathbf{w}_{jn}^{\text{P}}[t] & \left( \underbrace{\sqrt{\rho_{jn}[t]} d_{s_{v_{j,n}^{\text{hyb}}[t]}}[\tau]}_{\text{voxel-indexed probing}} + \underbrace{\sqrt{1 - \rho_{jn}[t]} s_n^{(\text{data})}[\tau]}_{\text{communication data}} \right) \\ & + \underbrace{\sum_{q=1}^{Q_j^{(s)}[t]} \sqrt{P_{j,q}^{(s)}[t]} \mathbf{f}_{jv_{j,q}^{(s)}[t]}^{(s)}[t] d_{s_{v_{j,q}^{(s)}[t]}}[\tau]}_{\text{dedicated sensing}}, \quad (7.64) \end{aligned}$$

where  $d_{s_v[t]}[\tau]$  is the deterministic waveform assigned to voxel  $v$  by the sequence plan of Chapter 6, and  $s_n^{(\text{data})}[\tau]$  is the data symbol with  $\mathbb{E}|s_n^{(\text{data})}[\tau]|^2 = 1$ . The first sum is the payload branch carrying both communication data and probing. The second sum is the dedicated sensing branch.

## 7.6 DOWNLINK RECEPTION MODEL AND METRICS

Consider a single-antenna DL UE  $n \in \mathcal{U}_d[t]$ . Let  $\mathbf{h}_{jn}[t] \in \mathbb{C}^N$  denote the channel between DL AP  $j$  and UE  $n$ , and let  $g_{kn}[t] \in \mathbb{C}$  denote the scalar inter-UE channel from simultaneously transmitting UL UE  $k \in \mathcal{U}_u[t]$  to DL UE  $n$ . If the UL UEs transmit at power  $P_u^{\max}$ , the received baseband signal is

$$r_{d,n}[\tau] = \sum_{j \in \mathcal{A}_d[t]} \mathbf{h}_{jn}^T[t] \mathbf{x}_{d,j}[\tau] + \sum_{k \in \mathcal{U}_u[t]} \sqrt{P_u^{\max}} g_{kn}[t] s_{u,k}[\tau] + w_{d,n}[\tau], \quad (7.65)$$

where  $w_{d,n}[\tau] \sim \mathcal{CN}(0, N_0)$  and  $s_{u,k}[\tau]$  is the unit-power UL symbol of UE  $k$ . The first term contains the desired payload contribution, DL multiuser interference, payload-attached probing leakage, and dedicated sensing leakage. The second term is the DTDD InUI. As in standard single-user reception, the UE treats all non-intended terms as nuisance and decodes only its own data stream.

### 7.6.1 Exact instantaneous DL SINR

Since only the data-bearing part of the intended payload streams is useful for communication, the exact instantaneous downlink SINR of UE  $n$  is

$$\text{SINR}_n^{\text{DL}}[t] = \frac{\underbrace{S_n^{\text{sig}}[t]}_{\text{desired data}}}{\underbrace{I_{n,\text{MU}}^{\text{DL}}[t]}_{\text{DL multiuser}} + \underbrace{I_{n,\text{P}}^{\text{DL}}[t]}_{\text{payload probing}} + \underbrace{I_{n,\text{VS}}^{\text{DL}}[t]}_{\text{dedicated sensing}} + \underbrace{I_{n,\text{UL}}[t]}_{\text{DTDD InUI}} + \underbrace{N_0}_{\text{noise}}}, \quad (7.66)$$

where

$$\begin{aligned}
S_n^{\text{sig}}[t] &\triangleq \left| \sum_{j \in \mathcal{S}_n^{(d)}[t]} \sqrt{P_{d,jn}^{(c)}[t](1 - \rho_{jn}[t])} \mathbf{h}_{jn}^H[t] \mathbf{w}_{jn}^P[t] \right|^2, \\
I_{n,\text{MU}}^{\text{DL}}[t] &\triangleq \sum_{k \in \mathcal{U}_d[t] \setminus \{n\}} \left| \sum_{j \in \mathcal{S}_k^{(d)}[t]} \sqrt{P_{d,jk}^{(c)}[t](1 - \rho_{jk}[t])} \mathbf{h}_{jn}^H[t] \mathbf{w}_{jk}^P[t] \right|^2, \\
I_{n,\text{P}}^{\text{DL}}[t] &\triangleq \sum_{j \in \mathcal{A}_d[t]} \sum_{k \in \mathcal{D}_j[t]} P_{d,jk}^{(c)}[t] \rho_{jk}[t] |\mathbf{h}_{jn}^H[t] \mathbf{w}_{jk}^P[t]|^2, \\
I_{n,\text{VS}}^{\text{DL}}[t] &\triangleq \sum_{j \in \mathcal{A}_d[t]} \sum_{q=1}^{Q_j^{(s)}[t]} P_{j,q}^{(s)}[t] |\mathbf{h}_{jn}^H[t] \mathbf{f}_{j,q}[t]|^2, \\
I_{n,\text{UL}}[t] &\triangleq \sum_{k \in \mathcal{U}_u[t]} P_u^{\text{max}} |g_{kn}[t]|^2.
\end{aligned}$$

The desired term  $S_n^{\text{sig}}[t]$  contains only the coherent data-bearing payload contribution intended for UE  $n$ . The interference terms  $I_{n,\text{MU}}^{\text{DL}}[t]$ ,  $I_{n,\text{P}}^{\text{DL}}[t]$ ,  $I_{n,\text{VS}}^{\text{DL}}[t]$ , and  $I_{n,\text{UL}}[t]$  correspond, respectively, to ordinary DL multiuser interference, payload-attached probing leakage, dedicated-sensing leakage, and DTDD InUI. The DL receiver does not exploit deterministic probing components for communication, so  $I_{n,\text{P}}^{\text{DL}}[t]$  accumulates their nuisance power over the active local payload branches. Probing and dedicated sensing both appear explicitly as nuisance terms at the DL receiver.

With one common pilot phase of length  $\tau_p$  and total slot length  $T_{\text{slot}}$ , the DL SE metric is

$$\text{SE}_n^{\text{DL}}[t] = \left(1 - \frac{\tau_p}{T_{\text{slot}}}\right) \log_2(1 + \text{SINR}_n^{\text{DL}}[t]). \quad (7.67)$$

## 7.7 FRONTHAUL AND COMPUTATIONAL SCALABILITY

The downlink stage remains compatible with the CF-ISAC scalability criterion introduced in Chapter 5, since every online step is bounded by local UE sets and by the CPU-provided candidate set  $\mathcal{V}_j^{\text{CPU}}[t]$ , whose cardinality is at most  $N$ . There is no dense all-AP beamformer solve and no slot-level dependence on all UEs or all requested voxels at each AP.

At slot level, DL AP  $j$  performs pilot-based channel estimation on the bounded set  $\mathcal{D}_j[t]$ , constructs the adaptive protected regularization matrix  $\mathbf{R}_j[t]$ , factorizes it once, evaluates hybrid utilities over  $\mathcal{V}_j^{\text{CPU}}[t]$ , computes one minimum probing fraction per admitted hybrid stream, constructs the served-UE complement projector, and generates dedicated sensing candidates only over the remaining local candidate set. A conservative grouped complexity bound is therefore

$$C_{\text{DL},j}^{\text{slot}}[t] = \mathcal{O}\left(|\mathcal{D}_j[t]|N\tau_p + (|\mathcal{D}_j[t]| + L_{\text{IA}})N^2 + N^3 + |\mathcal{D}_j[t]|V_j^{\text{CPU}}[t]N + |\mathcal{D}_j[t]|^2N + |\mathcal{D}_j[t]|^3 + V_j^{\text{rem}}[t]N^2 + R_j[t]\right), \quad (7.68)$$

where  $V_j^{\text{CPU}}[t] = |\mathcal{V}_j^{\text{CPU}}[t]|$ ,  $V_j^{\text{rem}}[t] = |\mathcal{V}_j^{\text{rem}}[t]|$ , and  $R_j[t]$  is the number of illumination reports returned by AP  $j$ . The per-payload-symbol generation cost is only

$$C_{\text{DL},j}^{\text{sym}}[t] = \mathcal{O}\left(N(|\mathcal{D}_j[t]| + Q_j^{(s)}[t])\right), \quad (7.69)$$

which is simply the weighted superposition of the active payload and dedicated-sensing vectors.

The fronthaul picture is equally favorable. On the CPU $\rightarrow$ DL side, the control plane sends the activation word, the local UE indices, the bounded candidate voxel list with sequence indices, and the scalar tuple  $\{P_{d,jn}[t], \Omega_{jn}^{\text{safe}}[t], \Upsilon_n^{\text{safe}}[t]\}_{n \in \mathcal{D}_j[t]}$ , while the payload plane sends the ordinary UE data symbols. A conservative per-slot count is

$$F_{C \rightarrow D}[t] = M_d[t]b_{\text{ctl}} + b_{\text{idx}} \sum_{j \in \mathcal{A}_d[t]} (|\mathcal{D}_j[t]| + V_j^{\text{CPU}}[t]) + 3b_{\text{sc}}S_d[t] + b_zTS_d[t], \quad (7.70)$$

where  $S_d[t] = \sum_{j \in \mathcal{A}_d[t]} |\mathcal{D}_j[t]|$ ,  $T = T_{\text{slot}} - \tau_p$ , and  $b_{\text{ctl}}$ ,  $b_{\text{idx}}$ ,  $b_{\text{sc}}$ ,  $b_{\text{C}}$ , and  $b_z$  denote the bit budgets for one control word, one index, one real scalar, one complex scalar, and one complex symbol, respectively. Importantly, no extra payload-plane stream is needed for the deterministic probing or dedicated sensing sequences, because the voxel-indexed DFT waveforms are fixed protocol objects generated locally once the slot state and sequence indices are known.

On the return path, DL AP  $j$  uploads the effective-gain scalar  $\Psi_{jn}[t]$  for each served payload direction and the effective illumination report in (7.63) for each

locally illuminated requested voxel. Let  $R_j[t] \triangleq |\mathcal{V}_j^{\text{ill}}[t]|$  be the number of returned illumination scalars. A conservative return-fronthaul count is

$$F_{D \rightarrow C}[t] = b_{\text{sc}} S_d[t] + \sum_{j \in \mathcal{A}_d[t]} R_j[t] (b_{\text{idx}} + b_{\mathbb{C}}), \quad (7.71)$$

where  $b_{\mathbb{C}}$  is the number of bits used to quantize one complex scalar  $\chi_{j,v}[t]$ . Thus the dominant downlink fronthaul term remains the ordinary communication payload distribution  $b_z T S_d[t]$ , while the sensing-specific additions stay on the control side and scale only with the bounded number of locally admitted voxel branches.





# 8

## UPLINK RECEPTION AND COHERENT DETECTION

The previous two chapters completed the transmit side of the proposed protocol. This chapter closes the receiver side of the architecture: it shows how active UL APs receive UE payloads under DTDD nuisance, extract sensing signatures, and run coherent target detection. Two receiver implementations are considered. The first is *distributed*: each UL AP performs local communication combining, voxel receive filtering, and scalar GLRT compression; the CPU fuses only low-dimensional summaries. The second is a *centralized* variant: each active UL AP forwards its antenna-domain payload block to the CPU, which performs communication combining, voxel receive filtering, matched filtering, and centralized GLRT. This mirrors the standard centralized/distributed trade-off in CF MMIMO: centralized processing can exploit richer joint covariance information, but it requires heavier fronthaul and network-wide computation [10, 12].

The chapter follows the receiver operation order. First, the common UL received signal is written for DTDD operation. Second, the distributed UL protocol is defined through an adaptive protected local L-MMSE combiner and CPU soft-symbol fusion. Third, the centralized UL protocol stacks raw antenna-domain blocks at the CPU and applies network-wide communication combining. Fourth, the distributed GLRT and its CFAR threshold are defined. Fifth, the centralized raw-uplink GLRT stacks antenna-domain matched-filter outputs and whitens them with a joint colored nuisance covariance. Finally, the chapter concludes with UL SINR/SE expressions and a fronthaul/computation analysis for the distributed and centralized implementations.

## 8.1 UPLINK RECEIVED SIGNAL UNDER DTDD OPERATION

At slot  $t$ , let  $\mathcal{A}_u[t]$  and  $\mathcal{A}_d[t]$  denote the UL and DL APs, let  $\mathcal{U}_u[t]$  denote the active UL users, and let  $\mathcal{U}_d[t]$  denote the active DL users. As in the previous chapters, one slot coincides with one coherence block of  $T_{\text{slot}}$  channel uses, of which the first  $\tau_p$  symbols are devoted to the common pilot phase and the remaining

$$T \triangleq T_{\text{slot}} - \tau_p \quad (8.1)$$

channel uses form the payload window.

During the payload interval, active UL user  $k \in \mathcal{U}_u[t]$  transmits

$$x_{u,k}[\tau] = \sqrt{P_u^{\max}} s_{u,k}[\tau], \quad \mathbb{E}\{|s_{u,k}[\tau]|^2\} = 1, \quad (8.2)$$

where  $P_u^{\max}$  is the slot-level UL power budget. The received vector at UL AP  $m$  is then

$$\mathbf{r}_{u,m}[\tau] = \sum_{k \in \mathcal{U}_u[t]} \sqrt{P_u^{\max}} \mathbf{h}_{mk}[t] s_{u,k}[\tau] + \sum_{j \in \mathcal{A}_d[t]} \tilde{\mathbf{G}}_{mj}[t] \mathbf{x}_{d,j}[\tau] + \mathbf{w}_{u,m}[\tau], \quad (8.3)$$

where  $\mathbf{h}_{mk}[t] \in \mathbb{C}^N$  is the reciprocal UE-to-AP channel,  $\tilde{\mathbf{G}}_{mj}[t] \in \mathbb{C}^{N \times N}$  is the effective InAI coupling matrix, and  $\mathbf{w}_{u,m}[\tau] \sim \mathcal{CN}(\mathbf{0}, N_0 \mathbf{I}_N)$  is thermal noise.

**TARGET-REFLECTED INTERFERENCE.** The signal model in (8.3) captures the three dominant contributions to the UL received field but omits a fourth physical effect inherent to every DTDD ISAC system: the echo of each DL transmission scattered off the physical targets in the surveillance volume and received at the UL APs, referred to here as Target-Reflected Interference (TRI). A portion of the transmitted energy reaches each physical target  $\ell$  in the surveillance volume, is scattered according to the target's bistatic RCS, and propagates to UL AP  $m$  as a bistatic echo. From the communication receiver's perspective, this echo is an unstructured nuisance. Formally, the expected TRI power at the output of combiner  $\mathbf{q}_{mk}[t]$  is

$$P_{\text{TRI},mk}[t] \triangleq \beta_0 \sum_{\ell} \sum_{j \in \mathcal{A}_d[t]} \frac{\sigma_{\ell}}{d_{j\ell}^2 d_{\ell m}^2} |\mathbf{q}_{mk}^H[t] \mathbf{b}_{m,\ell}[t]|^2 \mathbf{a}_{j,\ell}^H[t] \mathbf{Q}_j[t] \mathbf{a}_{j,\ell}[t], \quad (8.4)$$

where  $\beta_0 = \lambda^2/(4\pi)^3$  is the free-space bistatic propagation constant,  $\sigma_\ell$  is the expected RCS power of target  $\ell$ ,  $d_{j\ell}$  and  $d_{\ell m}$  are the DL-AP-to-target and target-to-UL-AP bistatic path lengths,  $\mathbf{b}_{m,\ell}[t] \in \mathbb{C}^N$  is the unit-norm echo arrival steering vector from target  $\ell$  to AP  $m$ ,  $\mathbf{a}_{j,\ell}[t] \in \mathbb{C}^N$  is the transmit steering vector from DL AP  $j$  toward target  $\ell$ , and  $\mathbf{Q}_j[t] \triangleq \mathbb{E}\{\mathbf{x}_{d,j}[\tau]\mathbf{x}_{d,j}^H[\tau]\}$  is the expected DL transmit covariance of AP  $j$  over the payload window.

Whether TRI must be included in (8.3) depends on its power relative to the other nuisance sources already present in the model. A systematic numerical study was carried out to quantify this ratio under the thesis operating conditions. The results at the thesis baseline RCS of 0 dBsm show that the median TRI power at the combiner output is at most 0.3% of the total interference-plus-noise floor, and the resulting median UL SE loss is 0.1% or less. InUI and thermal noise dominate throughout the sweep. As expected from (8.4), an RCS sweep study reveals that TRI scales linearly with  $\sigma_\ell$  and begins to degrade the median UL SE of more than approximately 5% loss only above  $\text{RCS} \approx 20$  dBsm. This is well outside the UAV RCS range considered in this thesis.

The signal model (8.3) omits TRI on this basis, and the SINR expressions of Section 8.7 do not carry a corresponding TRI denominator term. This simplification is quantitatively justified over the entire operating RCS range of this thesis. Should a future application require surveillance of targets with RCS substantially above 10–20 dBsm, the TRI contribution should be incorporated explicitly into the interference covariance model and treated on the same footing as the residual InAI.

**SIGNAL DECOMPOSITION.** To make the communication and sensing roles more explicit, define the locally served UL-user set of AP  $m$

$$\mathcal{U}_m[t] \triangleq \{k \in \mathcal{U}_u[t] : m \in \mathcal{C}_k\}. \quad (8.5)$$

Then (8.3) can be decomposed as

$$\begin{aligned} \mathbf{r}_{u,m}[\tau] = & \underbrace{\sum_{k \in \mathcal{U}_m[t]} \sqrt{P_u^{\max}} \mathbf{h}_{mk}[t] s_{u,k}[\tau]}_{\text{locally served UL streams}} + \underbrace{\sum_{k \in \mathcal{U}_u[t] \setminus \mathcal{U}_m[t]} \sqrt{P_u^{\max}} \mathbf{h}_{mk}[t] s_{u,k}[\tau]}_{\text{other-user UL interference}} \\ & + \underbrace{\sum_{j \in \mathcal{A}_d[t]} \tilde{\mathbf{G}}_{mj}[t] \mathbf{x}_{d,j}[\tau]}_{\text{InAI}} + \mathbf{w}_{u,m}[\tau]. \quad (8.6) \end{aligned}$$

In the communication branch, the second and third terms act as nuisance for local UL decoding. In the sensing branch, the same terms are absorbed into the equivalent clutter-plus-noise model under the principle of treating unknown UL signals and residual DTDD leakage as interference for sensing processing [15].

## 8.2 DISTRIBUTED UPLINK PROTOCOL

A full network-wide MMSE combiner would require channel-vector sharing and covariance exchange across all active UL APs, while a fully waveform-aware receiver against the time-varying InAI would require detailed knowledge of the current DL transmissions. Both options violate the scalable philosophy. The receiver therefore adopts a protected local L-MMSE combiner: each UL AP uses local channel estimates, a fixed sparse list of dominant DL nuisance directions, and one adaptive scalar that increases only when the active DL→UL threat is actually large. This design is the UL counterpart of the protected precoder of Chapter 7: the directional penalty shape is based on stored nuisance descriptors, while the scalar strength follows the actual active DL threat level.

The nuisance model begins from a slow-timescale set of sparse directional descriptors. For each UL AP  $m$ , let

$$\mathcal{N}_m^{\text{DL}} \triangleq \text{Top}_{L_{\text{IA}}} \{j : \zeta_{mj}^{\text{res}}\} \quad (8.7)$$

denote the bounded list of DL-side neighbors with the strongest residual coupling toward  $m$ . Here  $L_{\text{IA}}$  is a small protocol constant and  $\zeta_{mj}^{\text{res}} \geq 0$  is the residual AP–AP nuisance coefficient after path loss and residual suppression. For each stored

neighbor  $j \in \mathcal{N}_m^{\text{DL}}$ , AP  $m$  stores a unit-norm receive steering proxy  $\mathbf{u}_{mj} \in \mathbb{C}^N$  pointing toward the physical location of AP  $j$ . In slot  $t$ , only the active DL members of this stored list are relevant:

$$\mathcal{B}_m[t] \triangleq \mathcal{A}_d[t] \cap \mathcal{N}_m^{\text{DL}}. \quad (8.8)$$

Instead, the normalization reference is the full stored list,

$$Z_m^{\text{store}} \triangleq \sum_{r \in \mathcal{N}_m^{\text{DL}}} \zeta_{mr}^{\text{res}}, \quad (8.9)$$

and the stored-list weights are

$$\mu_{mj}^{\text{store}} \triangleq \frac{\zeta_{mj}^{\text{res}}}{\max\{Z_m^{\text{store}}, \epsilon_\zeta\}}, \quad j \in \mathcal{N}_m^{\text{DL}}. \quad (8.10)$$

The active directional penalty shape is then

$$\mathbf{P}_m^{\text{IA}}[t] \triangleq \sum_{j \in \mathcal{B}_m[t]} \mu_{mj}^{\text{store}} \mathbf{u}_{mj} \mathbf{u}_{mj}^H. \quad (8.11)$$

Its active mass satisfies

$$\sum_{j \in \mathcal{B}_m[t]} \mu_{mj}^{\text{store}} = \frac{\sum_{j \in \mathcal{B}_m[t]} \zeta_{mj}^{\text{res}}}{\sum_{r \in \mathcal{N}_m^{\text{DL}}} \zeta_{mr}^{\text{res}}} \leq 1, \quad (8.12)$$

up to the numerical floor in (8.10). Thus the receiver does not apply a full-strength directional penalty merely because one weak neighbor happens to be active. If only a small or weak subset of the stored nuisance list is active, the penalty shape itself is small.

The scalar strength is also adaptive. Let the active residual InAI-to-noise threat at UL AP  $m$  be

$$\Gamma_{\text{IA},m}^{\text{UL}}[t] \triangleq \frac{\sum_{j \in \mathcal{B}_m[t]} \zeta_{mj}^{\text{res}} P_{d,j}^{\text{tot}}[t]}{N_0}, \quad (8.13)$$

where  $P_{d,j}^{\text{tot}}[t]$  is the total power radiated by active DL AP  $j$  in the payload window.

The effective receiver-side protection coefficient is

$$\lambda_{\text{IA},m}^{\text{UL,eff}}[t] = \bar{\lambda}_{\text{IA}}^{\text{UL}} \min \left\{ \frac{\Gamma_{\text{IA},m}^{\text{UL}}[t]}{\Gamma_{\text{IA}}^{\text{UL,ref}}}, 1 \right\}, \quad (8.14)$$

with maximum receiver-side strength  $\bar{\lambda}_{\text{IA}}^{\text{UL}}$  and design reference  $\Gamma_{\text{IA}}^{\text{UL,ref}} > 0$ . A practical implementation can use the same reference scale as the DL protected precoder, or a separately tuned receiver-side value.

To keep the covariance scale comparable across loads, AP  $m$  computes the local receive scale

$$s_m^{\text{UL}}[t] \triangleq \frac{1}{N} \text{tr} \left( P_u^{\max} \sum_{\ell \in \mathcal{U}_m[t]} \hat{\mathbf{h}}_{m\ell}[t] \hat{\mathbf{h}}_{m\ell}^H[t] + N_0 \mathbf{I}_N \right), \quad \bar{s}_m^{\text{UL}}[t] \triangleq \max\{s_m^{\text{UL}}[t], s_{\min}^{\text{UL}}\}. \quad (8.15)$$

The directional InAI covariance surrogate used by the receiver is therefore

$$\mathbf{R}_m^{\text{InAI,dir}}[t] \triangleq \lambda_{\text{IA},m}^{\text{UL,eff}}[t] \bar{s}_m^{\text{UL}}[t] \mathbf{P}_m^{\text{IA}}[t]. \quad (8.16)$$

For each locally served user  $k \in \mathcal{U}_m[t]$ , AP  $m$  builds the local interference-plus-noise covariance surrogate

$$\mathbf{R}_{m,k}^{\text{loc,dir}}[t] \triangleq P_u^{\max} \sum_{\ell \in \mathcal{U}_m[t] \setminus \{k\}} \hat{\mathbf{h}}_{m\ell}[t] \hat{\mathbf{h}}_{m\ell}^H[t] + N_0 \mathbf{I}_N + \mathbf{R}_m^{\text{InAI,dir}}[t] + \bar{\nu}_{\text{UL}} \bar{s}_m^{\text{UL}}[t] \mathbf{I}_N + c_{\text{UL}} N_0 \mathbf{I}_N, \quad (8.17)$$

where  $\hat{\mathbf{h}}_{mk}[t]$  is the pilot-based local channel estimate,  $\bar{\nu}_{\text{UL}}$  is a small scale-normalized diagonal loading coefficient, and  $c_{\text{UL}} > 0$  is a small noise-referenced loading coefficient. The signal-proportional term  $\bar{\nu}_{\text{UL}} \bar{s}_m^{\text{UL}}[t] \mathbf{I}_N$  scales with the local traffic and channel energy, so it under-regularizes lightly loaded or geometrically weak APs while over-loading strong ones; the complementary flat term  $c_{\text{UL}} N_0 \mathbf{I}_N$  provides a signal-independent regularization floor that robustifies the per-AP matrix inversion against channel-estimation error without penalizing strong APs, thereby improving the quality of the soft symbols forwarded to the CPU for fusion. The local combiner is

$$\mathbf{v}_{mk}[t] = (\mathbf{R}_{m,k}^{\text{loc,dir}}[t])^{-1} \hat{\mathbf{h}}_{mk}[t], \quad (8.18)$$

and the resulting local soft symbol is

$$z_{mk}[\tau] \triangleq \mathbf{v}_{mk}^H[t] \mathbf{r}_{u,m}[\tau]. \quad (8.19)$$

If the active residual InAI is mild, then  $\lambda_{\text{IA},m}^{\text{UL,eff}}[t]$  and the active mass of  $\mathbf{P}_m^{\text{IA}}[t]$  are small, so (8.18) is close to an ordinary local L-MMSE combiner. If strong neighboring DL transmitters are active, the combiner adds a low-rank directional covariance term and suppresses those arrival directions. The post-combining effective gain

$$\tilde{g}_{mk}[t] \triangleq |\mathbf{v}_{mk}^H[t] \hat{\mathbf{h}}_{mk}[t]|^2, \quad (8.20)$$

will later be used by the CPU to weight the soft-symbol fusion.

### 8.2.1 Protection trade-offs and applicability

The protected local L-MMSE combiner of (8.18) is the receive-side complement of the protected DL precoder developed in Section 7.2, and it is governed by the same fundamental trade-off between interference suppression and useful signal gain.

A standard local L-MMSE combiner includes only the multi-user interference covariance and the thermal noise floor in its matrix inversion: the protected design adds the term  $\mathbf{R}_m^{\text{InAI,dir}}[t]$  in (8.16) to the covariance model, placing an explicit directional cost on receive energy in the directions of the dominant active DL neighbors. The result is that the receive beam is actively steered away from those spatial signatures, suppressing the residual InAI rather than absorbing it into a larger noise term. As on the transmit side, the scalar  $\lambda_{\text{IA},m}^{\text{UL,eff}}[t]$  in (8.14) gates the penalty strength according to the actual active residual threat, so the modification is applied only when the leakage level warrants it.

The receiver-side protection carries the same categories of cost, such as the combining gain loss, the required inter-AP knowledge, and the computational overhead; and the same optimal scenarios discussed for the transmitter side in Section 7.2.2.

Together, the transmitter-side protected precoder and the receiver-side protected combiner form a complementary pair: the precoder reduces the interference power generated toward vulnerable neighbors, and the combiner suppresses whatever residual leakage escapes that transmit-side reduction.

### 8.2.2 CPU soft fusion for uplink communication

After local combining, UL APs forward low-dimensional summaries to the CPU, which performs a second-stage soft fusion. For slot  $t$ , AP  $m \in \mathcal{A}_u[t]$  uploads two objects:

$$\{\tilde{g}_{mk}[t]\}_{k \in \mathcal{U}_m[t]} \quad \text{and} \quad \{z_{mk}[\tau]\}_{k \in \mathcal{U}_m[t], \tau = \tau_p + 1, \dots, T_{\text{slot}}}. \quad (8.21)$$

For user  $k \in \mathcal{U}_u[t]$ , define the active serving-AP set

$$\mathcal{A}_k^u[t] \triangleq \mathcal{C}_k \cap \mathcal{A}_u[t]. \quad (8.22)$$

The CPU forms the fused soft symbol

$$\hat{s}_{u,k}[\tau] = \sum_{m \in \mathcal{A}_k^u[t]} a_{mk}[t] z_{mk}[\tau], \quad (8.23)$$

with normalized effective-gain weights

$$a_{mk}[t] \triangleq \frac{\sqrt{\tilde{g}_{mk}[t]}}{\sum_{i \in \mathcal{A}_k^u[t]} \sqrt{\tilde{g}_{ik}[t]}}, \quad m \in \mathcal{A}_k^u[t]. \quad (8.24)$$

For later SINR analysis in Section 8.7, it is convenient to define the stacked network combiner. Let  $\mathcal{A}_u[t] = \{m_1, \dots, m_{M_u[t]}\}$  and set

$$\mathbf{q}_{mk}[t] \triangleq \begin{cases} a_{mk}[t] \mathbf{v}_{mk}[t], & m \in \mathcal{A}_k^u[t], \\ \mathbf{0}_N, & m \notin \mathcal{A}_k^u[t], \end{cases} \quad (8.25)$$

so that the stacked fused combiner becomes

$$\mathbf{q}_k[t] \triangleq [\mathbf{q}_{m_1 k}^T[t], \dots, \mathbf{q}_{m_{M_u[t]} k}^T[t]]^T. \quad (8.26)$$

If  $\mathbf{r}_u[\tau]$  denotes the network-stacked received vector formed by stacking  $\{\mathbf{r}_{u,m}[\tau]\}_{m \in \mathcal{A}_u[t]}$ , then (8.23) can be written compactly as

$$\hat{s}_{u,k}[\tau] = \mathbf{q}_k^H[t] \mathbf{r}_u[\tau]. \quad (8.27)$$

### 8.3 CENTRALIZED UPLINK PROTOCOL

This thesis also proposes a centralized UL implementation in which the active receiving APs do not compute local communication combiners or local sensing statistics. Instead, each UL AP forwards its antenna-domain received block to the CPU, which performs all subsequent baseband processing.

Let the active UL set be ordered as  $\mathcal{A}_u[t] = \{m_1, \dots, m_{M_u[t]}\}$ . For each payload symbol, define the stacked received vector

$$\mathbf{r}_u^c[\tau] \triangleq [\mathbf{r}_{u,m_1}^T[\tau], \dots, \mathbf{r}_{u,m_{M_u[t]}}^T[\tau]]^T \in \mathbb{C}^{NM_u[t]}. \quad (8.28)$$



Likewise, define the stacked user channel and stacked residual InAI channel as

$$\mathbf{h}_k^c[t] \triangleq [\mathbf{h}_{m_1 k}^T[t], \dots, \mathbf{h}_{m_{M_u[t]} k}^T[t]]^T, \quad \tilde{\mathbf{G}}_j^c[t] \triangleq [\tilde{\mathbf{G}}_{m_1 j}^T[t], \dots, \tilde{\mathbf{G}}_{m_{M_u[t]} j}^T[t]]^T. \quad (8.29)$$

Stacking (8.3) gives

$$\mathbf{r}_u^c[\tau] = \sum_{k \in \mathcal{U}_u[t]} \sqrt{P_u^{\max}} \mathbf{h}_k^c[t] s_{u,k}[\tau] + \sum_{j \in \mathcal{A}_d[t]} \tilde{\mathbf{G}}_j^c[t] \mathbf{x}_{d,j}[\tau] + \mathbf{w}_u^c[\tau], \quad (8.30)$$

where  $\mathbf{w}_u^c[\tau] \sim \mathcal{CN}(\mathbf{0}, N_0 \mathbf{I}_{NM_u[t]})$ . The raw payload block available at the CPU is

$$\mathbf{Y}_u^c[t] \triangleq [\mathbf{r}_u^c[\tau_p + 1], \dots, \mathbf{r}_u^c[T_{\text{slot}}]] \in \mathbb{C}^{NM_u[t] \times T}. \quad (8.31)$$

If centralized channel estimation is also assumed, the despread pilot samples are transported in the same way and the CPU computes the channel estimates from the stacked pilot observations. In the notation below, the resulting estimates are denoted by  $\hat{\mathbf{h}}_k^c[t]$ .

The centralized communication combiner is defined from a network-wide covariance surrogate. Let

$$\mathbf{C}_d^c[t] \triangleq \text{blkdiag}(\mathbf{C}_{d,m_1}[t], \dots, \mathbf{C}_{d,m_{M_u[t]}}[t]), \quad (8.32)$$

where the per-AP blocks  $\mathbf{C}_{d,m}[t]$  are later defined in (8.95). A centralized protected L-MMSE covariance can then be written as

$$\mathbf{R}_{\text{UL}}^c[t] \triangleq P_u^{\max} \sum_{\ell \in \mathcal{U}_u[t]} \hat{\mathbf{h}}_\ell^c[t] (\hat{\mathbf{h}}_\ell^c[t])^H + \mathbf{C}_d^c[t] + N_0 \mathbf{I}_{NM_u[t]} + \nu_c \bar{s}_c^{\text{UL}}[t] \mathbf{I}_{NM_u[t]}, \quad (8.33)$$

where  $\nu_c$  is a small diagonal-loading coefficient and

$$\bar{s}_c^{\text{UL}}[t] \triangleq \max \left\{ \frac{1}{NM_u[t]} \text{tr} \left( P_u^{\max} \sum_{\ell \in \mathcal{U}_u[t]} \hat{\mathbf{h}}_\ell^c[t] (\hat{\mathbf{h}}_\ell^c[t])^H + \mathbf{C}_d^c[t] + N_0 \mathbf{I}_{NM_u[t]} \right), s_{\min}^c \right\} \quad (8.34)$$

is the centralized receive scale. The corresponding network-wide combiner for user  $k$  is

$$\mathbf{q}_k^c[t] = (\mathbf{R}_{\text{UL}}^c[t])^{-1} \hat{\mathbf{h}}_k^c[t], \quad (8.35)$$

and the centralized soft symbol is

$$\hat{s}_{u,k}^c[\tau] = (\mathbf{q}_k^c[t])^H \mathbf{r}_u^c[\tau]. \quad (8.36)$$

The centralized combiner can suppress InUI and residual InAI using a joint covariance, whereas the distributed path only exchanges scalar soft-symbol summaries after local combining.

The same raw block in (8.31) is also used by the centralized sensing processor. Therefore communication and sensing share the same uploaded antenna-domain samples, and the difference lies only in the CPU processing applied to them.

## 8.4 VOXEL OBSERVATIONS

The scheduling layer assigns to each active UL AP  $m$  a voxel-processing list  $\mathcal{V}_m^{\text{sens}}[t]$ . During the payload interval, the antenna-domain block at UL AP  $m$  is

$$\mathbf{Y}_m[t] \triangleq [\mathbf{r}_{u,m}[\tau_p + 1], \dots, \mathbf{r}_{u,m}[T_{\text{slot}}]] \in \mathbb{C}^{N \times T}. \quad (8.37)$$

In the distributed UL protocol, this block is cached and processed locally at AP  $m$ . In the centralized UL protocol, the same block is uploaded to the CPU and becomes one sub-block of  $\mathbf{Y}_u^c[t]$  in (8.31). The mathematical sensing signature is common to both implementations; only the place where the receive projection, matched filtering, and likelihood maximization are performed changes.

For each  $v \in \mathcal{V}_m^{\text{sens}}[t]$ , let  $\mathbf{b}_{m,v}[t] \in \mathbb{C}^N$  denote the unit-norm arrival steering vector from voxel  $v$  to AP  $m$ . The distributed sensing branch uses the same adaptive directional protection principle as the communication combiner. Specifically, AP  $m$  builds the shared sensing covariance surrogate

$$\mathbf{R}_m^{\text{sens,dir}}[t] \triangleq P_u^{\max} \sum_{\ell \in \mathcal{U}_m[t]} \hat{\mathbf{h}}_{m\ell}[t] \hat{\mathbf{h}}_{m\ell}^H[t] + N_0 \mathbf{I}_N + \mathbf{R}_m^{\text{InAI,dir}}[t] + \bar{\nu}_s \bar{s}_m^{\text{UL}}[t] \mathbf{I}_N, \quad (8.38)$$

where  $\bar{\nu}_s$  is a small sensing-side diagonal loading coefficient. The inverse of (8.38) is computed once per slot and reused for all processed voxel hypotheses. The voxel-dependent protected receive beam is

$$\tilde{\mathbf{u}}_{m,v}[t] = (\mathbf{R}_m^{\text{sens,dir}}[t])^{-1} \mathbf{b}_{m,v}[t], \quad \mathbf{u}_{m,v}[t] = \frac{\tilde{\mathbf{u}}_{m,v}[t]}{\|\tilde{\mathbf{u}}_{m,v}[t]\|}, \quad (8.39)$$

so that  $\|\mathbf{u}_{m,v}[t]\| = 1$ . The corresponding beamformed scalar stream is

$$y_{m,v}[\tau] \triangleq \mathbf{u}_{m,v}^H[t] \mathbf{r}_{u,m}[\tau], \quad \tau = \tau_p + 1, \dots, T_{\text{slot}}. \quad (8.40)$$

In the centralized uplink implementation, (8.39) and (8.40) are evaluated at the CPU from the uploaded  $\mathbf{Y}_m[t]$ .

For a fixed voxel  $v$ , the final deterministic illumination is obtained from the effective scalar reports returned by the transmitting DL APs in (7.63). Specifically, for every DL AP  $j$ , the CPU knows the complex scalar  $\chi_{j,v}[t]$  in (7.61), with  $\chi_{j,v}[t] = 0$  if AP  $j$  did not transmit a deterministic branch toward voxel  $v$ . The informative branch set can therefore be represented by the reported nonzero scalars

$$\mathcal{B}_v[t] \triangleq \{j \in \mathcal{A}_d[t] : \chi_{j,v}[t] \neq 0\}, \quad (8.41)$$

and the informative voxel set is

$$\mathcal{V}_{\text{inf}}[t] \triangleq \{v : \mathcal{B}_v[t] \neq \emptyset\}. \quad (8.42)$$

When  $v \in \mathcal{V}_{\text{inf}}[t]$ , all deterministic branches contributing to  $\{\chi_{j,v}[t]\}_{j \in \mathcal{B}_v[t]}$  use the same voxel sequence  $d_{s_v[t]}[\tau]$  assigned in Chapter 6.

The scalar deterministic signature used by the distributed detector is then

$$g_{m,v}[t] \triangleq \sqrt{T} \mathbf{u}_{m,v}^H[t] \mathbf{b}_{m,v}[t] \sum_{j \in \mathcal{B}_v[t]} \chi_{j,v}[t]. \quad (8.43)$$

Equivalently, for the centralized raw-antenna detector, define the antenna-domain signature vector

$$\mathbf{s}_{m,v}^a[t] \triangleq \sqrt{T} \mathbf{b}_{m,v}[t] \sum_{j \in \mathcal{B}_v[t]} \chi_{j,v}[t] \in \mathbb{C}^N, \quad (8.44)$$

so that  $g_{m,v}[t] = \mathbf{u}_{m,v}^H[t] \mathbf{s}_{m,v}^a[t]$ . Equations (8.43) and (8.44) contain only known quantities: the assigned sequence length, the receive steering/projection, and the effective DL illumination scalars reported by the transmitting APs. If  $\mathcal{B}_v[t] = \emptyset$ , then  $g_{m,v}[t] = 0$  and  $\mathbf{s}_{m,v}^a[t] = \mathbf{0}_N$ , so voxel  $v$  is not informative in that slot.

The distributed voxel matched filter is formed for every  $v \in \mathcal{V}_m^{\text{sens}}[t]$  as

$$y_{m,v}^{\text{MF}}[t] \triangleq \frac{1}{\sqrt{T}} \sum_{\tau=\tau_p+1}^{T_{\text{slot}}} d_{s_v[t]}^*[\tau] y_{m,v}[\tau] = \frac{1}{\sqrt{T}} \mathbf{u}_{m,v}^H[t] \mathbf{Y}_m[t] \mathbf{d}_{s_v[t]}^*, \quad (8.45)$$

where  $\mathbf{d}_{s_v[t]} \in \mathbb{C}^T$  collects the deterministic sequence samples over the payload window. For the centralized detector, the CPU keeps the antenna-domain matched-filter output before local scalar projection,

$$\mathbf{y}_{m,v}^a[t] \triangleq \frac{1}{\sqrt{T}} \mathbf{Y}_m[t] \mathbf{d}_{s_v[t]}^* \in \mathbb{C}^N. \quad (8.46)$$

The scalar observation in (8.45) is recovered from (8.46) as

$$y_{m,v}^{\text{MF}}[t] = \mathbf{u}_{m,v}^H[t] \mathbf{y}_{m,v}^a[t]. \quad (8.47)$$

If another distant voxel reuses the same sequence index, its residual contribution is treated as part of the effective nuisance. The anchor-cell reuse rule is designed so that such leakage is spatially separated from voxel  $v$  and is therefore much weaker than same-voxel coherent illumination.

#### 8.4.1 Rank-one coherent voxel model

The matched-filtered sensing sample (8.45) admits a simple and operationally meaningful local model if one adopts the dominant-target approximation already implicit in the voxelized design. The interpretation is that, within one scanned voxel and one slot, the coherent return observed at receiver  $m$  is represented by one unknown complex echo amplitude multiplying the known deterministic signature in (8.43). Accordingly, the voxel model at AP  $m$  is written as the binary hypothesis test

$$\mathcal{H}_0 : y_{m,v}^{\text{MF}}[t] = n_{m,v}[t], \quad (8.48)$$

$$\mathcal{H}_1 : y_{m,v}^{\text{MF}}[t] = \gamma_{m,v}[t] g_{m,v}[t] + n_{m,v}[t], \quad (8.49)$$

where  $\gamma_{m,v}[t] \in \mathbb{C}$  is unknown and  $n_{m,v}[t]$  is the scalar effective nuisance, modeled as

$$n_{m,v}[t] \sim \mathcal{CN}(0, \sigma_{m,v}^2[t]). \quad (8.50)$$

The coefficient  $\gamma_{m,v}[t]$  absorbs the true bistatic propagation, target RCS, target orientation, residual phase mismatch, and any difference between the geometric design model and the actual echo.

The scalar variance is built from two components. The first is a slow-timescale background map  $\sigma_{m,v,\text{bg}}^2$  that captures static clutter and the receiver noise floor for direction  $(m, v)$ . The second is the directional InAI contribution induced by the active DL set,

$$\sigma_{m,v}^2[t] \triangleq \sigma_{m,v,\text{bg}}^2 + \mathbf{u}_{m,v}^H[t] \mathbf{R}_m^{\text{InAI,dir}}[t] \mathbf{u}_{m,v}[t]. \quad (8.51)$$

Equation (8.51) corrects the directionality of the dominant residual InAI without requiring voxel-resolved knowledge of the entire DL waveform, while the unknown UL payload leakage remains absorbed into the equivalent sensing nuisance under the treat-unknown-signals-as-interference principle [15]. For performance analysis only, the conditional local coherent SCNR is

$$\rho_{m,v}[t] \triangleq \frac{|\gamma_{m,v}[t]|^2 |g_{m,v}[t]|^2}{\sigma_{m,v}^2[t]}. \quad (8.52)$$

This quantity affects the conditional detection probability under  $\mathcal{H}_1$ , but it is not needed to set the CFAR threshold.

## 8.5 DISTRIBUTED GLRT

The distributed detector keeps the antenna-domain and covariance-aware processing at the APs: each receiving AP evaluates a local likelihood ratio for every assigned informative voxel and forwards only the resulting scalar statistic to the CPU. This is the same architectural trade-off adopted in scalable CF reception, where local processing is preferred whenever the information loss is small compared with the fronthaul and computation saved by avoiding fully centralized baseband transport [12, 15].

The local GLRT is evaluated only for assigned voxel–receiver pairs with a nonzero deterministic signature, i.e.,  $v \in \mathcal{V}_m^{\text{sens}}[t]$  and  $|g_{m,v}[t]| > 0$ . Under (8.48)–(8.49), local detection is a Gaussian GLRT with unknown complex amplitude  $\gamma_{m,v}[t]$ . Maximizing the likelihood under  $\mathcal{H}_1$  gives

$$\hat{\gamma}_{m,v}[t] = \frac{g_{m,v}^*[t] y_{m,v}^{\text{MF}}[t]}{|g_{m,v}[t]|^2}, \quad g_{m,v}[t] \neq 0. \quad (8.53)$$

Substitution into the likelihood ratio yields

$$\Lambda_{m,v}[t] = \frac{2|g_{m,v}^*[t]y_{m,v}^{\text{MF}}[t]|^2}{\sigma_{m,v}^2[t]|g_{m,v}[t]|^2} = \frac{2|y_{m,v}^{\text{MF}}[t]|^2}{\sigma_{m,v}^2[t]}, \quad g_{m,v}[t] \neq 0. \quad (8.54)$$

Thus, in the distributed scalar GLRT, the magnitude and phase of  $g_{m,v}[t]$  cancel from the statistic whenever the signature is nonzero.

Each active UL AP sends one scalar  $\Lambda_{m,v}[t]$  for every assigned voxel  $v \in \mathcal{V}_m^{\text{sens}}[t]$  that is retained for distributed sensing. The scheduled receiver set for voxel  $v$  is inherited from Chapter 6,

$$\mathcal{M}_v[t] = \{m \in \mathcal{A}_u[t] : v \in \mathcal{V}_m^{\text{sens}}[t]\}. \quad (8.55)$$

The contributing receiver set used by the detector is computed at the CPU from the known receive-side geometry and from the effective illumination scalars  $\{\chi_{j,v}[t]\}$  reported by the DL APs:

$$\mathcal{M}_v^{\text{det}}[t] \triangleq \{m \in \mathcal{M}_v[t] : |g_{m,v}[t]|^2 > 0\}, \quad M_v^{\text{det}}[t] \triangleq |\mathcal{M}_v^{\text{det}}[t]|, \quad (8.56)$$

If  $M_v^{\text{det}}[t] = 0$ , voxel  $v$  is not informative in slot  $t$  and no detector threshold is evaluated for it. Otherwise, the CPU fuses the distributed coherent statistics by simple summation,

$$\Lambda_v[t] \triangleq \sum_{m \in \mathcal{M}_v^{\text{det}}[t]} \frac{2|y_{m,v}^{\text{MF}}[t]|^2}{\sigma_{m,v}^2[t]}. \quad (8.57)$$

The decision rule is

$$\Lambda_v[t] \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta_{\text{det},v}[t]. \quad (8.58)$$

Figure 8.1 illustrates the two-branch distributed receiver architecture at one UL AP, where the same received vector  $\mathbf{r}_{u,m}[\tau]$  is processed in parallel for communication and coherent sensing, and the local statistics  $\{\Lambda_{m,v}[t]\}$  are uploaded to the CPU for distributed voxel detection.

### 8.5.1 Detection threshold and sensing metrics

The local statistic in (8.54) admits exact null and conditional alternative laws. Under  $\mathcal{H}_0$ ,

$$\Lambda_{m,v}[t] = \frac{2|n_{m,v}[t]|^2}{\sigma_{m,v}^2[t]} \sim \chi_2^2, \quad (8.59)$$

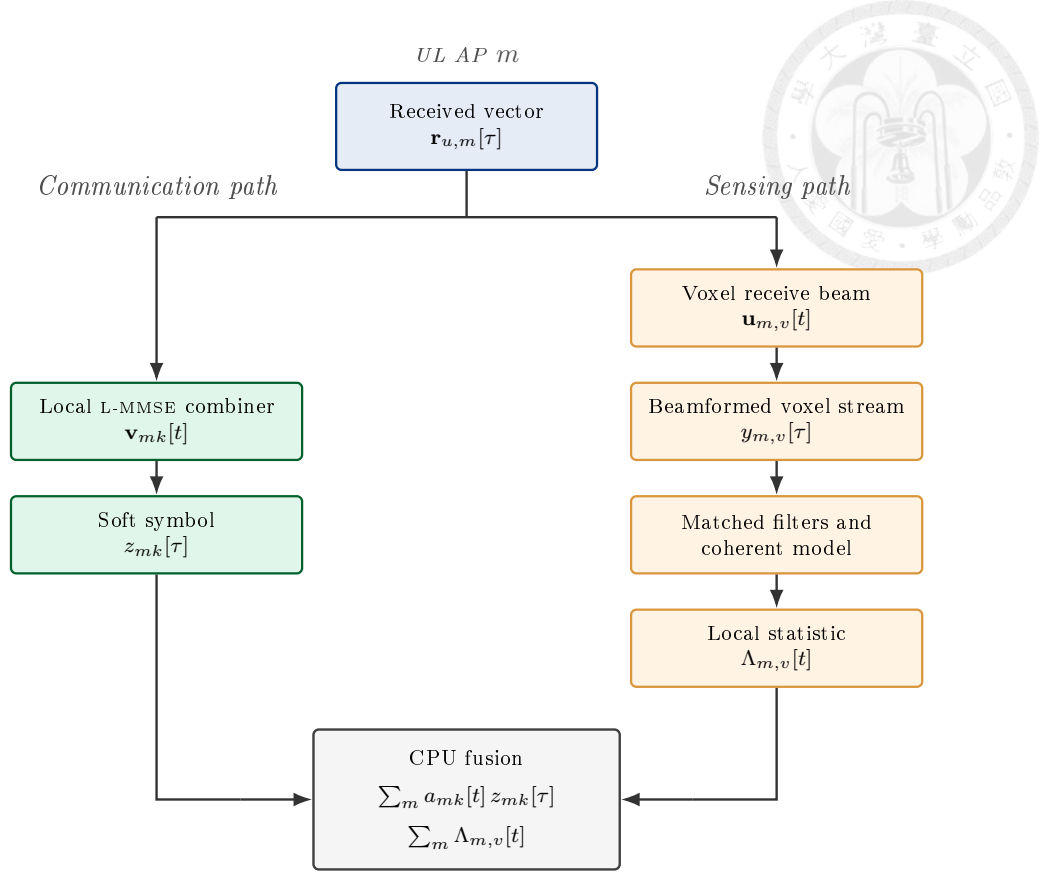


Figure 8.1: Two-branch distributed receiver architecture at UL AP. The same received vector  $\mathbf{r}_{u,m}[\tau]$  is processed in parallel for communication and coherent sensing. The communication branch applies the adaptive protected local L-MMSE combiner to generate soft symbols  $z_{mk}[\tau]$ , which are forwarded to the CPU and fused for payload decoding. The sensing branch applies a voxel-dependent receive beam, forms beamformed sensing streams, and computes the local coherent GLRT statistic  $\Lambda_{m,v}[t]$ , which is uploaded to the CPU for distributed voxel detection.

where  $\chi_2^2$  denotes the chi-square distribution with two DoF. Under  $\mathcal{H}_1$ , conditioned on  $\gamma_{m,v}[t]$  and the known signature  $g_{m,v}[t]$ ,

$$\Lambda_{m,v}[t] \sim \chi_2'^2(2\rho_{m,v}[t]), \quad (8.60)$$

with  $\rho_{m,v}[t]$  defined in (8.52). These laws make explicit the separation between thresholding and target strength: the central law under  $\mathcal{H}_0$  determines the CFAR threshold, while  $\gamma_{m,v}[t]$  and the physical target strength appear only in the noncentrality under  $\mathcal{H}_1$  [23, 26].

Since the local statistics are independent across the contributing UL APs under the adopted distributed model, the fused statistic in (8.57) satisfies

$$\mathcal{H}_0 : \Lambda_v[t] \sim \chi_{2M_v^{\text{det}}[t]}^2, \quad (8.61)$$

$$\mathcal{H}_1 : \Lambda_v[t] \sim \chi_{2M_v^{\text{det}}[t]}^2(2\rho_v^{\text{eff}}[t]), \quad (8.62)$$

with conditional fused coherent SCNR  $\rho_v^{\text{eff}}[t] \triangleq \sum_{m \in \mathcal{M}_v^{\text{det}}[t]} \rho_{m,v}[t]$ . The exact voxel-wise  $P_{\text{FA}}$  is therefore

$$P_{\text{FA},v}[t] = \Pr(\Lambda_v[t] > \eta_{\text{det},v}[t] \mid \mathcal{H}_0) = Q\left(M_v^{\text{det}}[t], \frac{\eta_{\text{det},v}[t]}{2}\right), \quad (8.63)$$

where  $Q(a, x) = \Gamma(a, x)/\Gamma(a)$  is the regularized upper incomplete Gamma function.

Thus the CFAR threshold for target false-alarm probability  $P_{\text{FA},v}^*$  is

$$\eta_{\text{det},v}[t] = 2Q^{-1}(M_v^{\text{det}}[t], P_{\text{FA},v}^*). \quad (8.64)$$

The corresponding conditional detection probability is

$$P_{\text{D},v}[t] = 1 - F_{\chi_{2M_v^{\text{det}}[t]}^2(2\rho_v^{\text{eff}}[t])}(\eta_{\text{det},v}[t]), \quad (8.65)$$

where  $F_{\chi_{\nu}^2(\lambda)}(\cdot)$  denotes the noncentral chi-square cumulative distribution function.

## 8.6 CENTRALIZED GLRT

In the centralized approach, the UL APs upload their antenna-domain payload blocks and the CPU forms the matched-filtered antenna vectors in (8.46). Receive projection, whitening, and likelihood maximization are all performed after the CPU has access to the raw UL samples. The model follows the centralized GLRT philosophy used in DTDD CF-ISAC, where the CPU stacks the received samples and accounts for colored sensing nuisance induced by UL users and residual InAI [15].

Let  $\mathcal{M}_v^{\text{det}}[t] = \{m_1, \dots, m_{M_v^{\text{det}}[t]}\}$  be the contributing receiver set already defined in (8.56). The CPU stacks the antenna-domain matched-filtered observations as

$$\mathbf{y}_v^{\text{a}}[t] \triangleq [(\mathbf{y}_{m_1,v}^{\text{a}}[t])^T, \dots, (\mathbf{y}_{m_{M_v^{\text{det}}[t]},v}^{\text{a}}[t])^T]^T \in \mathbb{C}^{NM_v^{\text{det}}[t]}. \quad (8.66)$$

Similarly, define the antenna-domain signature matrix

$$\mathbf{S}_v^a[t] \triangleq \text{blkdiag}(\mathbf{s}_{m_1,v}^a[t], \dots, \mathbf{s}_{m_{M_v^{\text{det}}[t]},v}^a[t]) \in \mathbb{C}^{NM_v^{\text{det}}[t] \times M_v^{\text{det}}[t]}, \quad (8.67)$$

and the unknown link-amplitude vector

$$\boldsymbol{\gamma}_v[t] \triangleq [\gamma_{m_1,v}[t], \dots, \gamma_{m_{M_v^{\text{det}}[t]},v}[t]]^T. \quad (8.68)$$

The centralized antenna-domain voxel hypothesis is

$$\mathcal{H}_0 : \mathbf{y}_v^a[t] = \mathbf{n}_v^a[t], \quad (8.69)$$

$$\mathcal{H}_1 : \mathbf{y}_v^a[t] = \mathbf{S}_v^a[t]\boldsymbol{\gamma}_v[t] + \mathbf{n}_v^a[t], \quad (8.70)$$

with

$$\mathbf{n}_v^a[t] \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_v^a[t]). \quad (8.71)$$

The covariance  $\boldsymbol{\Sigma}_v^a[t]$  is the antenna-domain nuisance covariance after sequence matching. Its block diagonals contain receiver noise, residual UL payload leakage, residual clutter, and directional InAI; its off-diagonal blocks represent correlation across UL APs induced by common UL payload symbols, common DL leakage, or correlated clutter. When these off-diagonal blocks are negligible and the covariance is projected locally by  $\{\mathbf{u}_{m,v}[t]\}$ , the centralized model reduces to the distributed scalar model.

The centralized likelihood must use a positive-definite covariance. In a practical implementation this point is important because the empirical cross-receiver nuisance covariance is assembled from a finite number of active UL UE streams and residual DL-originated leakage terms. Such a covariance has the generic form

$$\hat{\boldsymbol{\Sigma}}_v^a[t] = \sum_{q=1}^{Q_v[t]} \mathbf{c}_{q,v}[t] \mathbf{c}_{q,v}^H[t] + \boldsymbol{\Sigma}_{v,\text{floor}}^a[t], \quad (8.72)$$

where  $\mathbf{c}_{q,v}[t]$  denotes one projected nuisance contribution and  $\boldsymbol{\Sigma}_{v,\text{floor}}^a[t]$  denotes the residual receiver-noise and clutter floor. If the empirical part is used without a sufficiently representative floor, then

$$\text{rank} \left( \sum_{q=1}^{Q_v[t]} \mathbf{c}_{q,v}[t] \mathbf{c}_{q,v}^H[t] \right) \leq Q_v[t], \quad (8.73)$$

so near-null eigenmodes can appear whenever the detector dimension is larger than the number of effective nuisance sources. Since the centralized GLRT whitens the observation with  $\Sigma_v^{-1}$ , these near-null modes would be assigned unrealistically large weight. Real receivers always retain an uncorrelated thermal-plus-clutter component, and classical adaptive radar detectors are known to be sensitive to covariance estimation errors when the covariance is estimated from limited data [51].

To make this floor explicit, the centralized detector applies a small diagonal loading to the correlation matrix associated with  $\Sigma_v^a[t]$ . Let

$$\mathbf{D}_v^a[t] \triangleq \text{diag}\left(\sqrt{[\Sigma_v^a[t]]_{1,1}}, \dots, \sqrt{[\Sigma_v^a[t]]_{NM_v^{\text{det}}[t], NM_v^{\text{det}}[t]}}\right), \quad (8.74)$$

and define the empirical correlation matrix

$$\mathbf{R}_v^a[t] \triangleq (\mathbf{D}_v^a[t])^{-1} \Sigma_v^a[t] (\mathbf{D}_v^a[t])^{-1}. \quad (8.75)$$

For a loading coefficient  $\alpha_c \in [0, 1)$ , the regularized correlation matrix is

$$\mathbf{R}_{v,\alpha}^a[t] \triangleq (1 - \alpha_c) \mathbf{R}_v^a[t] + \alpha_c \mathbf{I}_{NM_v^{\text{det}}[t]}, \quad (8.76)$$

and the covariance used by the centralized GLRT is

$$\Sigma_{v,\alpha}^a[t] \triangleq \mathbf{D}_v^a[t] \mathbf{R}_{v,\alpha}^a[t] \mathbf{D}_v^a[t]. \quad (8.77)$$

Equivalently, (8.77) shrinks the empirical covariance toward its diagonal part,

$$\Sigma_{v,\alpha}^a[t] = (1 - \alpha_c) \Sigma_v^a[t] + \alpha_c \text{diag}(\Sigma_v^a[t]). \quad (8.78)$$

This form is scale-invariant across voxels because the loading is applied after normalizing by the local marginal variances. It also preserves the per-antenna nuisance powers, since the diagonal of  $\Sigma_{v,\alpha}^a[t]$  is identical to the diagonal of  $\Sigma_v^a[t]$ . If  $\mathbf{R}_v^a[t] \succeq \mathbf{0}$ , then  $\lambda_{\min}(\mathbf{R}_{v,\alpha}^a[t]) \geq \alpha_c$ ; therefore the whitening operation cannot exploit a zero-noise direction.

The same regularization is applied after the receive-beam projection, where the CPU uses the scalar stack  $\mathbf{y}_v^s[t] = \mathbf{U}_v^H[t] \mathbf{y}_v^a[t]$ . Writing

$$\Sigma_v^s[t] \triangleq \mathbf{U}_v^H[t] \Sigma_v^a[t] \mathbf{U}_v[t], \quad \mathbf{D}_v^s[t] \triangleq \text{diag}\left(\sigma_{m_{1,v}}[t], \dots, \sigma_{m_{M_v^{\text{det}}[t],v}}[t]\right), \quad (8.79)$$

the scalar correlation matrix is  $\boldsymbol{\rho}_v[t] = (\mathbf{D}_v^s[t])^{-1} \boldsymbol{\Sigma}_v^s[t] (\mathbf{D}_v^s[t])^{-1}$  and the loaded covariance is

$$\boldsymbol{\Sigma}_{v,\alpha}^s[t] = \mathbf{D}_v^s[t] \left( (1 - \alpha_c) \boldsymbol{\rho}_v[t] + \alpha_c \mathbf{I}_{M_v^{\text{det}}[t]} \right) \mathbf{D}_v^s[t]. \quad (8.80)$$

From this point onward, the centralized detector treats  $\boldsymbol{\Sigma}_{v,\alpha}^a[t]$  as the nuisance covariance model used for whitening and threshold evaluation. The centralized GLRT maximizes the likelihood under  $\mathcal{H}_1$  over  $\boldsymbol{\gamma}_v[t]$ . Using standard Gaussian linear-model GLRT theory [25, 26], the weighted least-squares estimate is

$$\hat{\boldsymbol{\gamma}}_v^a[t] = \left( (\mathbf{S}_v^a[t])^H (\boldsymbol{\Sigma}_{v,\alpha}^a[t])^{-1} \mathbf{S}_v^a[t] \right)^\dagger (\mathbf{S}_v^a[t])^H (\boldsymbol{\Sigma}_{v,\alpha}^a[t])^{-1} \mathbf{y}_v^a[t], \quad (8.81)$$

where  $(\cdot)^\dagger$  denotes the Moore–Penrose pseudoinverse. Substitution into the log-likelihood ratio gives the centralized raw-uplink statistic

$$\Lambda_v^a[t] = 2 (\mathbf{y}_v^a[t])^H (\boldsymbol{\Sigma}_{v,\alpha}^a[t])^{-1} \mathbf{S}_v^a[t] \left( (\mathbf{S}_v^a[t])^H (\boldsymbol{\Sigma}_{v,\alpha}^a[t])^{-1} \mathbf{S}_v^a[t] \right)^\dagger \cdot (\mathbf{S}_v^a[t])^H (\boldsymbol{\Sigma}_{v,\alpha}^a[t])^{-1} \mathbf{y}_v^a[t]. \quad (8.82)$$

The centralized decision rule is

$$\Lambda_v^a[t] \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \eta_{a,v}[t]. \quad (8.83)$$

Define the whitened projection matrix

$$\mathbf{P}_v^a[t] \triangleq (\boldsymbol{\Sigma}_{v,\alpha}^a[t])^{-1/2} \mathbf{S}_v^a[t] \left( (\mathbf{S}_v^a[t])^H (\boldsymbol{\Sigma}_{v,\alpha}^a[t])^{-1} \mathbf{S}_v^a[t] \right)^\dagger (\mathbf{S}_v^a[t])^H (\boldsymbol{\Sigma}_{v,\alpha}^a[t])^{-1/2} \quad (8.84)$$

and let

$$r_v^a[t] \triangleq \text{rank}(\mathbf{S}_v^a[t]). \quad (8.85)$$

Under the regularized Gaussian nuisance model and  $\mathcal{H}_0$ , the statistic is central chi-square with  $2r_v^a[t]$  DoF,

$$\Lambda_v^a[t] \sim \chi_{2r_v^a[t]}^2. \quad (8.86)$$

Under  $\mathcal{H}_1$ , with mean vector

$$\boldsymbol{\mu}_v^a[t] \triangleq \mathbf{S}_v^a[t] \boldsymbol{\gamma}_v[t], \quad (8.87)$$

the statistic is noncentral chi-square,

$$\Lambda_v^a[t] \sim \chi_{2r_v^a[t]}^2(\lambda_v^a[t]), \quad \lambda_v^a[t] = 2(\boldsymbol{\mu}_v^a[t])^H (\boldsymbol{\Sigma}_{v,\alpha}^a[t])^{-1/2} \mathbf{P}_v^a[t] (\boldsymbol{\Sigma}_{v,\alpha}^a[t])^{-1/2} \boldsymbol{\mu}_v^a[t]. \quad (8.88)$$

Therefore the centralized CFAR threshold at a target false-alarm probability  $P_{\text{FA},v}^*$  is

$$\eta_{a,v}[t] = 2Q^{-1}(r_v^a[t], P_{\text{FA},v}^*), \quad (8.89)$$

and the corresponding conditional detection probability is

$$P_{\text{D},v}^a[t] = 1 - F_{\chi_{2r_v^a[t]}^2(\lambda_v^a[t])}(\eta_{a,v}[t]). \quad (8.90)$$

The relation between the raw-uplink statistic and the distributed statistic is useful for interpretation. Let

$$\mathbf{U}_v[t] \triangleq \text{blkdiag}(\mathbf{u}_{m_1,v}[t], \dots, \mathbf{u}_{m_{M_v^{\text{det}}[t]},v}[t]). \quad (8.91)$$

If the CPU first projects  $\mathbf{y}_v^a[t]$  through  $\mathbf{U}_v^H[t]$ , the resulting low-dimensional observation is

$$\mathbf{y}_v^s[t] = \mathbf{U}_v^H[t] \mathbf{y}_v^a[t] = [y_{m_1,v}^{\text{MF}}[t], \dots, y_{m_{M_v^{\text{det}}[t]},v}^{\text{MF}}[t]]^T. \quad (8.92)$$

If the loaded projected covariance  $\mathbf{U}_v^H[t] \boldsymbol{\Sigma}_{v,\alpha}^a[t] \mathbf{U}_v[t]$  is diagonal with entries  $\{\sigma_{m,v}^2[t]\}$  and all contributing signatures are nonzero, then (8.82) reduces to

$$\Lambda_v^a[t] = \sum_{m \in \mathcal{M}_v^{\text{det}}[t]} \frac{2|g_{m,v}^*[t] y_{m,v}^{\text{MF}}[t]|^2}{\sigma_{m,v}^2[t] |g_{m,v}[t]|^2} = \sum_{m \in \mathcal{M}_v^{\text{det}}[t]} \frac{2|y_{m,v}^{\text{MF}}[t]|^2}{\sigma_{m,v}^2[t]} = \Lambda_v[t]. \quad (8.93)$$

Thus the centralized raw-uplink detector is statistically stronger only when the CPU can exploit antenna-domain or cross-receiver covariance information that the distributed scalar reports have already discarded.

## 8.7 UPLINK SINR AND SPECTRAL EFFICIENCY

Let  $\mathbf{h}_k[t]$  denote the network-stacked channel of UL user  $k$  toward the active receiving APs, and let the hybrid DL-originated nuisance covariance seen by the stacked combiner be

$$\mathbf{C}_d[t] \triangleq \text{blkdiag}(\mathbf{C}_{d,1}[t], \dots, \mathbf{C}_{d,M_u}[t]), \quad (8.94)$$

with per-AP block

$$\mathbf{C}_{d,m}[t] \triangleq \sum_{j \in \mathcal{A}_d[t]} \tilde{\mathbf{G}}_{mj}[t] \left( \sum_{n \in \mathcal{D}_j[t]} P_{d,jn}^{(c)}[t] \mathbf{w}_{jn}^P[t] (\mathbf{w}_{jn}^P[t])^H + \sum_{q=1}^{Q_j^{(s)}[t]} P_{j,q}^{(s)}[t] \mathbf{f}_{j,q}[t] \mathbf{f}_{j,q}^H[t] \right) \tilde{\mathbf{G}}_{mj}^H[t]. \quad (8.95)$$

The payload term in (8.95) employs  $P_{d,jn}^{(c)}[t]$  because the deterministic probing and the random data components share the same protected payload beam and their powers sum to the communication-side payload power. For the distributed UL path, the exact instantaneous UL SINR of user  $k$  is

$$\text{SINR}_k^{\text{UL}}[t] = \frac{P_u^{\max} |\mathbf{q}_k^H[t] \mathbf{h}_k[t]|^2}{P_u^{\max} \sum_{i \in \mathcal{U}_u[t] \setminus \{k\}} |\mathbf{q}_k^H[t] \mathbf{h}_i[t]|^2 + \mathbf{q}_k^H[t] \mathbf{C}_d[t] \mathbf{q}_k[t] + N_0 \|\mathbf{q}_k[t]\|^2}. \quad (8.96)$$

The first denominator term is ordinary UL multiuser interference. The second term is the hybrid DL→UL nuisance induced by the same payload and dedicated sensing beams that generated the sensing signatures. The third term is receiver noise after combining.

For the centralized UL protocol, the same expression is evaluated with the network-wide combiner  $\mathbf{q}_k^c[t]$  and the stacked channel vectors of Section 8.3:

$$\text{SINR}_k^{\text{UL},c}[t] = \frac{P_u^{\max} |(\mathbf{q}_k^c[t])^H \mathbf{h}_k^c[t]|^2}{P_u^{\max} \sum_{i \in \mathcal{U}_u[t] \setminus \{k\}} |(\mathbf{q}_k^c[t])^H \mathbf{h}_i^c[t]|^2 + (\mathbf{q}_k^c[t])^H \mathbf{C}_d^c[t] \mathbf{q}_k^c[t] + N_0 \|\mathbf{q}_k^c[t]\|^2}. \quad (8.97)$$

The instantaneous UL SE metric with payload window  $T = T_{\text{slot}} - \tau_p$  symbols is

$$\text{SE}_k^{\text{UL}}[t] = \left( 1 - \frac{\tau_p}{T_{\text{slot}}} \right) \log_2(1 + \text{SINR}_k^{\text{UL}}[t]), \quad (8.98)$$

for the distributed path, and

$$\text{SE}_k^{\text{UL},c}[t] = \left( 1 - \frac{\tau_p}{T_{\text{slot}}} \right) \log_2(1 + \text{SINR}_k^{\text{UL},c}[t]) \quad (8.99)$$

for the centralized protocol.

## 8.8 FRONTHAUL AND COMPUTATIONAL SCALABILITY

This section separates the fronthaul and computational burden of the two receiver protocols, and also separates the communication and sensing contributions within each path.

### 8.8.1 Distributed uplink and distributed GLRT

In the distributed UL, every online operation at UL AP  $m$  is local, bounded, or sparse. The communication part includes pilot-based channel estimation on  $\mathcal{U}_m[t]$ , construction of the adaptive directional nuisance covariance (8.16), and protected local L-MMSE combining for the served UL users. The sensing part includes one shared protected sensing covariance inverse and then one voxel-dependent receive beam and matched-filter accumulator for each voxel in the bounded list

$$V_m^{\text{sens}}[t] \triangleq |\mathcal{V}_m^{\text{sens}}[t]|, \quad V_m^{\text{sens}}[t] \leq B_u^{\text{max}}. \quad (8.100)$$

A conservative slot-level complexity count is therefore [12, 15, 17]

$$C_{\text{UL},m}^{\text{slot,dist}}[t] = \mathcal{O} \left( \underbrace{|\mathcal{U}_m[t]|N\tau_p + (|\mathcal{U}_m[t]| + |\mathcal{N}_m^{\text{DL}}|)N^2 + |\mathcal{U}_m[t]|N^3}_{\text{distributed communication front-end}} + \underbrace{N^3 + V_m^{\text{sens}}[t](N^2 + NT)}_{\text{distributed sensing front-end}} \right). \quad (8.101)$$

The first group corresponds to pilot processing, nuisance-covariance construction, and local communication combiners. The second group corresponds to the shared sensing-side inverse in (8.38) and the bounded voxel-wise receive-filter/matched-filter operations. If the matched filters are accumulated symbol by symbol, the per-payload-symbol work can be written as

$$C_{\text{UL},m}^{\text{sym,dist}}[t] = \mathcal{O}(N(|\mathcal{U}_m[t]| + V_m^{\text{sens}}[t])), \quad (8.102)$$

which remains bounded per AP because  $|\mathcal{U}_m[t]|$  is pilot-capped and  $V_m^{\text{sens}}[t] \leq B_u^{\text{max}}$ .

The distributed fronthaul also separates naturally into control, communication return, and sensing return. The CPU→UL-AP direction carries a short activation

word, user identifiers, the voxel identifiers in  $\mathcal{V}_m^{\text{sens}}[t]$ , and bounded active-InAI descriptors for  $\mathcal{B}_m[t]$ , such as active neighbor identifiers and  $P_{d,j}^{\text{tot}}[t]$  values or, equivalently, the precomputed scalar  $\lambda_{\text{IA},m}^{\text{UL,eff}}[t]$ . The return direction carries one effective gain  $\tilde{g}_{mk}[t]$  per served user, soft symbols  $z_{mk}[\tau]$ , and at most one detection statistic  $\Lambda_{m,v}[t]$  per processed voxel. Per-slot distributed fronthaul is summarized as

$$F_{C \rightarrow U}^{\text{dist}}[t] = \underbrace{M_u[t]b_{\text{ctl}}}_{\text{activation}} + b_{\text{idx}} \underbrace{\sum_{m \in \mathcal{A}_u[t]} (|\mathcal{U}_m[t]| + V_m^{\text{sens}}[t])}_{\text{user and voxel identifiers}} + b_{\text{sc}} \underbrace{\sum_{m \in \mathcal{A}_u[t]} |\mathcal{B}_m[t]|}_{\text{active InAI descriptors}}, \quad (8.103)$$

$$F_{U \rightarrow C}^{\text{dist,comm}}[t] = \underbrace{b_{\text{sc}}S_u[t]}_{\{\tilde{g}_{mk}[t]\}} + \underbrace{b_zTS_u[t]}_{\{z_{mk}[\tau]\}}, \quad (8.104)$$

$$F_{U \rightarrow C}^{\text{dist,sens}}[t] = \underbrace{(b_{\text{idx}} + b_{\text{sc}}) \sum_{m \in \mathcal{A}_u[t]} V_m^{\text{sens}}[t]}_{\{(v, \Lambda_{m,v}[t])\}}, \quad (8.105)$$

where  $S_u[t] \triangleq \sum_{m \in \mathcal{A}_u[t]} |\mathcal{U}_m[t]|$  and  $b_{\text{ctl}}$ ,  $b_{\text{idx}}$ ,  $b_{\text{sc}}$ , and  $b_z$  denote the bit budgets for one control word, one index, one scalar, and one complex soft symbol, respectively.

### 8.8.2 Centralized uplink and centralized GLRT

The centralized uplink moves all baseband processing to the CPU: active UL APs act as radio heads that forward antenna-domain samples, while the CPU performs channel estimation, combining, sensing matched filtering, and GLRT evaluation. This is practical when the fronthaul can carry raw samples, but increases fronthaul load compared with AP-side compression.

Scalability is preserved under the scheduling protocol of Chapter 6 because each receiving AP only serves a bounded local user set  $\mathcal{U}_m[t]$  after pilot coordination and service capping, as established in (6.27). Let

$$\mathcal{K}_u^{\text{adm}}[t] \triangleq \bigcup_{m \in \mathcal{A}_u[t]} \mathcal{U}_m[t] \quad (8.106)$$

be the users admitted by at least one active AP in slot  $t$ . From (6.27) we have

$$|\mathcal{K}_u^{\text{adm}}[t]| \leq \sum_{m \in \mathcal{A}_u[t]} |\mathcal{U}_m[t]| \leq \tau_p M_u[t], \quad (8.107)$$

so the per-AP UL workload stays bounded as the total user population grows; extra users are either handled by other local clusters or not admitted in the current slot.

If the CPU performs only payload processing centrally and receives channel estimates through another mechanism, the raw return fronthaul is

$$F_{U \rightarrow C}^{\text{cent,pay}}[t] = \underbrace{b_z N T M_u[t]}_{\{\mathbf{Y}_m[t]\}_{m \in \mathcal{A}_u[t]}}. \quad (8.108)$$

If the pilot observations are also transported for centralized channel estimation, the additional term is

$$F_{U \rightarrow C}^{\text{cent,pil}}[t] = \underbrace{b_z N \tau_p M_u[t]}_{\text{raw pilot observations}}. \quad (8.109)$$

The same uploaded raw block supports both centralized communication and centralized sensing; no separate scalar sensing report is needed. The only UL-side sensing-specific metadata is the informative voxel list and sequence/receiver book-keeping:

$$F_{U \rightarrow C}^{\text{cent,sens-meta}}[t] = \underbrace{b_{\text{idx}} |\mathcal{V}_{\text{inf}}[t]|}_{\text{voxel list}} + \underbrace{b_{\text{idx}} \sum_{v \in \mathcal{V}_{\text{inf}}[t]} M_v^{\text{det}}[t]}_{\text{voxel-receiver associations}}, \quad (8.110)$$

provided that the CPU already knows the deterministic sequence plan and the scheduled UL receiver list. Compared with (8.105), centralized sensing replaces one scalar per AP–voxel pair by raw antenna-domain sample transport for every active UL AP. This is a much stronger fronthaul assumption than the distributed receiver path.

The centralized communication computation is dominated by the covariance factorization over the active UL antenna stack in (8.33). Using one shared covariance inverse for all admitted UL users gives the conservative count

$$C_{\text{CPU}}^{\text{cent,comm}}[t] = \mathcal{O}((N M_u[t])^3 + |\mathcal{K}_u^{\text{adm}}[t]| (N M_u[t])^2 + T |\mathcal{K}_u^{\text{adm}}[t]| N M_u[t]). \quad (8.111)$$

The cubic term is the cost of factorizing the centralized covariance, the second term forms the user combiners, and the third term applies them over the payload block. Using (8.107), the user-dependent part is upper bounded by a function of  $\tau_p$  and  $M_u[t]$ . Consequently, under the bounded local-service rule of Chapter 6, the centralized uplink does not violate the user-scalability criterion of Definition 5.1.

For sensing, the antenna-domain observation for voxel  $v$  has dimension  $NM_v^{\text{det}}[t]$ , and a conservative CPU count is

$$C_{\text{CPU}}^{\text{cent,sens}}[t] = \mathcal{O} \left( \sum_{v \in \mathcal{V}_{\text{inf}}[t]} \left[ (NM_v^{\text{det}}[t])^3 + NM_v^{\text{det}}[t]T + (NM_v^{\text{det}}[t])^2 \right] \right). \quad (8.112)$$

The cubic term represents whitening or factorization of the voxel-specific regularized antenna-domain nuisance covariance in (8.77). The remaining terms represent sequence matching, stacking, and evaluation of the quadratic form. If  $M_v^{\text{det}}[t]$  is bounded by the voxel-neighborhood design, the per-voxel sensing dimension remains bounded. Hence, the centralized uplink is also scalable in the sensing hypothesis dimension under the bounded receiver-assignment rule.

Table 8.1 summarizes the key operational points.

| ELEMENT                       | ONLINE BURDEN                                                  | SCALABILITY                                                                         |
|-------------------------------|----------------------------------------------------------------|-------------------------------------------------------------------------------------|
| Distributed UL at AP $m$      | Form local covariance and solve one $N \times N$ system per UE | Scales with local UE and nuisance sets only                                         |
| Distributed sensing at AP $m$ | One protected inverse and one matched filter per voxel         | Uses only locally assigned voxels, $\leq B_u^{\text{max}}$                          |
| Distributed CPU fusion        | Sum soft symbols and scalar GLRTs                              | One soft-symbol stream per UE-AP and one scalar per AP-voxel                        |
| Centralized UL                | Upload raw antenna blocks and invert network covariance        | Scales with admitted UEs but requires raw-sample transport and global factorization |
| Centralized sensing           | Evaluate (8.82) on matched-filter stacks                       | Richer covariance information but needs centralized raw data                        |

Table 8.1: Scalability summary of the UL and detection layer. The proposed distributed path keeps the heaviest online processing local at the APs, while the centralized UL path quantifies the cost and benefit of raw-sample CPU processing.





# 9 | NUMERICAL RESULTS

The previous chapters introduced the proposed DTDD CF-ISAC protocol from three complementary viewpoints. Chapter 6 defined the slow- and fast-timescale control architecture, including the AP mode assignment, requested-voxel scheduler, deterministic sensing-sequence reuse, and adaptive grid logic. Chapter 7 then described the DL waveform: protected payload precoding, deterministic superimposed probing, and dedicated virtual sensing beams admitted only when the communication constraints leave sufficient spatial and power slack. Finally, Chapter 8 developed the UL receiver split, the distributed coherent GLRT, the centralized raw-uplink variant, and the corresponding UL SE expressions.

The purpose of this chapter is to connect those design blocks to the numerical behavior of the complete system. The results show first that opportunistic payload illumination is not enough for reliable drone surveillance: dedicated sensing-aware DL resources are needed to move the detector away from the false-alarm floor. Among the sensing-aware options, the hybrid waveform gives the strongest sensing performance when communication constraints leave enough residual power and spatial freedom, while naturally falling back toward Deterministic probing when the communication threshold or the user load becomes high. The chapter also makes the hard communication constraints of Chapter 5 explicitly verifiable, by reporting the fraction of scheduled UEs that actually meet their per-user SINR threshold, and by showing that the unsatisfied minority is the geometric infeasibility floor of a random CF deployment rather than a violation introduced by sensing.

Then, the chapter shows that residual InAI must be handled on both sides of the DTDD split. The protected DL precoder sacrifices a small amount of communication rate to preserve UL operation, and the protected UL combiner prevents the remaining leakage from dominating the receiver. For sensing, the centralized detector remains a useful upper benchmark, and the gap it opens over the distributed GLRT is regime-dependent rather than uniform: it is largest for weak targets and in the steep transition region of the detector, where cross-AP correlation carries the most information, and it nearly closes once the illumination is strong enough to drive the distributed statistic into saturation. The distributed GLRT therefore retains the practically relevant high- $P_D$  behavior in the strong-illumination regime, with much lower fronthaul and processing requirements, but it is not interchangeable with the centralized detector across all operating points. A dedicated experiment then shows that, under different DTDD UL/DL traffic ratio, the communication SE and sensing retention are jointly near-optimal at a moderately DL-leaning split. Finally, the tracking and hardware-budget results show that the proposed surveillance task is mainly limited by revisit budget, target separability, and distributed geometry: spreading the antenna budget over many APs is more valuable than concentrating the same antennas into a few large arrays.

Throughout the chapter, detection performance is reported relative to operational references rather than to a protocol-independent fundamental limit. The centralized receiver acts as an upper benchmark that quantifies the value of cross-AP covariance information, while the opportunistic waveform and the false-alarm floor act as lower references.

The full ROC curves supporting the  $P_{FA}$  points of the main chapter are reported in Appendix A, while Appendix B validates the adaptive voxel-grid controller used by the scan-and-track layer.

## 9.1 SIMULATION METHODOLOGY

All results are generated from the system model of Chapter 5 and from the protocol blocks developed in Chapters 6–8. Unless otherwise stated, the baseline deployment contains  $M = 36$  distributed APs, each equipped with  $N = 16$  antennas arranged as a virtual  $4 \times 4$  Uniform Planar Array (UPA), over a  $500 \times 500$  m wrap-around service area. The default communication load is  $K = 24$  single-antenna UEs, split equally between DL and UL traffic. The surveillance region covers the same horizontal area and the altitude interval  $[5, 15]$  m; the target process contains 20 moving drones with 0 dBsm RCS, unless the target count is explicitly swept, and the drone target model is detailed in Section 9.1.1. Figure 9.1 shows a representative snapshot of the simulated scenario, including the UEs, APs, target positions and voxel-grid structure in a single Monte-Carlo drop and slot. The baseline radio parameters are 1.9 GHz carrier frequency, 20 MHz bandwidth, 1 W maximum DL power per AP, and 0.2 W UL payload power per UE. Each slot contains  $T_{\text{slot}} = 200$  symbols, of which  $\tau_p = 10$  are used for pilots and  $T = 190$  are available for data and sensing. The numerical results are averaged over 1000 Monte-Carlo drops and 200 evaluated slots per drop, except for the adaptive-grid validation in Appendix B, where the horizon is increased when needed to show convergence. The use of wrap-around geometry allows boundary effects to be removed without requiring a physically larger simulated network [12].

The protocol uses the adaptive voxel grid of Section 6.2. The default coarse grid is initialized at  $12 \times 12 \times 2$  voxels, and each coarse voxel contains  $2 \times 2 \times 2$  fine children. The controller then observes the informative scan budget  $Q_{\text{eff}}[t]$  in (6.3), subtracts the active track-maintenance demand, and resizes the coarse grid according to (6.7)–(6.9).

### 9.1.1 Surveillance target model

The swept target parameters are chosen to represent low-altitude UAV traffic. The baseline RCS is 0 dBsm. From Table 3.1, a small UAV typically ranges from  $-20$  to 0 dBsm; the baseline therefore models a larger platform or a small drone observed

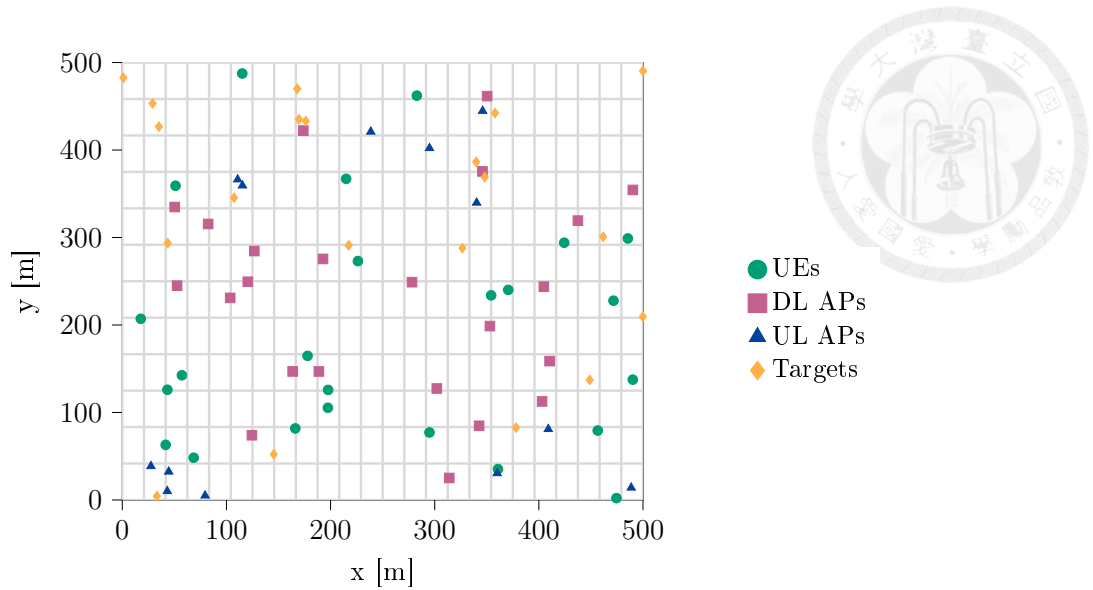


Figure 9.1: Representative simulation scenario. The figure shows one Monte-Carlo drop in the  $500 \times 500$  m wrap-around area, including communication UEs, DL APs, UL APs, and target positions. The role split changes from slot to slot according to the DTDD scheduler of Section 6.4.

at a favorable aspect, while the RCS sweep of Figure 9.8 extends from  $-40$  dBsm, representing low-observability micro-drones, up to  $10$  dBsm for large fixed-wing platforms.

Drone mobility is described by a variable-speed planar motion model with a configurable heading process. The baseline speed is  $20$  m/s, which represents the cruising regime of civilian platforms, while in Figure 9.9 the velocity is swept up to an aggressive stress point, well above common civilian UAV speeds, retained to map the robustness envelope of the tracker rather than to model a specific aircraft. Two heading processes are studied, corresponding to two representative agility classes. In the *fixed-heading* model each drone draws a heading uniformly in  $[0, 2\pi)$  at the start of the drop and keeps it for the entire horizon, so the trajectory is a straight line at constant heading; this represents a low-agility, forward-cruising platform such as a fixed-wing or glider-like UAV flying along a corridor. In the *random-heading* model each drone re-draws its heading uniformly in  $[0, 2\pi)$  at random instants, and holds it between changes; this represents a highly agile multi-rotor drone that turns and changes direction frequently. In both models the speed is independently re-jittered around the drone mean every slot and projected onto the current heading. The

fixed-heading model underlies the retention heatmap of Section 9.7, where the two models are also compared.

However, it is important to note that the protocol itself is target-agnostic: it operates on voxel-quantized echoes and makes no UAV-specific assumption, so it can detect and track any low-altitude target with comparable RCS and mobility, such as birds, light aircraft, or other small aerial vehicles. The drone framing fixes representative altitude, speed, and RCS ranges; it does not restrict the class of detectable targets.

### 9.1.2 *Protocol variants and receiver split*

The figures compare the three DL sensing variants summarized in Table 9.1. They share the same scheduler, the same requested-voxel distribution, and the same protected payload-precoding pipeline. The difference is only which sensing resources are allowed after the communication beams have been constructed. The opportunistic sensing variant forces both the superimposed probing fraction  $\rho_{jn}[t]$  and the dedicated sensing split  $\beta_{jv}[t]$  to zero. It is therefore an opportunistic baseline: target illumination occurs only when a communication payload beam happens to illuminate a requested voxel. The deterministic sensing variant allows deterministic probing on selected payload streams but forces the dedicated branch to zero. The hybrid sensing variant is the proposed design; it admits both superimposed probing and dedicated virtual sensing beams whenever the feasibility tests of Sections 7.3 and 7.4 allow them.

Each sensing variant is evaluated under the distributed and centralized UL receiver protocols described in Chapter 8. The split does not modify the DL waveform. Consequently, any distributed–centralized difference in DL SE is expected to be zero, while the UL SE and the sensing detector may differ. Unless otherwise stated clearly in the figure, the simulation uses the distributed UL protocol.

| VARIANT               | CONSTRAINT                                  | INTERPRETATION                                                                                                                                            |
|-----------------------|---------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------|
| Opportunistic sensing | $\rho_{jn}[t] = 0$ ,<br>$\beta_{jv}[t] = 0$ | Communication payload beams are transmitted with no deterministic sensing and no dedicated sensing beam. Sensing is incidental to the communication field |
| Deterministic sensing | $\beta_{jv}[t] = 0$                         | Deterministic sensing signatures may be superimposed on communication beams, but no dedicated sensing beams are created                                   |
| Hybrid sensing        | Full protocol                               | This variant uses superimposed probing and, when residual power and DoF are available, dedicated sensing beams                                            |

Table 9.1: Protocol variants used in the numerical results.

### 9.1.3 Communication and sensing metrics

For Monte-Carlo drop  $d$  and slot  $t$ , let  $\mathcal{U}_d^{(d)}[t]$  and  $\mathcal{U}_u^{(d)}[t]$  denote the scheduled DL and UL UE sets. The instantaneous DL SE follows (7.67), and the distributed and centralized UL SE metrics follow (8.98) and (8.99). The plotted averages are

$$\overline{\text{SE}}_{\text{DL}} \triangleq \frac{1}{N_{\text{MC}} N_s} \sum_{d=1}^{N_{\text{MC}}} \sum_{t=1}^{N_s} \sum_{k \in \mathcal{U}_d^{(d)}[t]} \text{SE}_{d,k}^{(d)}[t], \quad (9.1)$$

$$\overline{\text{SE}}_{\text{UL}} \triangleq \frac{1}{N_{\text{MC}} N_s} \sum_{d=1}^{N_{\text{MC}}} \sum_{t=1}^{N_s} \sum_{k \in \mathcal{U}_u^{(d)}[t]} \text{SE}_{u,k}^{(d)}[t], \quad (9.2)$$

where  $N_{\text{MC}}$  is the number of drops and  $N_s$  is the number of evaluated slots. When a figure reports aggregate communication performance, it uses

$$\overline{\text{SE}}_{\text{sum}} \triangleq \overline{\text{SE}}_{\text{DL}} + \overline{\text{SE}}_{\text{UL}}. \quad (9.3)$$

The sensing results are evaluated after the initial scan transient. Let  $W$  be the revisit window in slots from (6.4); the steady-state slot set is

$$\mathcal{T}_{\text{ss}} \triangleq \{t \in \{1, \dots, N_s\} : t > W\}. \quad (9.4)$$

For a detector architecture  $\chi \in \{\text{dist}, \text{cent}\}$ , let  $\Lambda_v^\chi[d, t]$  be the statistic assigned to voxel  $v$  in drop  $d$  and slot  $t$ , using either the distributed statistic in (8.57) or the centralized statistic in (8.82). Let  $\mathcal{E}_1^{(d,t)}$  be the set of target-present voxels in the request set  $\mathcal{V}_{\text{req}}^{(d,t)}$ , and  $\mathcal{E}_0^{(d,t)}$  the set of non-target voxels in the same request.

$\mathbf{1}\{\Lambda_v^x[d, t] > \eta\}$  is gated to zero for voxels that the DL plan does not illuminate, so non-illuminated target voxels count as missed detections. For a fixed threshold  $\eta$ , the empirical  $P_D$  and  $P_{FA}$  are

$$\hat{P}_D^x(\eta) = \frac{\sum_d \sum_{t \in \mathcal{T}_{ss}} \sum_{v \in \mathcal{E}_1^{(d,t)}} \mathbf{1}\{\Lambda_v^x[d, t] > \eta\}}{\sum_d \sum_{t \in \mathcal{T}_{ss}} |\mathcal{E}_1^{(d,t)}|}, \quad (9.5)$$

$$\hat{P}_{FA}^x(\eta) = \frac{\sum_d \sum_{t \in \mathcal{T}_{ss}} \sum_{v \in \mathcal{E}_0^{(d,t)}} \mathbf{1}\{\Lambda_v^x[d, t] > \eta\}}{\sum_d \sum_{t \in \mathcal{T}_{ss}} |\mathcal{E}_0^{(d,t)}|}. \quad (9.6)$$

The main figures report  $\hat{P}_D^x$  at the threshold corresponding to the target false-alarm probability  $P_{FA}^* = 10^{-2}$ , unless another value is explicitly shown. This operating-point view is complemented by the full ROC curves in Appendix A.

For the tracking analysis in Figure 9.9 and Figure 9.7, the reported quantity is not a one-slot  $P_D$ . Let  $\mathcal{G}[d, t]$  be the set of physical targets in drop  $d$  that are assigned to exactly one active track at slot  $t$ , with no other target assigned to the same track. The unique-share metric is

$$S_{\text{uniq}}[d, t] \triangleq \frac{|\mathcal{G}[d, t]|}{N_{\text{tar}}}, \quad (9.7)$$

where  $N_{\text{tar}}$  is the number of targets. The plots report the average over the last  $L_{\text{tail}}$  revisit windows,

$$\bar{S}_{\text{uniq}} \triangleq \frac{1}{N_{\text{MC}} |\mathcal{T}_{\text{tail}}|} \sum_{d=1}^{N_{\text{MC}}} \sum_{t \in \mathcal{T}_{\text{tail}}} S_{\text{uniq}}[d, t]. \quad (9.8)$$

This definition penalizes both missed tracks and ambiguous many-target-to-one-track assignments. It is therefore stricter than a raw detection count and better aligned with the surveillance objective of maintaining separate targets over time.

## 9.2 SENSING–COMMUNICATION TRADE-OFF

The first analysis in Figure 9.2 varies the communication threshold  $\Gamma^{\min}$  used by the DL feasibility logic. This experiment is the most direct numerical illustration of the best-effort sensing formulation of Section 5.4.1: as  $\Gamma^{\min}$  grows, the hard communication constraints (5.8)–(5.9) absorb progressively more of the available

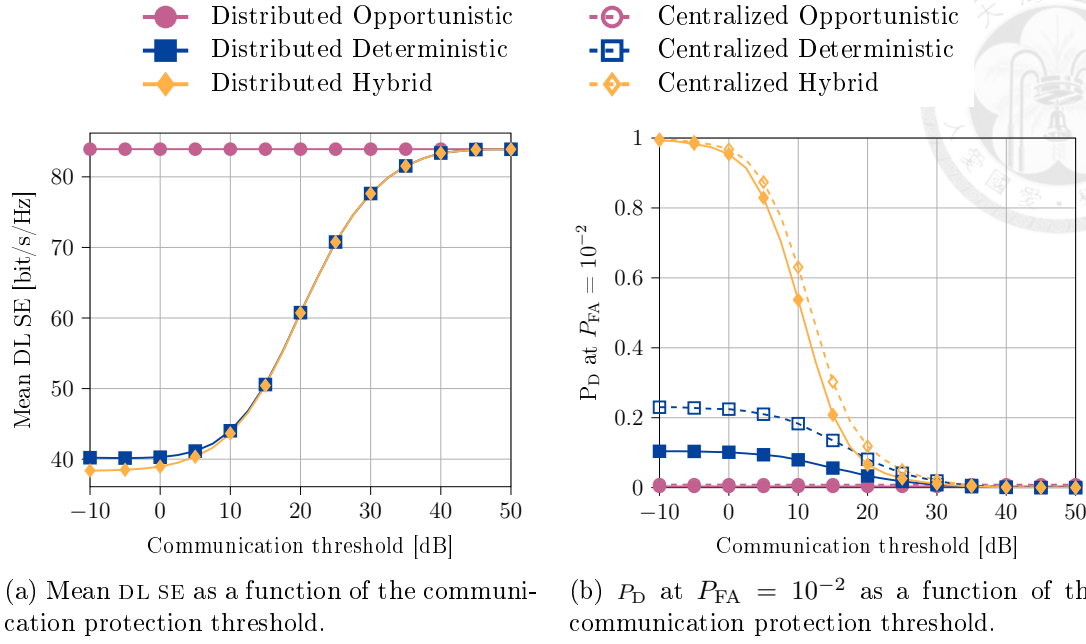


Figure 9.2: Sensing-communication trade-off over the communication threshold. The Opportunistic sensing baseline keeps all DL power on communication payload beams. The Deterministic and Hybrid variants progressively return power to communication as the protection threshold increases. The centralized and distributed curves coincide in the DL SE panel because the receiver split does not change the DL waveform, while the  $P_D$  panel shows the effect of centralized colored-covariance fusion.

power budget, leaving less for the soft sensing objective. This is the central trade-off encoded by the superimposed probing ceiling in Section 7.3 and by the residual dedicated power in (7.48).

The Opportunistic sensing curve is flat in both panels. Since it enforces  $\rho_{jn}[t] = 0$  and  $\beta_{jv}[t] = 0$ , the communication threshold has no path through which it can alter the waveform: the entire DL power budget remains on payload beams. Consequently, the corresponding sensing performance is close to the false-alarm floor. This confirms that opportunistic payload illumination is not sufficient for reliable drone surveillance in the baseline geometry. The user-directed beams occasionally cross the surveillance volume, but they are not designed to create a stable non-centrality in the voxel GLRT statistic of (8.54).

The sensing-aware variants show the opposite behavior in the DL SE panel. At low threshold, they are allowed to allocate a large fraction of the slack power to sensing, and the DL SE falls to about 40 bit/s/Hz. As  $\Gamma^{\min}$  increases, the feasibility

condition behind the probing split becomes stricter: the support term required to preserve (7.66) grows with  $\sqrt{\Gamma^{\min}}$ , so the admissible probing fraction decreases. In parallel, the minimum payload power approaches the per-AP power budget, so the dedicated ceiling in (7.48) vanishes. The sensing-aware DL curves therefore climb monotonically and meet the Opportunistic sensing value at the high-threshold end of the sweep. This convergence is not accidental; once both  $\rho_{jn}[t]$  and  $\beta_{jv}[t]$  are forced to zero by the constraints, all three variants reduce to the same protected communication precoder.

The  $P_D$  panel shows why the hybrid branch is useful. At  $\Gamma^{\min} = 0$  dB, distributed deterministic sensing reaches  $P_D = 0.255$ , while distributed hybrid sensing reaches  $P_D = 0.980$ . The centralized detector raises the corresponding values to 0.568 and 0.999. The reason is the different power geometry of the two sensing resources. The superimposed component is attached to a communication beam and is bounded by the communication SINR constraint. The dedicated beam, in contrast, uses the residual power after the minimum payload support has been paid and is shaped toward the voxel subject to the protected null-space constraints of Section 7.4. At low threshold this residual power is large, so the local SCNR in (8.52) moves the target-present statistic far into the noncentral regime.

The centralized detector is consistently above the distributed one in Figure 9.2b, but the gap between the two architectures is strongly regime-dependent. Its size is governed by how much exploitable cross-AP structure the centralized detector can recover beyond the distributed scalar sum. This is why the gain depends jointly on the variant and on the threshold. A representative weak-illumination point is  $\Gamma^{\min} = 0$  dB: for the Deterministic variant, whose superimposed-only illumination produces a weak per-voxel noncentrality, the centralized detector lifts  $P_D$  from 0.255 to 0.568: a gain of more than 0.31. For the Hybrid variant across the sweep, the centralized gain becomes largest in the transition region between roughly 15 and 25 dB, where the distributed  $P_D$  is neither saturated nor at the floor and the colored covariance has the most to exploit.

At  $\Gamma^{\min} \geq 45$  dB, the sensing-aware curves collapse toward the false-alarm floor. The communication requirement is so strict that the precoder admits essentially no superimposed or dedicated sensing voxel; the target-present voxels stay in  $\mathcal{E}_1^{(d,t)}$  but are not illuminated, so by (9.5) they count as misses and  $\hat{P}_D^x \rightarrow P_{FA}^*$ . The opportunistic detector still evaluates each requested voxel paired with its best incidental communication beam, but the resulting non-centrality is too small to lift the statistic above  $\eta^*$  on average.

The curves above quantify the average communication and sensing cost of the threshold sweep, but they do not by themselves show whether the strict per-UE constraints (5.8)–(5.9) are actually met. Section 9.2.1 complements them by reporting, over the same threshold sweep, the fraction of scheduled UEs that satisfy their individual SINR target.

### 9.2.1 Communication threshold satisfaction

Figure 9.3 reports the fraction of scheduled UEs whose effective SINR actually clears the threshold, as a function of the communication threshold  $\Gamma^{\min}$ .

A subtle but important feature of this experiment is that  $\Gamma^{\min}$  plays a dual role, exactly as in Section 9.2. It is at the same time the *design target* that each DL AP uses to size the probing ceiling  $\bar{\rho}_{jn}[t]$  in (7.25) and the residual dedicated ceiling  $\bar{\beta}_j[t]$  in (7.48), and the *evaluation level* used for the satisfaction curves. Every point on the curve is therefore a self-consistent operating point: the precoder is built to protect  $\Gamma^{\min}$ , and the network is then measured at the same  $\Gamma^{\min}$ . This coupling is the reason the sensing variants and the Opportunistic baseline must eventually coincide at large  $\Gamma^{\min}$ , as discussed below.

DOWNLINK. At the  $\Gamma^{\min} = 5$  dB operating point used in the load and hardware experiments, the full Hybrid waveform satisfies the DL constraint in (5.8) for 84.1% of the scheduled DL UEs, the Deterministic variant for 84.9%, and the Opportunistic baseline for 86.9%. The hard constraint is therefore met for the large majority of UEs,

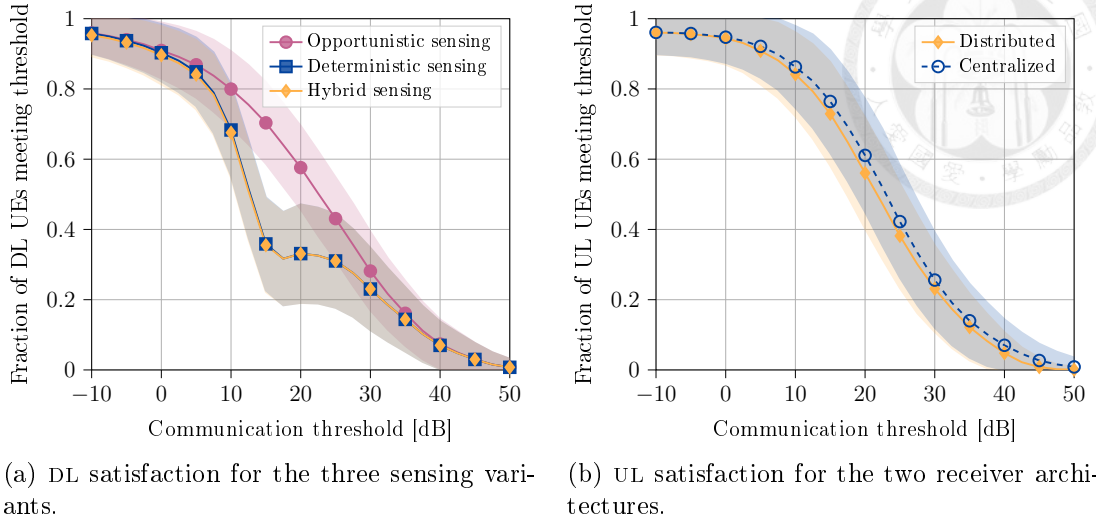


Figure 9.3: Communication SINR-threshold satisfaction. Each curve reports the fraction of scheduled UEs satisfying the hard constraints (5.8)–(5.9) as a function of the communication threshold  $\Gamma^{\min}$ . (a) The sensing cost is non-monotone in  $\Gamma^{\min}$ , rising to a maximum near 15 dB before the design target switches sensing off; the Deterministic and Hybrid curves are nearly coincident and both rejoin the Opportunistic baseline at high  $\Gamma^{\min}$ . (b) The distributed and centralized UL receivers, which are independent of the sensing variant. Shaded bands denote the  $\pm 1\sigma$  spread across deployments.

and the price paid by sensing at this operating point is only about 2.8 percentage points of satisfied DL UEs. Across the full sweep, however, this cost is markedly non-monotone.

The steep collapse of the two sensing-aware curves between  $\Gamma^{\min} = 10$  and 15 dB, from about 68% to about 35%, is the most prominent feature of the panel. Its cause is the way deterministic probing enters the DL SINR of (7.66). The dominant cost is not the  $(1 - \rho_{jn}[t])$  scaling of the desired term  $S_n^{\text{sig}}[t]$ , which removes less than 1 dB of useful power at the probing fractions observed here. It is the probing-leakage term  $I_{n,P}^{\text{DL}}[t]$ , which accumulates  $P_{d,jk}^{(c)}[t]\rho_{jk}[t]|\mathbf{h}_{jn}^H[t]\mathbf{w}_{jk}^P[t]|^2$  over *all* active payload branches in the network. Each served UE is illuminated not only by its own probing but by the probing superimposed on every other UE’s payload beam, so deterministic probing raises a network-wide interference floor. This produces the two large single-step drops of about 16 and 15 percentage points in the sensing-aware curves, against only about 4 to 5 percentage points for the Opportunistic baseline over the same two steps.

The falling edge is the protocol withdrawing sensing. Because  $\Gamma^{\min}$  is also the design target, the support term  $(\sqrt{\Gamma_n^{\min} \Upsilon_n^{\text{safe}}[t]} - \Omega_{jn}^{\text{safe}}[t])_+^2$  in (7.25) grows with  $\sqrt{\Gamma^{\min}}$ , so  $\bar{\rho}_{jn}[t] \rightarrow 0$ ; in parallel the minimum protected payload power  $P_{d,jn}^{\min}[t]$  in (7.46) approaches the per-AP budget and the dedicated ceiling  $\bar{\beta}_j[t]$  in (7.48) vanishes. With  $\rho_{jn}[t] \rightarrow 0$  and  $\beta_{jv}[t] \rightarrow 0$  the Deterministic and Hybrid precoders become physically identical to the Opportunistic one, and the three DL curves merge. Where the QoS demand is gentle the cost of sensing is negligible, and where it is stringent the protocol has already returned all power to communication.

The change of satisfaction along the sweep is the sum of two opposing effects: a threshold effect, which is negative because the bar rises and whose magnitude is the density of UEs sitting exactly at  $\Gamma^{\min}$ , and a relief effect, which is positive because raising the design target drives  $\bar{\rho}_{jn}[t]$  down, lowers the network-wide probing floor, and shifts the whole SINR distribution upward. Past about 22.5 dB the probing fraction has decayed toward zero, the relief effect saturates, and the monotone decline returns.

The Deterministic and Hybrid DL curves are nearly coincident over the whole range, separated by less than one percentage point. The only difference between them is the dedicated sensing branch, whose leakage onto served UEs is the term  $I_{n,\text{VS}}^{\text{DL}}[t]$  in (7.66). This leakage is small because the dedicated beams are formed in the complement of the served-UE subspace through the projector  $\mathbf{\Pi}_j^\perp[t]$  of (7.39)–(7.42) and are steered toward three-dimensional airspace voxels that are angularly separated from the ground-level UEs. The meaningful DL distinction in this figure is therefore between sensing and no sensing; the dedicated beams cost essentially nothing in constraint satisfaction, consistent with the small DL SE difference between the two variants seen elsewhere in this chapter.

Both ends of the sweep also expose geometric floors that no waveform choice can move. At  $\Gamma^{\min} = -10$  dB about 4–5% of the DL UEs remain unsatisfied even with sensing disabled, and at  $\Gamma^{\min} = 50$  dB all three curves collapse to about 1%. These limits are set by the random deployment geometry: a UE with very low aggregate

path gain cannot reach a high SINR even when its serving APs spend their entire budget on it, and a 50 dB DL SINR is simply unreachable in an interference-limited CF network of this size. For exactly those UEs, the local ceiling (7.25) has already set  $\rho_{jn}[t] = 0$  and, through (7.46)–(7.48), has left no power for dedicated sensing, so the protocol is already spending the entire available budget on communication and still cannot lift those links above the threshold. This is precisely the behavior anticipated by the best-effort sensing philosophy of Section 5.4.1: the communication constraints are enforced first, with all available power, and sensing consumes only what remains.

UPLINK. The UL panel contains only two curves since the sensing variant does not alter the UL receiver comparison. The centralized receiver gives a higher mean satisfaction fraction throughout the sweep, although the two architectures are nearly indistinguishable at loose thresholds: at  $\Gamma^{\min} = -10$  dB both satisfy about 96% of the scheduled UL UEs, while at the baseline value of 5 dB the centralized and distributed fractions are 92.1% and 90.8%, respectively. The absolute gain becomes most pronounced in the intermediate-to-high threshold region, reaching about five percentage points at 20 dB, where centralized processing satisfies 61.1% of the UEs compared with 56.0% under distributed processing; beyond this region both curves decay toward zero as the target enters the extreme SINR tail. This consistent ordering follows from the network-wide combiner’s ability to jointly exploit the antenna-domain observations.

### 9.3 TRADE-OFF UNDER COMMUNICATION LOAD

The second analysis in Figure 9.4 keeps the communication protection threshold fixed at 5 dB and sweeps the number of active UEs. This experiment asks whether the hybrid advantage survives when the system becomes increasingly communication-dominated. The UE split remains balanced between DL and UL traffic, so increasing

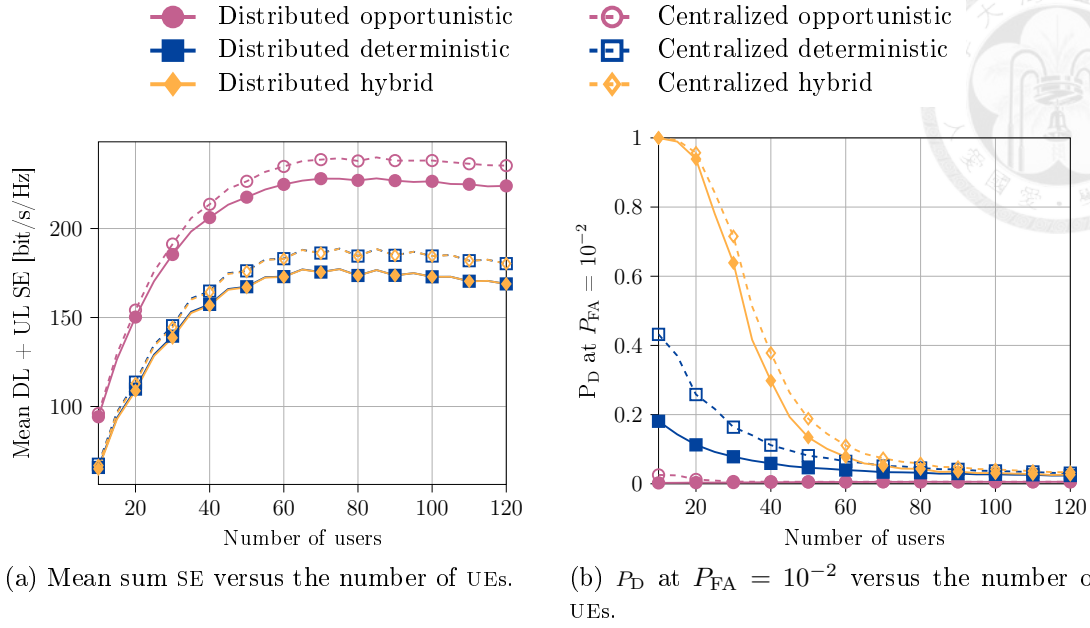


Figure 9.4: Trade-off under increasing communication load. The left panel reports  $\overline{SE}_{\text{sum}}$  from (9.3); the right panel reports  $P_D$  at  $P_{FA} = 10^{-2}$ . As the number of UEs increases, Opportunistic sensing remains the communication upper envelope, while the hybrid sensing advantage gradually disappears because the dedicated sensing power ceiling is consumed by communication support. The sweep extends to  $K = 120$  to expose the full degradation envelope; the region  $K \geq M = 36$  represents a stress-test condition in which the Cell-Free premise  $K \ll M$  is violated, rather than an intended operating point.

$K$  simultaneously increases the number of served streams, the pilot-reuse pressure, and InUI.

The sum SE grows sub-linearly and then saturates. The centralized hybrid curve is about 7% higher because the network-wide UL combiner improves the UL part of (9.3). Opportunistic sensing remains the communication upper envelope because it never diverts DL power to sensing.

Three effects explain the saturation. First, the pilot length is fixed at  $\tau_p = 10$ , so loads above  $K = 10$  require pilot reuse and therefore degrade the channel estimates as described in Section 2.2. Second, the user-centric service sets remain locally bounded for scalability, so each AP has only  $N = 16$  antenna dimensions with which to serve and protect its active streams. Third, every additional UE adds desired signal but also adds multi-user interference, InUI, and scheduling pressure. At large

$K$ , the marginal throughput of the new stream is smaller than the extra interference it creates, which explains the mild droop at the right end of the curves.

The sensing panel reveals the central limitation of the hybrid design. At low load, the dedicated sensing branch has substantial residual power. By  $K = 120$ , distributed hybrid and distributed superimposed-only sensing are almost indistinguishable. The centralized curves follow the same trend, and the spacing between the two architectures repeats the regime-dependent pattern seen in the threshold sweep. At low load the dedicated branch saturates the distributed statistic, so both detectors sit near the ceiling and the centralized gain is small; at very high load both collapse toward the false-alarm floor as the hybrid waveform degenerates into superimposed-only probing, so the gap closes from the other side; the separation is widest at intermediate load. The reason is direct: as the number of communication streams increases, the minimum payload power  $P_{d,j}^{\min}[t]$  needed to satisfy the SINR protection consumes nearly the full per-AP budget, and (7.48) leaves little or no residual dedicated sensing power. The hybrid waveform then degenerates toward the superimposed-only waveform.

**STRESS-TEST BOUNDARY.** The degradation observed across both panels must be read against the fundamental operating premise of Cell-Free Massive MIMO, which is that the number of active UEs is much smaller than the number of APs,  $K \ll M$ . For the baseline deployment with  $M = 36$  APs, this premise erodes progressively as  $K$  approaches  $M$ : at  $K = 30$ – $36$ , the available pilot sequences are already reused three to four times on average and the user-centric clusters of different UEs overlap across a large share of the AP pool. The load range  $K \geq M = 36$  in Figure 9.4 should therefore be read as a stress test rather than an intended operating point. The purpose of extending the sweep is to expose the shape of the degradation curve and to confirm that the rapid performance collapse at the right end of both panels is not a shortcoming of the proposed protocol, but the expected and predictable consequence of violating the distributed cooperation premise of Cell-Free Massive MIMO.

## 9.4 RESIDUAL INTER-AP INTERFERENCE PROTECTION

The DTDD architecture creates residual InAI, represented in the UL received signal model of (8.3) and in the directional covariance terms used by the protected receivers. The next analysis sweeps the residual AP-to-AP factor from  $-35$  to  $0$  dB and compares different protection choices.

### 9.4.1 Downlink precoder sweep

Figure 9.5 compares three DL precoders while keeping the UL combiner fixed to the protected local L-MMSE design. The first is the proposed protected LP-MMSE precoder, which includes the adaptive InAI penalty of (7.10). The second is the same local LP-MMSE structure without the InAI protection term. The third is MR precoding toward the served user directions.

At the lowest leakage level, all three precoders produce the same UL SE. The residual InAI is then below the effective post-combining noise floor, so the exact DL radiation pattern is not limiting the UL link. As the residual factor increases, the protected precoder rotates the DL beams away from the strongest protected InAI directions. This costs DL signal gain: the protected DL SE drops from  $40.34$  to  $37.73$  bit/s/Hz between  $-35$  and  $0$  dB. The unprotected LP-MMSE curve remains flat in DL, because its design does not depend on the residual InAI level, but this is precisely why its UL performance collapses. MR is less sensitive than the unprotected LP-MMSE solution at high leakage, but its lower multi-user suppression makes it inferior to the protected design in sum SE.

### 9.4.2 Uplink combiner sweep

Figure 9.6 performs the complementary experiment. The DL precoder is fixed to the protected LP-MMSE design, while the UL combiner is changed between protected local L-MMSE, ordinary local L-MMSE, and MR. Since the DL waveform is fixed, only the UL SE curves are informative.

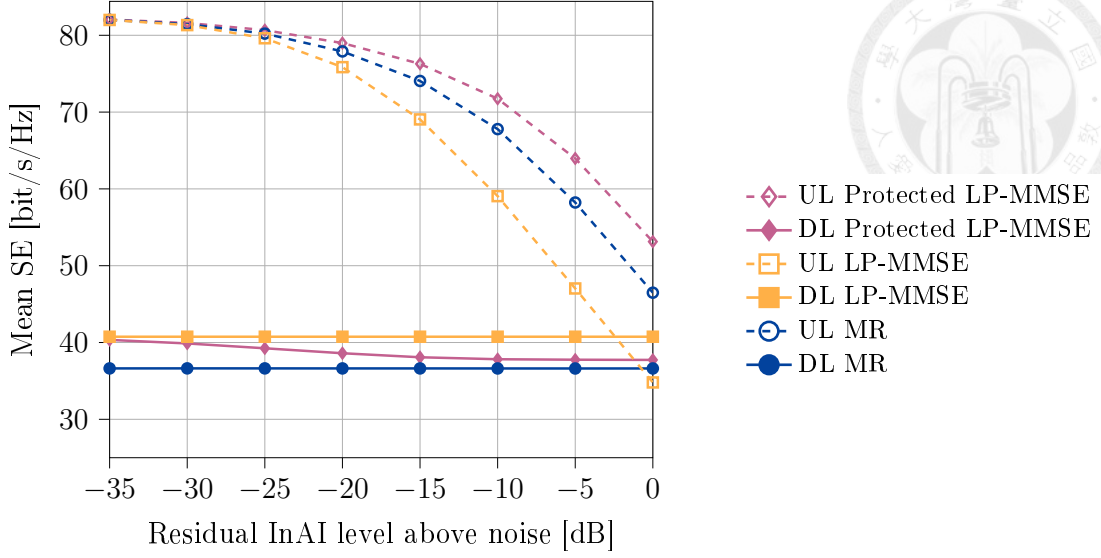


Figure 9.5: Effect of protected DL precoding under residual InAI. The UL combiner is fixed to the protected local L-MMSE design. The protected LP-MMSE precoder sacrifices a small amount of DL SE as the residual InAI level increases, but it preserves much more UL SE than the unprotected local LP-MMSE design.

The protected and ordinary L-MMSE curves overlap at  $-35$  dB. At  $0$  dB, the protected combiner obtains  $53.13$  bit/s/Hz, ordinary L-MMSE obtains  $44.32$  bit/s/Hz, and MR obtains  $37.05$  bit/s/Hz. The protection gain therefore grows with the residual InAI level. MR is structurally lower at weak leakage because it cannot suppress multi-user interference as effectively, but it loses less relative performance over the sweep because it does not rely on an inverted covariance model. Together, Figure 9.5 and Figure 9.6 show that DL-side and UL-side protection are complementary. The precoder reduces the interference emitted toward receiving APs, while the combiner suppresses the leakage that remains after scheduling and precoding.

#### 9.4.3 Trade-off analysis and operating conditions

Both experiments compare the proposed adaptive directional designs against benchmarks that either ignore the coupling direction, such as ordinary LP-MMSE and L-MMSE, or sacrifice multi-user suppression entirely, like MR. The performance gap between the protected and unprotected designs opens precisely when the residual coupling becomes large enough to make the unmodeled leakage a dominant nuisance term. The fact that the ordinary L-MMSE combiner is outperformed by the

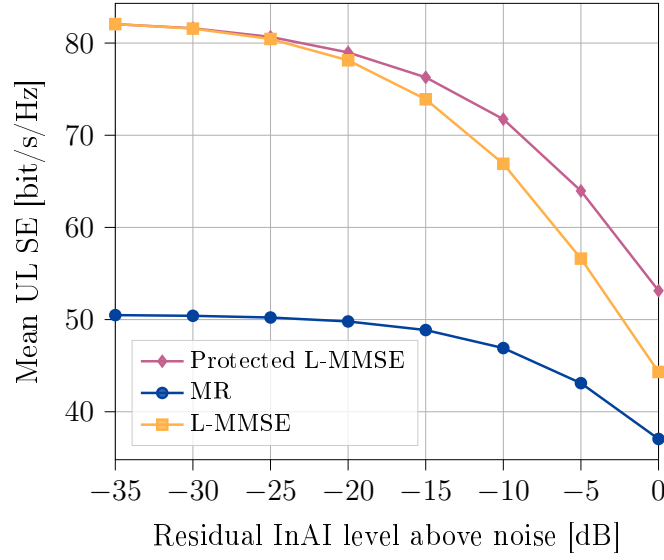


Figure 9.6: Effect of protected UL combining under residual InAI. The DL precoder is fixed to the protected LP-MMSE design. Ordinary local L-MMSE combining is nearly optimal when residual InAI is weak, but the protected version becomes necessary as the leakage level increases.

protected design, despite both having access to the same local channel estimates, demonstrates that the directional penalty adds spatial information beyond what a scalar noise floor can represent.

The DL SE loss visible in Figure 9.5 represents the beamforming gain sacrificed to keep energy away from the UL neighbors. The corresponding recovery in UL SE is about 18 bit/s/Hz at 0 dB residual coupling, so the sum SE is substantially improved whenever the leakage is strong. Crucially, the cost vanishes at low residual coupling: both the precoder and the combiner recover to their unprotected baselines automatically through the adaptive scalars  $\lambda_{IA,j}^{\text{eff}}[t]$  and  $\lambda_{IA,m}^{\text{UL,eff}}[t]$ .

## 9.5 PERFORMANCE OVER DIFFERENT DL AND UL SPLITS

The experiments so far fixed the communication traffic at an even split, with half of the UEs served in the DL and half in the UL. However, since the operating direction of each AP is decided from the local UL/DL traffic by the DTDD rule of Section 6.4, the traffic split controls indirectly how many APs illuminate the surveillance volume and how many receive its echoes.

Figure 9.7 sweeps the UL UE fraction

$$\phi_{\text{ul}} \triangleq \frac{K_{\text{u}}}{K}, \quad K_{\text{u}} = |\mathcal{U}_{\text{u}}[t]|, \quad K_{\text{d}} = |\mathcal{U}_{\text{d}}[t]| = K - K_{\text{u}}, \quad (9.9)$$

from 0.1 to 0.9, while keeping the baseline deployment and the distributed hybrid waveform fixed. The left axis reports the mean UL, DL, and sum SE; the right axis reports the unique target share  $\bar{S}_{\text{uniq}}$  of (9.8).

Two features organize the whole figure. First, the UL and DL SE curves cross at  $\phi_{\text{ul}} \approx 0.40$ , where both equal about 52.7 bit/s/Hz, and the retention curve reaches its maximum,  $\bar{S}_{\text{uniq}} = 0.955$ . The communication-balanced operating point and the sensing-optimal operating point therefore coincide, and they lie on the DL-leaning side. Second, communication and sensing pull in opposite directions, but asymmetrically: the sum SE rises almost monotonically toward the UL-heavy side, from 93 to 164 bit/s/Hz, whereas the retention curve is a plateau followed by a cliff, staying above 0.85 for every split up to  $\phi_{\text{ul}} = 0.60$  and then collapsing to near zero by 0.70–0.90. The remainder of this section traces both features to a single scheduling mechanism.

Every trend in Figure 9.7 originates in how the DTDD scheduler of Algorithm 6.3 maps the user-side fraction  $\phi_{\text{ul}}$  onto the AP mode partition. This mapping is strongly nonlinear. Indeed, between  $\phi_{\text{ul}} = 0.30$  and 0.40, the mean UL-AP count jumps from 2.7 to 8.2, because once  $K_{\text{u}}$  exceeds roughly ten of the twenty-four UEs, enough traffic neighborhoods become UL-dominated to flip a large block of APs at once. Below that threshold the network stays almost entirely in DL mode, and only the repair pass injects the handful of UL APs needed for voxel feasibility. This single transition accounts for the two inflection points seen in the figure: the jump in sum SE and the peak in retention at  $\phi_{\text{ul}} = 0.40$ .

COMMUNICATION SPECTRAL EFFICIENCY. While UL SE grows monotonically and the DL SE falls monotonically, the sum SE is *not* monotone. This is the most subtle feature of the communication curves, and it is explained by the same phase transition. In the DL-heavy regime each UL UE is crammed onto only two or three UL APs, so

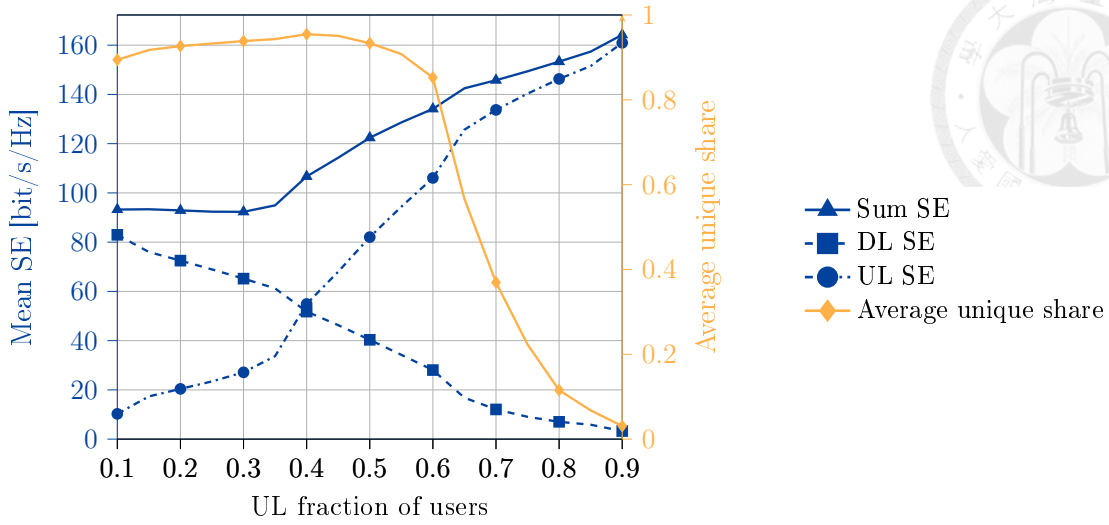


Figure 9.7: Network performance over the UL/DL traffic split. The left axis reports the mean UL, DL, and sum SE; the right axis reports the unique target share of (9.8). The UL and DL SE balance at  $\phi_{\text{ul}} \approx 0.40$ , which also maximizes retention. Tracking degrades gracefully toward the DL-heavy side but collapses for  $\phi_{\text{ul}} \geq 0.70$ . The sum SE rises almost monotonically toward the UL-heavy side.

adding UL UEs raises contention faster than diversity and the per-UL-user SE actually falls. Meanwhile, the abundant DL-AP set still serves the DL UEs comfortably, so converting a DL UE into a UL UE is close to SE-neutral and the sum stays flat. At  $\phi_{\text{ul}} = 0.40$  the partition transitions, the UL UEs finally receive genuine macro-diversity, and the sum SE jumps; from there the UL side keeps improving while the DL side still has enough APs, so the sum climbs.

SENSING AND TRACK RETENTION. The retention curve is high and flat across the whole DL-heavy and balanced range and then falls off a cliff. Three resource pressures peak together at  $\phi_{\text{ul}} = 0.40$  and explain the optimum. First, the scan budget is largest there. The per-slot request budget of (6.17) scales with the DL-AP count, and at  $\phi_{\text{ul}} = 0.40$  there are still about 28 DL APs, so target discovery is nearly complete. Second, echo-reception diversity becomes adequate. The roughly eight UL APs that appear after the phase transition give the fused detector real spatial diversity. Third, the dedicated sensing power is largest. With about half a DL UE per DL AP, the residual dedicated ceiling  $\bar{\beta}_j[t]$  of (7.48) and the dedicated spatial budget  $(N - |\mathcal{D}_j[t]|)^+$  of (7.40) are both maximized. The peak falls at  $\phi_{\text{ul}} = 0.40$

rather than at 0.50 because, at the even split, the DL-AP count has already dropped to about twenty, which shrinks both the scan budget and the dedicated power.

Toward the DL-heavy end the retention eases only from 0.955 to 0.894 and never collapses. Even with about two UL APs, the repair pass guarantees that every requested voxel keeps at least one echo receiver, and the many DL APs flood the scene with illumination, so targets are still discovered and mostly tracked. On the other hand, the cliff between  $\phi_{ul} = 0.60$  and 0.70 is qualitatively different. The DL-AP count falls from about 13 to 4 across this step, so the scan budget of (6.17) collapses and the fraction of targets that the scheduler can even request for scan falls steeply. The collapse is one-sided because the scheduler can manufacture an echo receiver for a voxel that lacks one, but it cannot manufacture illumination capacity from a DL-AP set that has effectively vanished.

OPERATING REGIMES AND DESIGN IMPLICATIONS. The two curves divide Figure 9.7 into three regions of the UL fraction  $\phi_{ul}$ . The DL-heavy region  $\phi_{ul} \leq 0.30$  is worse in both metrics: the split  $\phi_{ul} = 0.40$  gives higher sum SE and higher retention, so a traffic mix below it is purely a loss. The UL-heavy region  $\phi_{ul} \geq 0.70$  is a sensing dead zone. The usable trade-off band is therefore  $\phi_{ul} \in [0.40, 0.60]$ : across it, sensing is strongest at the UL-light end 0.40 and communication strongest at the UL-heavy end 0.60, where retention still exceeds 0.85. The even split used as the baseline elsewhere in this chapter,  $\phi_{ul} = 0.50$ , lies comfortably inside this band, although it is not the sensing optimum.

## 9.6 DETECTION PROBABILITY OVER TARGET STRENGTH

Figure 9.8 studies the sensing detector independently of communication-threshold and user-load sweeps. The protocol is fixed to the hybrid waveform with the default configuration, while the target RCS is swept from  $-40$  to  $10$  dBsm. For each RCS value, the distributed and centralized detectors are evaluated at  $P_{FA} \in \{10^{-2}, 10^{-3}, 10^{-4}\}$ .

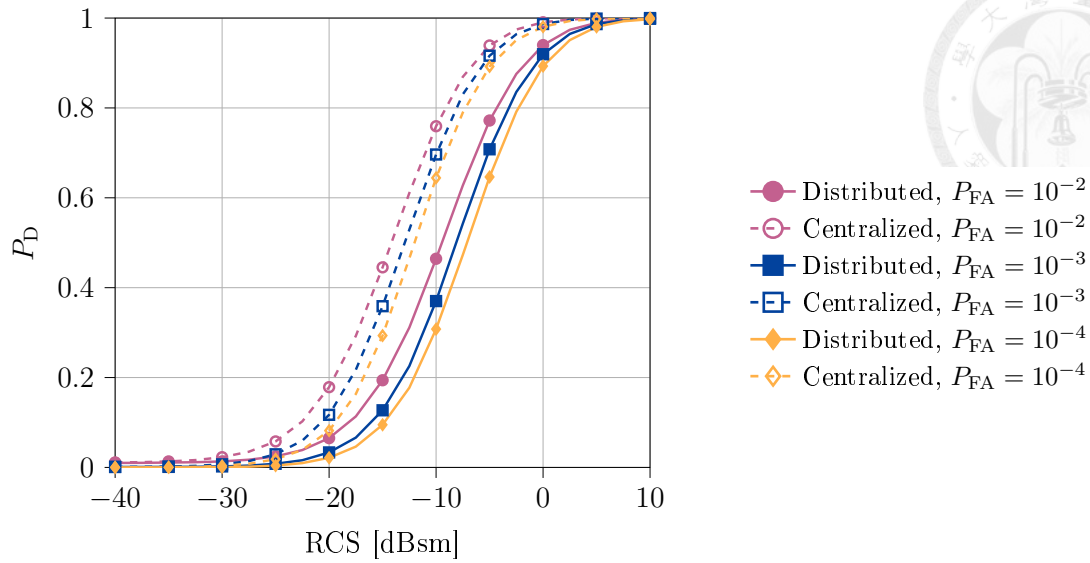


Figure 9.8: Detection probability over target RCS. The curves compare the distributed scalar-statistic GLRT with the centralized colored-covariance GLRT at three fixed false-alarm probabilities. Lower  $P_{FA}$  values require higher RCS to reach the same  $P_D$ .

For Figure 9.8, the denominator of (9.5) is restricted to target voxels that the precoder actually illuminated. This conditional view isolates the detector's intrinsic behavior from the precoder's coverage decisions, which would otherwise impose a ceiling below unity on every curve and obscure the detector-side trade-off that the figure is designed to study.

For very weak targets, roughly below  $-30$  dBsm, the statistic is dominated by the central chi-square behavior associated with (8.48) and the detector remains near the floor. For intermediate RCS values, the noncentrality created by the coherent voxel gain in (8.52) moves the statistic rapidly through the threshold region. For strong targets, roughly above  $-5$  dBsm, both detectors approach the ceiling. Reducing  $P_{FA}$  shifts the transition to the right because the threshold in (8.64) is higher.

The centralized detector is most useful in the transition region. At  $P_{FA} = 10^{-2}$ , the centralized gain over the distributed detector is about 0.032 at  $-25$  dBsm, 0.248 at  $-15$  dBsm, 0.305 at  $-10$  dBsm, and only 0.059 at 0 dBsm. The gap is therefore far from constant: it is negligible when the target is too weak for either detector and when the target is already easy for both, and it reaches roughly 0.3 in  $P_D$  where the distributed statistic sits on the steep part of its operating curve. This is the

same pattern observed in the threshold and load sweeps. This behavior is consistent with the established finding that well-designed distributed processing approaches centralizes performance in CF MMIMO.

The architectural conclusion is therefore that the distributed detector recovers most of the centralized performance in the regimes that matter for reliable surveillance, that is, when the target is strong enough or the network dense enough to saturate the fused statistic, while keeping the fronthaul benefits of local scalar reporting.

## 9.7 TRACK RETENTION UNDER MOBILITY AND DENSITY

The surveillance objective of the proposed protocol is persistent monitoring of moving low-altitude drones, as described in Section 1.3.2. The tracking experiment in Figure 9.9 therefore sweeps the mean drone speed and the number of simultaneously present drones, then reports the unique-share metric of (9.8), which is the end-to-end metric most directly aligned with the persistent surveillance goal. A drone contributes to the metric only if it is retained as a separate target, not merely detected somewhere in the grid. The heatmap of Figure 9.9 is generated under the fixed-heading motion model of Section 9.1.1, in which each drone flies a straight line at constant heading, as in a low-agility, fixed-wing or glider-like class. Section 9.7.2 then repeats the experiment under the random-heading model, which represents an agile multi-rotor drone, and compares the two.

### 9.7.1 Retention under the fixed-heading model

The heatmap of Figure 9.9 and the two degradation trends discussed in this subsection refer to the fixed-heading model, in which each drone cruises in a straight line.

At fixed target count, increasing speed reduces the retained unique share. The reason is the displacement accumulated over a revisit cycle. With  $T_{\text{slot}} = 200 \cdot 71.4 \mu\text{s}$

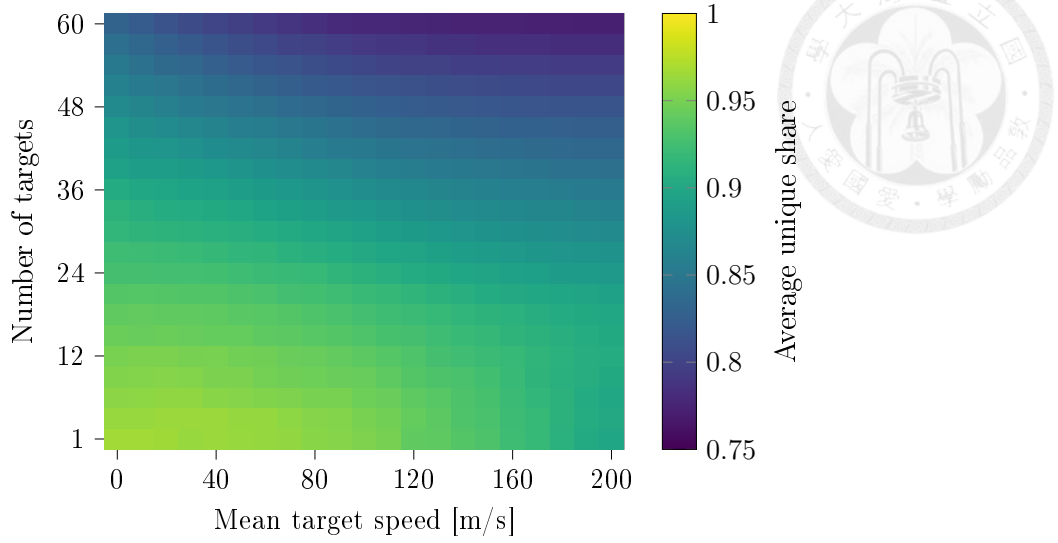


Figure 9.9: Target-retention heatmap under motion and density, for the fixed-heading (cruising) drone model. The color represents the average unique target share in the tail observation window. High values indicate that targets are both detected and kept separated by the track state, while low values indicate track loss, target merging, or both. The random-heading (agile) model yields a visually interchangeable heatmap; the two are compared in Section 9.7.2.

= 14.28 ms and a 40-slot revisit window, one revisit cycle lasts about 0.571 s. A target moving at 20 m/s therefore travels about 11.4 m over one revisit cycle, while a target moving at 200 m/s travels about 114 m. These distances are comparable with, or larger than, the coarse-voxel widths encountered by the adaptive grid. When a target crosses out of the support region expected by the track pointer before its next revisit, the scan list must reacquire it, and the target may temporarily lose its unique status. The heatmap therefore degrades gradually along the speed axis.

At fixed speed, increasing the number of targets also reduces the unique share. This is a grid-resolution effect combined with finite sensing budget. If  $N_{\text{tar}}$  targets are distributed over an effective set of  $V_{\text{eff}}$  distinguishable support cells, the expected number of unordered same-cell target pairs is approximately

$$\mathbb{E}\{C_{\text{pair}}\} \approx \frac{N_{\text{tar}}(N_{\text{tar}} - 1)}{2V_{\text{eff}}}. \quad (9.10)$$

This quadratic scaling explains why density becomes a strong limitation even when targets are stationary. If two targets occupy the same fine support region, the coherent detector reports a single occupied voxel and the assignment layer cannot

reliably maintain two unique tracks. This is the irreducible resolution–revisit trade-off behind the adaptive grid: a finer grid reduces the collision probability in (9.10), but it also increases the number of voxels that must be revisited within the same budget.

The upper-right region of Figure 9.9 combines the two failure modes. High target density increases the number of ambiguous support cells, while high speed increases the chance that a target leaves the expected support before the next scheduled revisit. The bottom-left region, with few targets and moderate speeds, remains close to full retention. The result also explains why the adaptive grid developed in Section 6.2 and validated in Appendix B is necessary: persistent surveillance depends not only on detector  $P_D$ , but also on whether the grid controller allocates enough revisit budget to maintain already detected targets while still discovering new ones.

### 9.7.2 Comparison with the random-heading model

To assess the protocol against a far more maneuverable drone, and to address whether smoother trajectories should track better, the same experiment is repeated under the random-heading model of Section 9.1.1: each drone re-draws its heading uniformly in  $[0, 2\pi)$  at random instants, spaced on average 40 slots apart, and holds it between changes. The speed model is identical, so the two configurations differ only in how the heading evolves within a drop.

Across the entire operating region the two models produce practically equivalent retention: the unique share differs by less than one percentage point (pp) at every swept grid point, and the heatmap structure is identical, so a single heatmap represents both. The residual difference is a small, systematic advantage of the random-heading model that grows with speed. First, at zero speed the two models are physically identical, because heading is irrelevant when the drone does not move. Second, averaging the retention difference  $\Delta \triangleq \bar{S}_{\text{uniq}}^{\text{rand}} - \bar{S}_{\text{uniq}}^{\text{fixed}}$  over four representative

target counts at different speed shows a monotone growth, from the noise floor up to about 0.7 pp, as shown in Table 9.2.

In Table 9.3, the same difference is averaged over speed at different target count, and is largest for a single drone and roughly halves once the network is congested. Heading is therefore a modulation on top of the first-order speed-by-density structure, not a primary driver of retention.

The reason of this effect is specific to the proposed tracker. The revisit gate of (6.14) is direction-agnostic: it revisits the last informative voxel together with an isotropic neighborhood of fine children and adjacent coarse voxels. The relevant geometric quantity is not the path length but the net displacement between successive informative revisits. For a fixed speed the two models sweep the same path length, but they differ in net displacement. The fixed-heading drone moves ballistically, so its distance from the last pinned voxel grows linearly in time and it tends to out-run its neighborhood. The random-heading drone performs a random walk, whose net displacement grows only as the square root of time, so it stays closer to where the track last localized it.

This separation explains the speed trend of Table 9.2. The ballistic net displacement grows linearly with speed while the neighborhood reach is fixed, so the fraction of fixed-heading drones that escape their support before the next revisit rises faster than for the random walk. The monotone-in-speed difference matches this mechanism.

The proposed scheduler-tracker is deliberately direction-agnostic, a choice made for scalability and fronthaul economy, so predictability confers no benefit and the operative axis is spatial confinement, under which the random walk is marginally favorable. The practical implication is that the protocol is robust to maneuvering, and its retention does not degrade when drones are agile rather than cruising. The cost is that it does not exploit trajectory smoothness, which would enhance the retention gain.

| MEAN SPEED (M/S)   | 0     | 40    | 80    | 120   | 160   | 200   |
|--------------------|-------|-------|-------|-------|-------|-------|
| Mean $\Delta$ (pp) | -0.04 | +0.35 | +0.44 | +0.66 | +0.69 | +0.65 |

Table 9.2: Mean retention difference  $\Delta = \text{random} - \text{fixed heading}$ , averaged over target counts at each speed. The  $v = 0$  column is the Monte-Carlo noise floor.

| NUMBER OF DRONES   | 1     | 24    | 48    | 60    |
|--------------------|-------|-------|-------|-------|
| Mean $\Delta$ (pp) | +0.85 | +0.29 | +0.35 | +0.33 |

Table 9.3: Mean retention difference  $\Delta = \text{random} - \text{fixed heading}$ , averaged over speed at drone count. The advantage is largest for a single drone and is compressed by congestion.

## 9.8 FIXED HARDWARE BUDGET

The last experiment shown in Figure 9.10 studies the role of the Cell-Free deployment itself. The total antenna budget is fixed at  $MN = 576$ , and the split between the number of APs and the number of antennas per AP is swept. The hybrid protocol and distributed UL receiver are used, with  $K = 24$  fixed. This experiment separates two possible sources of gain: per-AP array gain, which increases with  $N$ , and multi-static/macroscopic diversity, which increases with the number of geographically distributed APs.

The sensing result shows a clear interior optimum. The distributed  $P_D$  peaks near  $(M, N) = (48, 12)$ , then falls monotonically. The two-sided shape exposes the joint role of per-AP DoF and multi-static density that the fixed-budget trades off. On the left, very small arrays starve the dedicated sensing branch. The admission of Section 7.4 caps the dedicated beam count at  $N - |\mathcal{D}_j|$ , and the protected null-space penalty consumes further dimensions to orthogonalise against the strongest receiving APs, so each AP can only illuminate a handful of voxels. A large fraction of the target voxels in the request set therefore never receives any sensing power, and under the unconditional  $P_D$  of (9.5) they count as misses. On the right, very few APs collapse the multi-static geometry. The distributed GLRT in (8.57) loses both noncentrality contributions and viewing diversity. The peak at  $(M, N) = (48, 12)$  is the operating point where the dedicated sensing branch has enough local DoF to

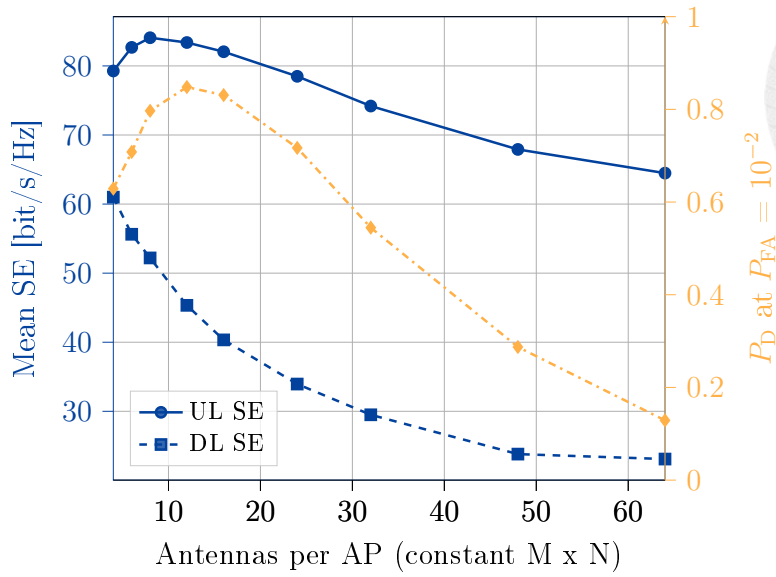


Figure 9.10: Performance over a fixed total antenna budget  $MN = 576$ . The left axis reports DL and UL SE; the right axis reports  $P_D$  at  $P_{FA} = 10^{-2}$ . The results show that distributing the antennas over many APs is more valuable for this ISAC task than concentrating them into a few large arrays.

admit many voxels and the network still has enough geographically distributed APs to cover the surveillance area with overlapping bistatic links. This conclusion is consistent with the general motivation for CF-ISAC: distributed multi-static sensing provides spatial diversity and coverage that a co-located architecture cannot easily reproduce [10, 15, 22].

The communication curves tell a related but not identical story. The DL SE decreases from 61.00 bit/s/Hz at (144, 4) to 23.09 bit/s/Hz at (9, 64). The dominant effect is macro-diversity: with many APs, most UEs have at least one nearby serving AP; with few APs, many UEs are far from all transmitters. The UL SE is non-monotone. This reflects a balance between local receive dimensions and macro-diversity. Small arrays provide many receiving sites but limited local interference suppression; very large arrays provide strong local combining but too few receiving sites.

The design implication is that the total antenna count alone is not a sufficient descriptor of a CF-ISAC deployment. DL communication is monotonically improved by spreading the antennas across more APs, but UL communication and sensing both

prefer a moderate split. The two optima nearly coincide because the underlying mechanisms align: UL combining benefits from enough per-AP antennas to suppress local interference while keeping a large number of receive sites for macro-diversity, and the hybrid sensing branch benefits from enough per-AP DoF to admit dedicated voxels while keeping a large number of bistatic transmit/receive pairs. The default configuration  $(M, N) = (36, 16)$  sits close to this optimum; it trades a small amount of  $P_D$  and UL SE for a deployment that is also reasonable in terms of AP count, fronthaul fan-out, and computational practicality.





# 10 | CONCLUSIONS AND FUTURE WORK

This chapter concludes the thesis by revisiting its central design problem: whether a CF MMIMO network can serve as a scalable ISAC platform for low-altitude drone surveillance, without abandoning the communication service for which the network is primarily deployed. The previous chapters addressed this problem constructively. The purpose of this final chapter is to extract the main takeaways of the work, clarify the limitations of the present implementation, and indicate future research directions that follow naturally from the proposed protocol architecture.

The main message is that drone surveillance in a DTDD CF-ISAC system requires a complete execution chain: a scan controller that decides which voxels deserve attention, a duplexing scheduler that creates simultaneous DL illumination and UL reception opportunities, a protected DL design that preserves communication feasibility, an UL receiver that remains robust to residual InAI, and a detector that compresses multi-static sensing evidence without creating unbounded fronthaul. This thesis has shown that such a chain can be organized in a scalable way, while still producing a meaningful sensing–communication trade-off.

## 10.1 MAIN CONCLUSIONS

The first conclusion is that the CF architecture is not only useful for communication coverage, but also fundamentally well matched to the sensing geometry of low-altitude surveillance. This is consistent with the broader motivation for CF-ISAC, where distributed infrastructure is expected to improve sensing coverage, robustness, and target-centric operation compared with purely co-located deployments [10]. The

fixed-hardware-budget results of Chapter 9 made this point concrete: for the considered surveillance task, spreading a fixed number of antennas over many APs was more valuable for  $P_D$  than concentrating the same antennas into a few large arrays.

The second conclusion is that opportunistic payload sensing alone is not enough for reliable drone surveillance in the considered geometry. The Opportunistic sensing baseline keeps all DL power on communication payload beams, and the numerical results showed that its  $P_D$  remains close to the false-alarm floor. This does not mean that communication signals are useless for sensing. Recent ISAC work has shown that data payloads contain sensing information and should not be discarded a priori [13, 14]. However, when the beams are selected only for UEs, their incidental illumination of a three-dimensional drone-surveillance grid is too irregular to support a stable voxel-wise GLRT. The network must intentionally shape at least part of the transmitted field toward the sensing task.

The third conclusion is that the hybrid DL design is the right compromise between communication-native sensing and dedicated sensing. Superimposed deterministic probing is attractive because it reuses the communication stream and introduces known sequences into the payload window. However, the probing fraction is necessarily limited by the communication SINR constraint. Dedicated sensing beams are more powerful because they can be directed toward requested voxels, but they must consume residual power and spatial DoF after communication feasibility has been paid. The proposed hybrid design combines these two mechanisms: it first tries to extract sensing value from payload directions and then activates dedicated beams only when the local power and null-space conditions allow it.

The fourth conclusion is that DTDD is useful but only if CLI is treated as a structural part of the design. The benefit of DTDD is that UL reception and DL transmission can coexist in the same time-frequency resource using spatially separated half-duplex APs, which is a promising alternative to full-duplex hardware for joint bi-directional communication and sensing [15]. However, residual InAI can dominate the UL receiver if the DL beams radiate strongly toward active receiving

APs, and it can also corrupt the sensing matched filters. The protected LP-MMSE precoder and the protected local L-MMSE combiner are therefore necessary to keep the UL link stable as the residual AP-to-AP leakage increases.

The fifth conclusion concerns the receiver split. The centralized raw-uplink receiver is an important upper benchmark because it can exploit cross-AP covariance information before scalar compression. On the other hand, the distributed receiver reduces the received antenna-domain block into soft communication symbols and scalar detection statistics, so that the CPU fuses low-dimensional summaries rather than raw samples. The results showed that the gap between the two architectures is regime-dependent: centralized detection delivers a substantial gain in the marginal RCS band and in the steep transition region of the detector, whereas the distributed GLRT approaches it closely once the waveform provides sufficient illumination to saturate the fused statistic. The distributed receiver is therefore a viable alternative in the relevant regimes that matter for reliable surveillance, which is what justifies it as the scalable default.

Finally, the thesis showed that persistent surveillance is more demanding than one-slot detection. The tracking heatmap in Chapter 9 demonstrated that target retention depends jointly on  $P_D$ , target density, target speed, grid resolution, and revisit time. A detector can be strong in a single slot and still fail at the surveillance level if the grid controller cannot revisit targets quickly enough, or if two nearby targets collapse into the same support region. The adaptive coarse/fine grid is therefore the mechanism that connects radio-layer sensing to the operational objective of maintaining separate moving targets over time.

## 10.2 LIMITATIONS OF THE PRESENT FRAMEWORK

The proposed framework is intentionally structured and scalable, but it still relies on simplifying assumptions that leave room for future work. The most important limitation is that the present protocol is mainly detection-driven. The state of a target is represented through voxel occupancy and track support, while richer

parameters such as velocity, direction of motion, RCS and target extent are not estimated explicitly. This is appropriate for a first scalable surveillance protocol, but it does not exploit all the information contained in a coherent multi-static observation. Classical radar processing and estimation theory show that echo measurements can support richer parameter estimation when the waveform and observation model are sufficiently informative [23, 24, 26].

A second limitation is the power policy. In the current design, active DL APs use the available power budget either for payload streams, superimposed probing, or dedicated sensing beams. This is useful for exploring the sensing–communication frontier, but it is not energy-aware. If a target is close, has high RCS, or is already detected with high reliability, additional transmit power may produce little operational benefit because the detector is already in the saturated part of the ROC curve. Similarly, if the communication load is light, the network may not need to illuminate with full power in every slot to maintain both QoS and track retention.

A third limitation concerns environmental realism. The simulator includes residual InAI, InUI, target motion, random layouts, and a voxel hierarchy, but clutter and propagation uncertainty are still simplified compared with a real urban low-altitude environment. In practice, reflections from buildings, ground, vegetation, vehicles, and non-target aerial objects can produce structured target-free returns. Recent CF-ISAC work explicitly models clutter or target-free channels and shows that they affect detection and power allocation [10, 18]. A deployment-oriented version of the protocol should therefore learn and update a background map, rather than treating the nuisance covariance as a mostly static descriptor.

A fourth limitation concerns the performance references used in the evaluation. The detection results are reported against operational benchmarks rather than against a fundamental performance limit. For the sensing layer, a canonical limit is hard to define, because the protocol is detection-driven rather than estimation-driven and because sensing operates on a best-effort residual of a communication-constrained allocation.

### 10.3 FUTURE WORK

The previous limitations suggest several research directions. The most natural extension is to make the protocol energy-aware. Instead of using the full AP power budget whenever it is available, the system could minimize the transmit energy needed to satisfy communication and sensing requirements. This direction connects the thesis protocol with the broader ISAC literature on power allocation under communication constraints [15, 18] and with the energy-efficiency motivation of future CF-ISAC networks [10].

A second future direction is to move from detection to joint detection, estimation, and tracking. A richer receiver could attach a state vector to each target,

$$\mathbf{s}_i[t] = [\mathbf{p}_i^T[t], \mathbf{v}_i^T[t], \sigma_i[t]]^T, \quad (10.1)$$

where  $\mathbf{p}_i[t]$  is position,  $\mathbf{v}_i[t]$  is velocity, and  $\sigma_i[t]$  is an effective RCS or reflectivity parameter. Velocity could be inferred from temporal evolution of voxel detections and, in wider-band or longer-coherent-processing configurations, from Doppler-sensitive matched filters. The RCS estimate could then be used not only as a target descriptor, but also as a feedback variable for the sensing scheduler. In the DTDD CF-ISAC model, the average sensing SCNR depends on the RCS statistics, the selected DL precoders, the AP scheduling, and the residual nuisance terms [15]. The same idea can be adapted to the present voxel model: once the network has an estimate of target strength and nuisance power, it can predict which target-voxel pairs require additional probing and which can be maintained with reduced power.

This naturally leads to a third direction: detection-probability-aware scheduling. Under the Gaussian coherent model used in Chapter 8, the voxel statistic has a central chi-square distribution under  $\mathcal{H}_0$  and a noncentral chi-square distribution under  $\mathcal{H}_1$ . If  $\lambda_v[t]$  denotes the predicted noncentrality parameter obtained from an SCNR or RCS estimate, then a local prediction of the detection probability can be written as

$$P_{D,v}[t] = 1 - F_{\chi^2_{\lambda_v[t]}}(\eta_v[t]), \quad (10.2)$$

where  $F_{\chi^2(\lambda_v[t])}(\cdot)$  is the cumulative distribution function of a noncentral chi-square random variable with  $\nu$  degrees of freedom and  $\eta_v[t]$  is the CFAR threshold associated with the target  $P_{\text{FA}}$ . Equation (10.2) would allow the scheduler to reason in terms of predicted sensing utility instead of only geometric coverage. For example, the scan controller could allocate more revisit opportunities to voxels with low predicted  $P_{\text{D}}$ , high track uncertainty, or high collision probability, while reducing the budget assigned to easy targets.

A complementary and more theoretical direction is the characterization of a fundamental sensing–communication performance limit for the proposed setting. Recent work has begun to formalize such limits through the CRB–rate region and related sensing mutual-information metrics [52], characterizing the Pareto trade-off between communication rate and an estimation-theoretic sensing accuracy. These results are, however, currently established for point-to-point or co-located MIMO models, and they measure sensing through estimation error rather than through detection probability. The protocol developed in this thesis is detection-driven, operates over a multi-static DTDD CF deployment, and treats sensing as a best-effort residual of a communication-constrained allocation. Extending the fundamental-limit analysis to this case remains an open problem and a natural continuation of this work.

A fifth direction concerns learning-assisted control. The proposed protocol already separates slow-timescale descriptors from slot-level execution, which creates a natural interface for data-driven methods. Learning should not replace the physical-layer rules that guarantee QoS, CFAR operation, and bounded complexity. However, it could improve the slow control objects: grid-size adaptation, bridge-request priorities, voxel-reuse distances, track-death parameters, and power-backoff policies. A promising approach would be to keep the analytically derived communication and detection constraints as safety filters, while a learning module proposes candidate scan priorities or long-term resource profiles. This is aligned with the emerging

role of Machine Learning in CF-ISAC, but it must be constrained by the scalability principles of user-centric CF processing [10, 12].

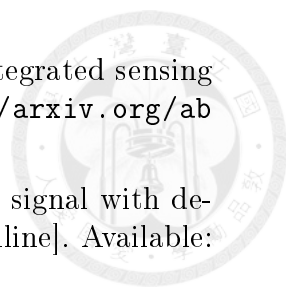
Finally, the framework should be validated beyond the present MATLAB simulator. A first step would be a more realistic channel and clutter simulator with imperfect CSI, finite-resolution fronthaul quantization, synchronization errors, and measured or ray-traced low-altitude propagation. A second step would be a small-scale testbed with a limited number of distributed radios, where the proposed distributed GLRT can be compared with a centralized raw-sample receiver under controlled drone-like targets. Such validation is important because the thesis has focused on protocol structure and performance trends, not on hardware impairments. Nevertheless, the explicit separation between slow control, local AP processing, and CPU fusion should make the proposed architecture suitable for gradual implementation.





## BIBLIOGRAPHY

- [1] Ericsson, "Ericsson mobility report november 2025," Tech. Rep., Nov. 2025. [Online]. Available: <https://www.ericsson.com/4aca6f/assets/local/reports-papers/mobility-report/documents/2025/ericsson-mobility-report-november-2025.pdf>
- [2] Z. Qadir et al., "Towards 6g internet of things: Recent advances, use cases, and open challenges," Nov. 2021.
- [3] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5g wireless networks: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 18, pp. 1617–1655, 3 Feb. 2016, ISSN: 1553-877X. DOI: 10.1109/COMST.2016.2532458
- [4] J. Parikh and A. Basu, "Lte advanced the 4g mobile broadband technology," *International Journal of Computer Applications*, vol. 13, pp. 17–21, 5 Jan. 2011, ISSN: 09758887. DOI: 10.5120/1776-2449
- [5] S. Dang et al., "What should 6g be?" *Nature Electronics*, vol. 3, no. 1, pp. 20–29, Jan. 2020, ISSN: 2520-1131. DOI: 10.1038/s41928-019-0355-6 [Online]. Available: <http://dx.doi.org/10.1038/s41928-019-0355-6>
- [6] M. Giordani et al., "Toward 6g networks: Use cases and technologies," *IEEE Communications Magazine*, vol. 58, pp. 55–61, 3 Mar. 2020, ISSN: 0163-6804. DOI: 10.1109/MCOM.001.1900411
- [7] M. H. Alsharif et al., "Sixth generation (6g) wireless networks: Vision, research activities, challenges and potential solutions," *Symmetry*, vol. 12, p. 676, 4 Apr. 2020, ISSN: 2073-8994. DOI: 10.3390/sym12040676
- [8] Y. Zhao et al., "A comprehensive survey of 6g wireless communications," Feb. 2021.
- [9] Y. Jiang et al., "6g non-terrestrial networks enabled low-altitude economy: Opportunities and challenges," May 2024.
- [10] D. Galappaththige et al., "Cell-free integrated sensing and communication: Principles, advances, and future directions," Mar. 2025. [Online]. Available: <http://arxiv.org/abs/2502.20345>
- [11] H. Q. Ngo et al., "Ultradense cell-free massive mimo for 6g: Technical overview and open questions," *Proceedings of the IEEE*, vol. 112, pp. 805–831, 7 Jul. 2024, ISSN: 0018-9219. DOI: 10.1109/JPROC.2024.3393514
- [12] E. Bjornson and L. Sanguinetti, "Scalable cell-free massive mimo systems," *IEEE Transactions on Communications*, vol. 68, pp. 4247–4261, 7 Jul. 2020, ISSN: 15580857. DOI: 10.1109/TCOMM.2020.2987311

- 
- [13] C. Xu et al., “Exploiting both pilots and data payloads for integrated sensing and communications,” Jun. 2025. [Online]. Available: <http://arxiv.org/abs/2506.15998>
  - [14] L. Xie et al., “Sensing mutual information for communication signal with deterministic pilots and random data payloads,” Jan. 2026. [Online]. Available: <http://arxiv.org/abs/2601.11149>
  - [15] A. Chowdhury, S. S. Thoota, and E. G. Larsson, “Joint sensing and bi-directional communication with dynamic tdd enabled cell-free mimo,” Aug. 2025. DOI: 10.1109/TWC.2025.3596622 [Online]. Available: <http://arxiv.org/abs/2508.03460><http://dx.doi.org/10.1109/TWC.2025.3596622>
  - [16] H. Q. Ngo et al., “Cell-free massive mimo versus small cells,” Jan. 2017.
  - [17] U. T. Demir, E. Björnson, and L. Sanguinetti, “Foundations of user-centric cell-free massive mimo,” *Foundations and Trends in Signal Processing*, vol. 14, pp. 162–472, 3-4 Jan. 2021, ISSN: 1932-8346. DOI: 10.1561/2000000109
  - [18] Z. Behdad et al., “Multi-static target detection and power allocation for integrated sensing and communication in cell-free massive mimo,” Mar. 2024. [Online]. Available: <http://arxiv.org/abs/2305.12523>
  - [19] U. Demirhan and A. Alkhateeb, “Cell-free isac mimo systems: Joint sensing and communication beamforming,” Feb. 2024.
  - [20] M. Elfiatoure et al., “Multiple-target detection in cell-free massive mimo-assisted isac,” *IEEE Transactions on Wireless Communications*, vol. 24, pp. 4283–4298, 5 May 2025, ISSN: 1536-1276. DOI: 10.1109/TWC.2025.3542206
  - [21] M. Mohammadi et al., “Virtually full-duplex cell-free massive mimo with access point mode assignment,” May 2022. [Online]. Available: <http://arxiv.org/abs/2205.06403>
  - [22] P. Jopanya, D. P. M. Osorio, and E. G. Larsson, “Drone surveillance via coordinated beam sweeping in mimo-isac networks,” Jan. 2026. [Online]. Available: <http://arxiv.org/abs/2601.08483>
  - [23] M. A. Richards, *Fundamentals of radar signal processing*. McGraw-Hill, 2014, ISBN: 9780071798327.
  - [24] M. Skolnik, *Radar Handbook, Third Edition*. McGraw-Hill Publishing, 2008, p. 1329, ISBN: 9780071485470.
  - [25] S. M. Kay, *Fundamentals of statistical signal processing : volume 1 Estimation Theory*. Prentice-Hall PTR, Jan. 1998, ISBN: 9780135041352.
  - [26] H. L. V. Trees, *Detection, Estimation, and Modulation Theory, Part I*. John Wiley & Sons, Inc., Sep. 2001, ISBN: 0471095176. DOI: 10.1002/0471221082
  - [27] T. L. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” *IEEE Transactions on Wireless Communications*, vol. 9, pp. 3590–3600, 11 Nov. 2010, ISSN: 15361276. DOI: 10.1109/TWC.2010.092810.091092
  - [28] T. L. Marzetta, E. G. Larsson, and H. Yang, *Fundamentals of massive MIMO*. Cambridge University Press, 2016, p. 242, ISBN: 9781316799895.

- [29] E. Björnson, J. Hoydis, and L. Sanguinetti, “Massive mimo networks: Spectral, energy, and hardware efficiency,” *Foundations and Trends in Signal Processing*, vol. 11, pp. 154–655, 3-4 2017, ISSN: 1932-8346. DOI: 10.1561/20000000093
- [30] T. S. Rappaport, *Wireless communications*. Prentice Hall PTR, 2001, pp. 163–166, ISBN: 9780130422323.
- [31] J. Zhang et al., “Cell-free massive mimo: A new next-generation paradigm,” *IEEE Access*, vol. 7, pp. 99 878–99 888, 2019, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2019.2930208
- [32] R. W. H. Jr and A. Lozano, *Foundations of MIMO Communication*. Cambridge University Press, Dec. 2018, ISBN: 9781139049276. DOI: 10.1017/9781139049276
- [33] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, “Energy and spectral efficiency of very large multiuser mimo systems,” *IEEE Transactions on Communications*, vol. 61, pp. 1436–1449, 4 Apr. 2013, ISSN: 0090-6778. DOI: 10.1109/TCOMM.2013.020413.110848
- [34] A. Ashikhmin, L. Li, and T. L. Marzetta, “Interference reduction in multi-cell massive mimo systems with large-scale fading precoding,” *IEEE Transactions on Information Theory*, vol. 64, pp. 6340–6361, 9 Sep. 2018, ISSN: 0018-9448. DOI: 10.1109/TIT.2018.2853733
- [35] R. Müller, L. Cottatellucci, and M. Vehkaperä, “Blind pilot decontamination,” Feb. 2014. DOI: 10.1109/JSTSP.2014.2310053
- [36] J. Ma and L. Ping, “Data-aided channel estimation in large antenna systems,” in *2014 IEEE International Conference on Communications (ICC)*, IEEE, Jun. 2014, pp. 4626–4631, ISBN: 978-1-4799-2003-7. DOI: 10.1109/ICC.2014.6884051
- [37] G. Interdonato, P. Frenger, and E. G. Larsson, “Scalability aspects of cell-free massive mimo,” in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, IEEE, May 2019, pp. 1–6, ISBN: 978-1-5386-8088-9. DOI: 10.1109/ICC.2019.8761828
- [38] E. Nayebe et al., “Precoding and power optimization in cell-free massive mimo systems,” *IEEE Transactions on Wireless Communications*, vol. 16, pp. 4445–4459, 7 Jul. 2017, ISSN: 1536-1276. DOI: 10.1109/TWC.2017.2698449
- [39] F. Li et al., “Scalable cell-free massive mimo with multiple cpus,” *Mathematics*, vol. 10, p. 1900, 11 Jun. 2022, ISSN: 2227-7390. DOI: 10.3390/math10111900
- [40] M. I. Skolnik, *Introduction to Radar Systems*. McGraw-Hill Education, 2002, ISBN: 0072881380.
- [41] M. A. Richards, J. Scheer, and W. A. Holm, *Principles of modern radar. Volume I, Basic principles*. SciTech Pub., 2010, p. 924, ISBN: 1891121529.
- [42] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation*. Wiley, 2001, p. 558, ISBN: 9780471416555.
- [43] S. Naoumi et al., “High-resolution sensing in communication-centric isac: Deep learning and parametric methods,” Nov. 2025.
- [44] L. Xie et al., “Sensing mutual information with random signals in gaussian channels: Bridging sensing and communication metrics,” Feb. 2024.

- 
- [45] S. Xu et al., “Mutual information-oriented isac beamforming design for large dimensional antenna array,” *Electronics*, vol. 14, p. 2515, 13 Jun. 2025, ISSN: 2079-9292. DOI: 10.3390/electronics14132515
- [46] Y. Liu et al., “Next-generation multiple access for integrated sensing and communications,” *Proceedings of the IEEE*, vol. 112, pp. 1467–1496, 9 Sep. 2024, ISSN: 0018-9219. DOI: 10.1109/JPROC.2024.3449807
- [47] G. Cheng et al., “Optimal coordinated transmit beamforming for networked integrated sensing and communications,” *IEEE Transactions on Wireless Communications*, vol. 23, pp. 8200–8214, 8 Aug. 2024, ISSN: 1536-1276. DOI: 10.1109/TWC.2023.3346457
- [48] T. H. Cormen et al., *Introduction to algorithms*. MIT Press, 2009, p. 1292, ISBN: 9780262533058.
- [49] G. H. Golub and C. F. V. Loan, *Matrix computations*. The Johns Hopkins University Press, 2013, p. 756, ISBN: 9781421407944.
- [50] E. G. Larsson, “Massive synchrony in distributed antenna systems,” *IEEE Transactions on Signal Processing*, vol. 72, pp. 855–866, 2024, ISSN: 1053-587X. DOI: 10.1109/TSP.2024.3358618
- [51] E. Kelly, “An adaptive detection algorithm,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-22, pp. 115–127, 2 Mar. 1986, ISSN: 0018-9251. DOI: 10.1109/TAES.1986.310745
- [52] Y. Xiong et al., “On the fundamental tradeoff of integrated sensing and communications under gaussian channels,” *IEEE Transactions on Information Theory*, vol. 69, pp. 5723–5751, 9 Sep. 2023, ISSN: 0018-9448. DOI: 10.1109/TIT.2023.3284449



## ROC CURVES FOR PROTOCOL VARIANTS



This appendix complements the fixed-false-alarm detection results in Chapter 9 by reporting the full ROC curves of the three protocol variants in Figure A.1. The operating point is the default scenario of Chapter 9:  $M = 36$ ,  $N = 16$ ,  $K = 24$ , 20 targets, and communication protection threshold  $\Gamma^{\min} = 5$  dB. The same three DL sensing variants are used: Opportunistic sensing, Deterministic sensing, and Hybrid. For each variant, both the distributed scalar-statistic detector and the centralized colored-covariance detector are shown.

The purpose of the appendix is to verify that the ranking observed at  $P_{\text{FA}} = 10^{-2}$  in Figure 9.2 is not an artifact of one selected threshold. In radar detection, a fixed operating point can hide whether two detectors cross at other  $P_{\text{FA}}$  levels; the ROC curve avoids this by sweeping the decision threshold over the full statistic range [23, 24, 25]. For detector architecture  $\chi \in \{\text{dist}, \text{cent}\}$  and threshold  $\eta$ , the empirical curve is

$$\left( \hat{P}_{\text{FA}}^{\chi}(\eta), \hat{P}_{\text{D}}^{\chi}(\eta) \right), \quad \eta \in \mathbb{R}_+, \quad (\text{A.1})$$

where the two probabilities are computed as in (9.5)–(9.6) over the steady-state slots of (9.4). The curve therefore reports the threshold-dependent separability of the target-present and target-absent statistic distributions.

The distributed opportunistic sensing curve is essentially diagonal. This means that the statistic distribution under the target-present hypothesis is almost indistinguishable from the statistic distribution under the target-absent hypothesis. The result agrees with the flat Opportunistic  $P_{\text{D}}$  values in Figure 9.2: when  $\rho_{jn}[t] = 0$  and

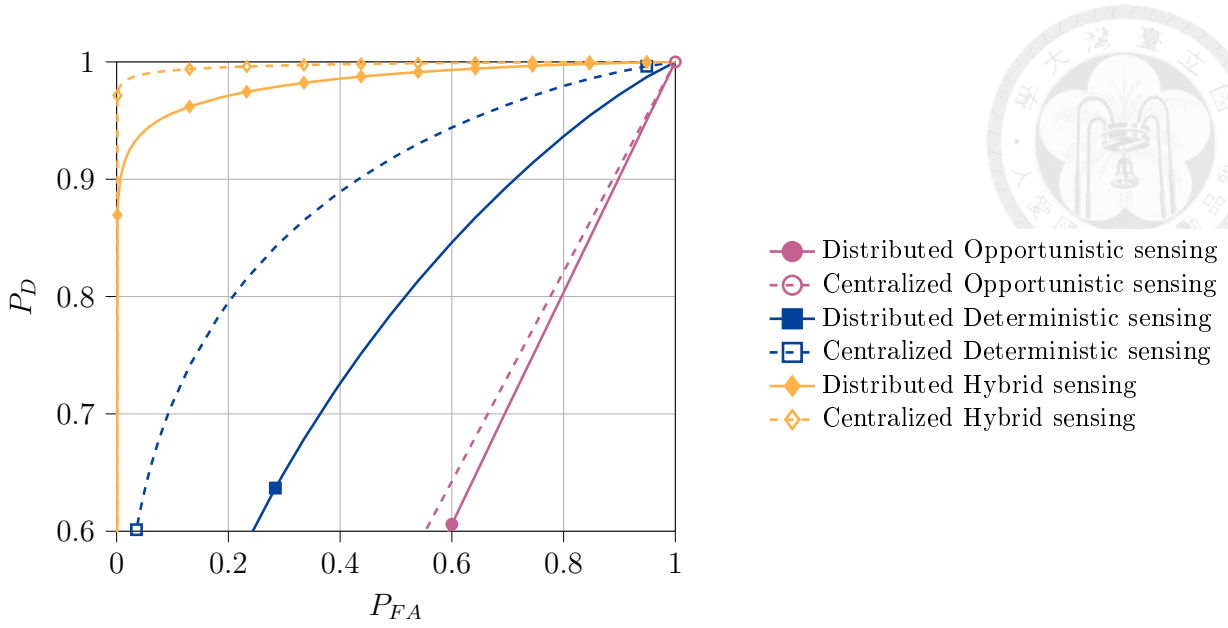


Figure A.1: Full ROC curves for the three protocol variants and the two UL receiver splits. Solid curves correspond to distributed scalar-statistic fusion, and dashed curves correspond to centralized colored-covariance fusion. The opportunistic sensing curve is close to the diagonal, the deterministic sensing curves show moderate noncentrality, and the hybrid curves remain near the top of the plot over a wide false-alarm range.

$\beta_{jv}[t] = 0$ , the detector sees only incidental payload illumination, and the average noncentrality in (8.52) is too small to produce reliable discrimination.

The Centralized Opportunistic curve lies above the distributed payload-only curve. However, the gain remains opportunistic and geometry-dependent. It is visible in the ROC curve, but it is much weaker than the gain created by explicit deterministic sensing resources.

The Deterministic sensing curves occupy the middle region of the plot. The deterministic signatures attached to payload beams create a real target-present shift, so the ROC is well above the Opportunistic baseline. At the same time, the probing fraction is constrained by the DL SINR protection logic of Section 7.3; it cannot grow arbitrarily without damaging communication. The resulting curve has the shape expected from a moderate-noncentrality coherent GLRT: increasing the false-alarm tolerance improves  $P_D$ , but the curve does not immediately saturate.

The hybrid curves are close to the upper boundary. The weak dependence on the threshold is the signature of a large target-present noncentrality: the dedicated

sensing beam supplies enough coherent energy that the statistic under (8.49) lies far from the target-absent distribution for most reasonable thresholds.

The ROC values at  $P_{\text{FA}} = 10^{-2}$  are close to, but not always identical to, the fixed-operating-point  $P_{\text{D}}$  values in Figure 9.2. This is expected because the main figure reports a conditional detection probability over illuminated target voxels, while the ROC curve is obtained by pooling statistic samples and sweeping the threshold over target-present and target-absent populations. When illumination is nearly uniform, as in the hybrid default case, the two estimates agree closely. When the illumination budget varies more strongly across slots and targets, as in the Deterministic sensing case, the two empirical normalizations can differ by a few percentage points.





# B | ADAPTIVE GRID CONVERGENCE

This appendix validates the adaptive voxel-grid controller introduced in Section 6.2. The controller is used throughout Chapter 9, but its behavior is most visible in the target-retention heatmap of Figure 9.9. The purpose here is to show that the grid is not an arbitrary simulation artifact. It follows a reproducible budget law: if the scan has more informative capacity than track-maintenance demand, the grid can refine; if the maintained tracks consume too much of the available budget, the grid coarsens.

The grid controller uses the informative scan budget  $Q_{\text{eff}}[t]$  in (6.3), the track-maintenance demand  $Q_{\text{trk}}[t]$ , and the moving averages in (6.6). The effective coarse-discovery budget is given in (6.7). The controller then applies the three-region rule of (6.9): coarsen if the current number of coarse voxels is above the budget, remain stable if it lies in the admissible band, and refine if the budget is large enough to support a finer grid.

## B.1 CONVERGENCE FROM DIFFERENT INITIALIZATIONS

Figure B.1a runs the same scenario from three initial coarse-grid sizes. The coarse-start initialization uses  $[6, 6, 1]$ , i.e., 36 coarse voxels. The medium-start initialization uses  $[18, 18, 2]$ , i.e., 648 coarse voxels. The fine-start initialization uses  $[30, 30, 2]$ , i.e., 1800 coarse voxels. The controller evaluates the resizing rule once per revisit window, so visible transitions occur every 40 slots. With the baseline slot duration of 14.28 ms, this corresponds to one possible grid update roughly every 0.571 s of simulated time.

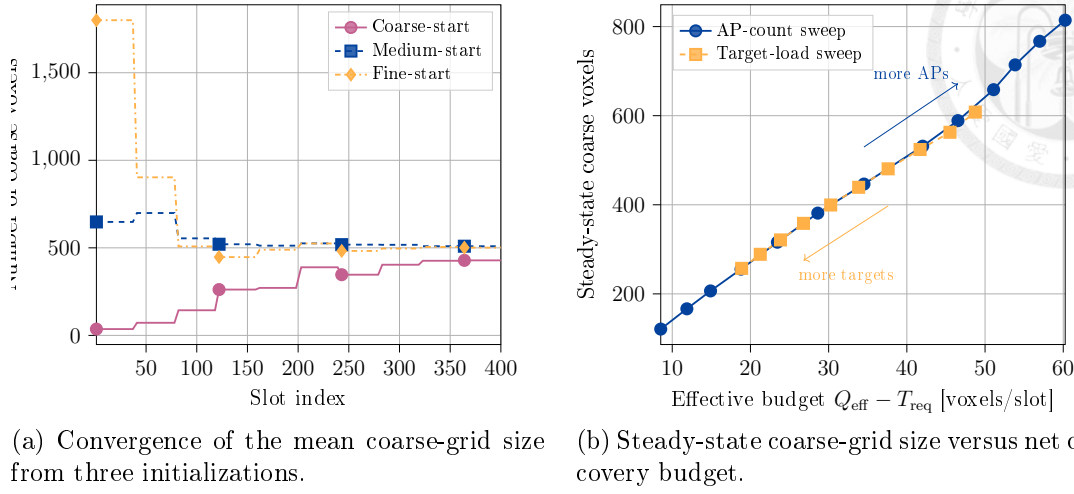


Figure B.1: Adaptive-grid convergence and steady-state budget law. Panel (a) shows the convergence behavior when the same scenario is initialized from coarse, medium, and fine grids. Panel (b) plots the steady-state number of coarse voxels against the net discovery budget  $\bar{Q}_{\text{eff}} - \bar{Q}_{\text{trk}}$  for two physically different sweeps. The near-collapse of the curves shows that the controller responds mainly to the same budget difference, regardless of whether the change comes from more sensing supply or more tracking demand.

The three trajectories move toward the same equilibrium band. The coarse-start case begins far below the admissible region and therefore refines repeatedly, before entering the neighborhood in which different drops may refine or coarsen depending on their realized informative budget. The fine-start case behaves in the opposite way: it begins above the budget and coarsens rapidly. The medium-start case begins near the upper part of the stable region and then descends toward the same band. By the latter part of the 400-slot horizon, all three initializations lie in the same approximate range, around 400–500 coarse voxels. The fine-start case reaches the equilibrium band faster than the coarse-start case because its logarithmic distance from the final band is smaller. The residual oscillation near equilibrium is a consequence of the discrete rule in (6.9).

Each individual drop always has an integer grid size, but the plotted traces are Monte-Carlo averages over multiple drops. Different drops experience different target positions, AP schedules, and informative budgets, so their resize events need

not occur in exactly the same direction at the same update. Averaging these integer trajectories produces non-integer means.

## B.2 STEADY-STATE GRID SIZE VERSUS EFFECTIVE BUDGET

Figure B.1b studies the steady-state law more directly. The horizontal axis is the net discovery budget

$$Q_{\text{net}} \triangleq \bar{Q}_{\text{eff}} - \bar{Q}_{\text{trk}}, \quad (\text{B.1})$$

and the vertical axis is the steady-state mean number of coarse voxels. Two different sweeps are used. In the first, the number of APs is varied, which changes the supply side of the sensing budget by modifying how many requested voxels become informative. In the second, the number of targets is varied, which changes the demand side by increasing the number of track-maintenance requests. The two mechanisms are physically different, but they enter (6.7) through the same difference  $Q_{\text{net}}$ .

The two curves nearly collapse over their shared range. This confirms that the adaptive grid is governed by a single effective budget rather than by the particular physical parameter being swept.

This appendix clarifies the interpretation of the tracking results: the protocol does not use a fixed grid when the scene becomes crowded. As target demand grows, more of the informative budget is spent on maintaining existing tracks, and the grid coarsens to keep the revisit process feasible. As sensing supply grows, the same rule permits a finer grid. The cost of coarsening is reduced spatial resolution and a higher probability of target collisions; the benefit is that the scan does not lose regularity.