Design and Optimization of Resource and Mobility Management in 6G Multi-band Networks

Mohammad Amin Saeidi

A Dissertation Submitted to the Faculty of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

Graduate Program in Electrical Engineering and Computer Science

York University

Toronto, Ontario

June 2026

© Mohammad Amin Saeidi, 2026

Abstract

Sixth-generation (6G) wireless networks require resource and mobility management methods that exploit different frequency bands while accounting for frequency-dependent propagation, blockage, beam behavior, mobility, and hardware constraints. This dissertation studies terahertz (THz) systems, multi-band network (MBN) architectures, cooperative upper mid-band (UMB)/THz MBNs, and reconfigurable intelligent surface (RIS)-assisted systems.

First, a molecular-absorption-aware resource allocation framework is developed for dense downlink THz networks with multi-connectivity. The framework jointly optimizes sub-band partitioning, user association, sub-band assignment, and power allocation while accounting for molecular absorption, blockage, interference, and beam squint. The assignment problem is solved through total unimodularity-based relaxation, and power allocation is addressed using centralized and distributed methods. The results show sum-rate gains over benchmarks and quantify the effects of base-station density, multi-connectivity, molecular absorption, blockage, and hardware impairments.

Second, MBNs are characterized through their main PHY- and MAC-layer challenges. Stand-alone and integrated architectures are compared. The results show that integrated MBNs better support mobility and simultaneous multi-band access,

but rely on more sophisticated and costly multi-band hardware, whereas stand-alone MBNs provide higher deployment flexibility through band-specific base-station densities. Third, a mobility-aware resource allocation framework is developed for cooperative UMB/THz MBNs with extremely large antenna arrays and hybrid beamforming. The framework captures near-field and velocity-aware channels, blockage, cooperative cluster size, and mobility. Joint user association and hybrid beamforming are optimized for stationary users, and two handover-aware methods are developed for mobile users. The results show that cooperative UMB/THz MBNs outperform THz-only networks, near-field beamforming is important for large arrays, and larger cooperative clusters are not always beneficial under mobility.

Finally, RIS-assisted multi-frequency systems are studied from array-level and hardware-level viewpoints. Beam-squint and beam-split effects are characterized under continuous and finite-resolution phase profiles, and a circuit-aware RIS model optimizes practical capacitances across multiple frequencies. The results show accurate beam prediction, sum-rate gains from beam-split-aware codeword selection, and the effects of capacitance range. Overall, this dissertation shows that 6G resource, mobility, and configuration management must account for frequency-dependent propagation, handovers, beam behavior, and hardware constraints. The GitHub code is available at here.

To my beloved parents,

Marjan and Hossein

Acknowledgments

I would like to express my deepest gratitude to my supervisor, Professor Hina Tabassum, for her guidance, support, patience, and encouragement throughout my PhD journey. Her mentorship helped me grow both as a researcher and as a person. I am sincerely grateful for the time, knowledge, and care she invested in my work, and for continuously motivating me to improve the quality of my research.

I would also like to thank my examining committee: Professor Hatem Abou-Zeid, Professor Rashid Bashir, Professor Neda Salahandish, Professor Manar Jammal, and Professor Andrew W. Eckford. I am grateful for their time, thoughtful questions, valuable comments, and constructive feedback.

I am especially grateful to my beloved parents, Marjan and Hossein, for their unconditional love, endless support, and sacrifices. Their belief in me has been a constant source of strength throughout my life. I would also like to thank my brother, Amirhossein, and his wife, Ronak, for their love, encouragement, and support during this journey.

I am thankful to my friends and colleagues who made this experience more meaningful and memorable. In particular, I would like to thank Aryan Manafi and Mehrazin Alizadeh for their friendship, support, and encouragement throughout the ups and downs of my PhD.

Finally, I would like to thank everyone who supported me, directly or indirectly, during this journey. This dissertation would not have been possible without the kindness, encouragement, and support I received from the people around me.

Contents

Abstract ii

Acknowledgments v

Contents vii

List of Tables xiii

List of Figures xiv

List of Acronyms xx

1 Introduction 1

1.1 What will 6G be? . . . 1

1.1.1 Key Performance Indicators . . . 2

1.1.2 Key Technologies . . . 2

1.1.3 Applications . . . 4

1.2 Network Resource Management . . . 4

1.3 Key 6G Technologies Considered in This Dissertation . . . 6

1.3.1 Multi-band 6G Networks . . . 6

1.3.2 THz and Upper Mid-Band Communications . . . 7

1.3.3 Massive Antenna Systems . . . 8

1.3.4 Cooperative Transmission . . . 8

1.3.5 Reconfigurable Intelligent Surfaces . . . 9

1.4 Challenges in Multi-band Resource and Mobility Management . . . 10

1.4.1 Terahertz Resource Allocation . . . 10

1.4.2 Multi-band Network Integration . . . 12

1.4.3 Mobility-aware Cooperative UMB/THz Networking . . . 14

1.4.4 RIS-assisted Multi-frequency Beam and Configuration Management . . . 15

1.5 Scope of the Dissertation . . . 16

1.5.1 Molecular Absorption-Aware Resource Allocation in Dense THz Networks with Multi-Connectivity . . . 16
1.5.2 Characterization of Multi-band Networks . . . 17
1.5.3 Resource Allocation in Cooperative MBNs in the Presence of Mobility . . . 18
1.5.4 Characterization of Beam-Squint and Beam-Split Effects in RIS-assisted Multi-Frequency Networks . . . 19
1.5.5 Circuit-aware Performance Optimization in RIS-Assisted Sub-6 GHz and Mid-Band Multi-Band Networks . . . 20
1.5.6 Conclusion and Future Work . . . 21
1.6 Publications List . . . 21

2 Literature Review and Preliminaries **23**
2.1 THz Communication . . . 23
2.2 Multi-band Networks . . . 26
2.3 Massive Antenna and Cooperative Systems . . . 28
2.4 Mobility Management . . . 29
2.5 RIS-assisted MBNs . . . 31
2.6 Mathematical Preliminaries . . . 34
2.6.1 Alternating Optimization . . . 34
2.6.2 Fractional Programming . . . 35
2.6.3 Scaled ADMM . . . 37
2.6.4 Total Unimodularity . . . 39
2.6.5 Successive Convex Approximation . . . 39
2.6.6 Projected ADAM . . . 40

3 Molecular Absorption-Aware User Assignment, Spectrum, and Power Allocation in Dense THz Networks with Multi-Connectivity **42**
3.1 Overview . . . 42
3.2 System Model and Assumptions . . . 46
3.2.1 Downlink Transmission Model . . . 46
3.2.2 Beam Squint Consideration and THz Channel Model . . . 48
3.2.3 THz SINR Model and Blockage-Aware Rate Model . . . 50
3.3 Problem Formulation, Constraints, and Transformations . . . 52
3.3.1 Problem Formulation . . . 52
3.3.2 Molecular Absorption Coefficient . . . 54
3.3.3 Convexifying the Molecular Absorption Coefficient . . . 55
3.3.4 Leading and Trailing Edge Bands of the TW . . . 56
3.3.5 Bound on the Number of THz Sub-bands with Beam-Squint . . . 57
3.3.6 Transformation of the Objective Function . . . 59

3.4 Proposed Algorithm for WSR Maximization in a Multi-Cell THz Network 60
3.4.1 Joint User Association and Sub-band Assignment 60
3.4.2 Centralized Power Allocation with Fractional Programming and Bisection Search 65
3.5 Low-Complexity Distributed Solution and Complexity Analysis . . . 68
3.5.1 Low-Complexity Distributed Solution 69
3.5.2 Complexity Analysis of the Proposed Algorithms 72
3.6 Numerical Results and Discussions 75
3.6.1 Simulation Set-up and Parameters 75
3.6.2 Impact of Antennas' Directionality 76
3.6.3 Impact of the Number of BSs 78
3.6.4 Effect of Molecular Absorption Coefficient 80
3.6.5 Blockage Density 82
3.6.6 Convergence Analysis 84
3.6.7 Number of Sub-bands, Available Bandwidth and Molecular Absorption Coefficient 86
3.6.8 Multi-Antenna Base Stations 87
3.6.9 Impact of Hardware Imperfections 89
3.6.10 Impact of Imperfect CSI 91
3.7 Complete Proof 92
3.7.1 Proof of Lemma 3.3 92
3.7.2 Proof of Convergence for **Algorithm-3** 94
3.8 Summary 96

4 Characterization of Multi-band Networks **98**
4.1 Overview 98
4.2 Fundamental Challenges in MBNs 101
4.2.1 Multi-band Transceiver Design 101
4.2.2 Multi-band Antenna Systems 102
4.2.3 Channel Estimation and Modeling 104
4.2.4 Network Deployment and Planning 104
4.2.5 User Offloading and Mobility Management 105
4.3 Resource Allocation for MBNs 106
4.3.1 Objective Functions 106
4.3.2 Design Constraints 108
4.4 User Offloading in RF/THz MBN: A Case Study 111
4.5 SA-MBN vs Int-MBN: A Case Study 113
4.5.1 Deployment Density 113
4.5.2 Association to TBSs 114
4.5.3 Spectral Efficiency and Achievable Rate 115

4.5.4 Deployment Cost Efficiency . . . 116
4.5.5 Handoff Probability in MBNs . . . 119
4.5.6 Significance of MBNs over Standalone Counterparts . . . 120
4.6 Summary . . . 121

5 Resource Allocation in Cooperative Mid-band/THz Networks in the Presence of Mobility 123
5.1 Overview . . . 123
5.2 System Model and Assumptions . . . 128
5.2.1 Near Field Communication Model . . . 129
5.2.2 Channel Model . . . 131
5.2.3 Downlink SINR and Rate Model . . . 134
5.3 Joint User Association and Hybrid Beamforming in Cooperative THz-UMB MBN . . . 136
5.3.1 Problem Formulation . . . 137
5.3.2 Analog Beamforming . . . 139
5.3.3 Problem Transformation . . . 141
5.3.4 Joint User Association and Beamforming . . . 143
5.3.5 Convergence and Complexity Analysis of Algorithm 1 . . . 147
5.4 Mobility-aware User Association and Hybrid Beamforming in Cooperative THz/UMB MBN . . . 150
5.4.1 HO Cost-based Approach . . . 151
5.4.2 MOOP-based Approach . . . 155
5.5 Numerical Results and Discussions . . . 156
5.5.1 Simulation Set-up and Parameters . . . 156
5.5.2 Benchmarks . . . 157
5.5.3 Convergence of Algorithm 1 . . . 158
5.5.4 Numerical Results for Stationary Users . . . 159
5.5.5 Numerical Results for Mobile Users . . . 165
5.6 Complete Proof . . . 173
5.6.1 Proof of Lemma 5.1 . . . 173
5.6.2 Proof of Lemma 5.2 . . . 174
5.7 Summary . . . 175

6 Characterization of Beam-Squint and Beam-Split Effects in RIS-assisted Multi-Frequency Networks 177
6.1 Overview . . . 177
6.2 System Model and Problem Statement . . . 181
6.3 Beam-squint and Beam-split Existence and Characterization: Continuous RIS Phase Profiles . . . 183

6.3.1 Beam-squint and Beam-split Characterization 183
6.3.2 Beam-Squint/Split Existence Conditions (Elevation Angles) . 185
6.3.3 Beam-Squint/Split Existence Conditions (Azimuth Angles) . . 187
6.4 Beam-squint and Beam-split Existence and Characterization: b-bit RIS Phase Profiles 189
6.4.1 Beam Pattern with b-bit Quantized RIS 190
6.4.2 Beam Pattern Transformation and Per-harmonic Peaks 192
6.4.3 Existence of Per Harmonic Beam-Squint/Split 193
6.5 Per-harmonic beam-squint/Split Angle Correction Mechanism 195
6.5.1 Decomposition of Harmonic of Interest and Perturbation . . . 196
6.5.2 Local Surrogate Function-based Optimization 198
6.5.3 Computational Complexity Analysis 199
6.6 Use-Case Study: Beam-Split Aware Scheduling in Multi-RIS, Multi-User, Multi-band Network 203
6.6.1 Quantized codebooks at $\rho = 1$ (home and cross incidence) . . 204
6.6.2 Beam-split-aware Codeword Selection 205
6.7 Numerical Results and Discussions 207
6.7.1 Beam-Squint/Split Analytical Validation 208
6.7.2 Beam-split Existence (Elevation and Azimuth) 209
6.7.3 Number of RIS Elements in b-bit Quantization 210
6.7.4 Beam-split-aware Scheduling 212
6.8 Complete Proof and Derivations 215
6.8.1 Proof of Lemma 6.1 215
6.8.2 Proof of Lemma 6.2 215
6.8.3 Proof of Lemma 6.3 217
6.8.4 Gradient and Hessian of J used in (6.40) 218
6.9 Summary 220

7 Circuit-aware Performance Optimization in RIS-Assisted Sub-6 GHz and Mid-band Multi-Band Networks 222
7.1 Overview 222
7.2 RIS Model Characterization 224
7.2.1 Array Response and Circuit-Based Reflection Model 225
7.2.2 RIS Supported Phase-Shift Range 226
7.3 Problem Formulation and Proposed Solution 228
7.3.1 Sum Array-Gain Maximization 229
7.3.2 Complexity Analysis 233
7.4 Numerical Results and Discussions 233
7.4.1 Convergence and Solver Comparison 234
7.4.2 Impact of Feasible Capacitance Window 235

7.4.3 Impact of Element Spacing 236
7.5 Summary 238

8 Conclusion and Future Work **242**
8.1 Conclusion 242
8.2 Future Work 247

Bibliography **251**

List of Tables

3.1 Curve-fitting parameters based on the approximated function of molecular absorption coefficient, $\bar{k}(f)$. 57

3.2 Comparison of time complexity between the proposed LP and binary solution of $\mathcal{P}_3$ 65

List of Figures

1.1 Transmission window based on molecular absorption coefficient 12

3.1 System model illustrating BSs and users in the presence of randomly distributed blockers. 47

3.2 Graphical illustration of $\mathbf{T}^{(2)}$ and $\mathbf{T}^{(3)}$. 64

3.3 Varying antennas' directionality on network sum-rate. $\Gamma_n^L = 1$, $B = 6$, $N = 12$. 76

3.4 Varying antennas' directionality on average multi-connectivity order. $\Gamma_n^L = 1$, $B = 6$, $N = 12$. 77

3.5 System sum-rate vs number of BSs, $\Gamma_n^L = 1$, $q = 0.2$, $N = 12$. 78

3.6 Average multi-connectivity order vs number of BSs, $\Gamma_n^L = 1$, $q = 0.2$, $N = 12$. 79

3.7 Network sum-rate vs molecular absorption coefficient, $\Gamma_n^L = 1$, $B = 6$, $N = 12$. 80

3.8 Average multi-connectivity order vs molecular absorption coefficient, $\Gamma_n^L = 1$, $B = 6$, $N = 12$. 81

3.9 System sum-rate vs blockage density, $\Gamma_n^L = 1$, $q = 0.2$, $N = 12$. 82

3.10 Average multi-connectivity order vs blockage density, $\Gamma_n^L = 1$, $q = 0.2$, $N = 12$. 83

3.11 Convergence of the proposed algorithms for power allocation, and the impact of varying the penalty factor in Algorithm-2. 84
3.12 Convergence of the proposed algorithm, Algorithm-3, for different setups, minimum required number of sub-bands for each user, i.e., Γ_n^L, and different initializations. 85
3.13 Total available bandwidth and optimal number of sub-bands for different TWs versus molecular absorption coefficient variation tolerance, ϵ. 87
3.14 Molecular absorption coefficient $k(f)$ for different TWs versus molecular absorption coefficient variation tolerance, ϵ. 88
3.15 System sum-rate vs the number of antennas at BSs, $q = 0.5$, $B = 10$, $N = 12$ 89
3.16 System sum-rate vs the level of HI at BSs and users ($k_t = k_r$), $q = 0.5$, $B = 10$, $N = 12$ 90
3.17 System sum-rate vs the accuracy of channel estimation (ζ), $q = 0.5$, $B = 10$, $N = 12$ 91

4.1 Vision of (a) Int-MBN architecture and (b) SA-MBN architecture in next generation wireless networks. 100
4.2 Network data rate versus molecular absorption coefficient for different number of BSs and UE-BS association metrics 111
4.3 The required number of BSs versus the target rate threshold. $R_{\max} = 400[m]$, $P_R^{tx} = 3$[Watts] 114
4.4 Impact of increasing the number of BSs at various THz bandwidth on THz association probability 115

4.5 Impact of increasing the number of BSs at various THz bandwidth on Spectral efficiency and data rate 117
4.6 Impact of increasing the number of BSs at various THz bandwidth on deployment cost efficiency 118
4.7 HO probability versus the velocity of user for different network architecture 119
4.8 Outage probability versus the molecular absorption coefficient for average network rate threshold of 1 GHz, $N_T = N_R = N_{\text{Hyb}} = 15$ 120
4.9 Average network data rate versus the molecular absorption coefficient 121

5.1 System model illustrating BSs and users and NFC model. 129
5.2 Convergence of Algorithm 1 for different system parameters 158
5.3 Network sum-rate at different molecular absorption coefficients, $R^{\text{Th}} = 0.5$ [Gbps], $K = 12$, $B^{\text{T}} = 4$, $B^{\text{U}} = 2$, $B^{\text{T}}_{\text{THzOn}} = 6$, $M^{\text{T}} = 504$, $M^{\text{U}} = 84$ 159
5.4 System sum-rate versus density of blockers, ξ. $R^{\text{Th}} = 0.5$ [Gbps], $K = 12$, $B^{\text{T}} = 4$, $B^{\text{U}} = 2$, $B^{\text{T}}_{\text{THzOn}} = 6$, $M^{\text{T}} = 504$, $M^{\text{U}} = 84$ 161
5.5 System sum-rate versus number of TBSs and UBSs. The total number of BSs is 10 and increasing TBSs decreases UBSs as B^{U} varies within the range of $[10, 0]$. $R^{\text{Th}} = 0.5$ [Gbps], $K = 12$, $M^{\text{T}} = 504$, $M^{\text{U}} = 84$. 162
5.6 System sum-rate versus the different number of antennas of UMB and THz bands for various velocities of users and antenna spacing. $R^{\text{Th}} = 0.1$ [Gbps], $K = 8$, $B^{\text{T}} = 4$, $B^{\text{U}} = 2$, $\mathcal{C}^{\text{T}}_k = 4$, $\mathcal{C}^{\text{U}}_k = 2$ 163
5.7 Normalized Array Gain at the THz band for different user velocity and antenna spacing, considering the distance from the transmitter is 40 [m], $\cos(\theta) = 0.5$ and $M^{\text{T}} = 504$ 165

5.8 Normalized Array Gain at the UMB for different user velocity and antenna spacing, considering the distance from the transmitter is 40 [m], $\cos(\theta) = 0.5$, and and $M^{\mathrm{U}} = 84$ 166

5.9 HO-aware sum-rate and the total number of HOs versus different cluster size ($\mathcal{C}_k^{\mathrm{T}} = \mathcal{C}_k^{\mathrm{U}}$), where $R^{\mathrm{Th}} = 0.5$ [Gbps], $K = 12$, $B^{\mathrm{T}} = 5$, $B^{\mathrm{U}} = 3$, $B^{\mathrm{T}}_{\mathrm{THzOn}} = 8$, and $\eta^{\mathrm{T}} = \eta^{\mathrm{U}} = 0.4$. 168

5.10 HO-aware sum-rate versus different HO cost ($\eta^{\mathrm{T}} = \eta^{\mathrm{U}}$), where $R^{\mathrm{Th}} = 0.5$ [Gbps], $K = 15$, $B^{\mathrm{T}} = 5$, $B^{\mathrm{U}} = 3$, $B^{\mathrm{T}}_{\mathrm{THzOn}} = 8$, and $\Xi = 1$. 169

5.11 Total number of HOs versus different HO cost ($\eta^{\mathrm{T}} = \eta^{\mathrm{U}}$), where $R^{\mathrm{Th}} = 0.5$ [Gbps], $K = 15$, $B^{\mathrm{T}} = 5$, $B^{\mathrm{U}} = 3$, $B^{\mathrm{T}}_{\mathrm{THzOn}} = 8$, and $\Xi = 1$. 170

5.12 HO-aware sum-rate versus minimum rate requirement R^{Th}, where $K = 15$, $B^{\mathrm{T}} = 5$, $B^{\mathrm{U}} = 3$, $B^{\mathrm{T}}_{\mathrm{THzOn}} = 8$, $\Xi = 1$, and $\eta^{\mathrm{T}} = \eta^{\mathrm{U}} = 0.4$ 171

5.13 HO-aware system sum-rate versus CSI uncertainty χ, where $K = 12$, $B^{\mathrm{T}} = 4$, $B^{\mathrm{U}} = 2$, $B^{\mathrm{T}}_{\mathrm{THzOn}} = 6$, $\Xi = 1$, and $\eta^{\mathrm{T}} = \eta^{\mathrm{U}} = 0.4$ 172

6.1 Two-RIS two-UE multi-band network: RIS 1 (RIS 2) is configured at f_1 (f_2); colored beams illustrate beam-split under cross-band evaluation, with angles (φ, θ) defined by the shown coordinate axes. 178

6.2 Continuous RIS beampattern under frequency mismatch for $\rho = 0.5$, $\alpha_y = \alpha_z = 1.5$, $N_y = 12$, $N_z = 12$, with incidence $(\varphi_I, \theta_I) = (-30°, -10°)$ and design direction $(\varphi_D, \theta_D) = (-24°, 44°)$. 189

6.3 Quantized RIS beampattern under frequency mismatch for $\rho = 0.65$, $\alpha_y = \alpha_z = 0.5$, $N_y = 4$, $N_z = 6$, and $b = 1$, with incidence $(\varphi_I, \theta_I) = (-30°, -10°)$ and design direction $(\varphi_D, \theta_D) = (-24°, 44°)$. 200

6.4 Quantized RIS beampattern under frequency mismatch for $\rho = 0.65$, $\alpha_y = \alpha_z = 0.5$, $N_y = 12$, $N_z = 16$, and $b = 1$, with incidence $(\varphi_I, \theta_I) = (-30°, -10°)$ and design direction $(\varphi_D, \theta_D) = (-24°, 44°)$. 201

6.5 Quantized RIS beampattern under frequency mismatch for $\rho = 0.5$, $\alpha_y = \alpha_z = 1.5$, $N_y = 12$, $N_z = 12$, and $b = 1$, with incidence $(\varphi_I, \theta_I) = (-30°, -10°)$ and design direction $(\varphi_D, \theta_D) = (-24°, 44°)$. 202

6.6 Mean error between the predicted peaks and the actual peaks for both continuous and quantization cases as a function of frequency ratio ρ for different element spacing, and $N_z = N_y = 10$. 210

6.7 Elevation split probability $\mathbb{P}(|\mathcal{M}_z| \geq 2)$ versus (ρ, α_z), where $\mathcal{M}_z$ is the feasible-integer set in (6.11); dashed lines mark $\frac{2\alpha_z}{\rho} = 1$ and $\frac{2\alpha_z}{\rho} = 2$. . 211

6.8 Azimuth split probability $\mathbb{P}(|\mathcal{M}_y| \geq 2 \mid \text{valid elevation})$ versus (ρ, α_y), where $\mathcal{M}_y$ is the feasible-integer set in (6.15); contours summarize thresholds of $\frac{2\alpha_y}{\rho} A_O$ under the same conditioning. 212

6.9 Mean error between the predicted peaks and the actual peaks for per-harmonic and corrected peaks versus the number of RIS elements, for different frequency ratio ρ and $\alpha_y = \alpha_z = 0.5$. 213

6.10 Average two-user sum-rate versus RIS size ($N_y = N_z$) in the two-RIS two-UE multi-band network, comparing beam-split-aware and beam-split-agnostic codeword selection for 1-bit quantization, i.e., $\mathcal{K}_\eta = \{-1, 0\}$, $\alpha \in \{0.75, 1\}$ and $Q_{\min} \in \{8, 16\}$. 214

7.1 Convergence behavior of the best run for each method in native-band RIS optimization: (a) ADAM-based methods and (b) GA. 235

7.2 Impact of the feasible capacitance window on the normalized sum array gain for the sub-6 RIS serving the sub-6 frequencies. 237
7.3 Impact of the feasible capacitance window on the normalized sum array gain for the sub-6 RIS serving the mid-band frequencies. 238
7.4 Impact of the feasible capacitance window on the normalized sum array gain for the mid-band RIS serving the mid-band frequencies. 239
7.5 Impact of the feasible capacitance window on the normalized sum array gain for the mid-band RIS serving the sub-6 frequencies. 240
7.6 Impact of the element spacing factor α_r on the normalized sum array gain for the sub-6 and mid-band RISs. 241

List of Acronyms

Acronym	Meaning
5G	Fifth-generation
6G	Sixth-generation
ADAM	Adaptive moment estimation
ADMM	Alternating direction method of multipliers
AI	Artificial intelligence
AO	Alternating optimization
AP	Access point
BS	Base station
CAPEX	Capital expenditure
CAV	Connected autonomous vehicle
CPU	Central processing unit
CSI	Channel state information
DCE	Deployment cost efficiency
EE	Energy efficiency
EHF	Extremely high frequency
ELAA	Extremely large antenna array
EM	Electromagnetic

Acronym	Meaning
FFC	Far-field communication
FP	Fractional programming
FSO	Free-space optics
Gbps	Gigabits per second
GHz	Gigahertz
HBF	Hybrid beamforming
HITRAN	High-resolution transmission molecular absorption database
HO	Handover
IMT	International Mobile Telecommunications
ISAC	Integrated sensing and communication
ITU	International Telecommunication Union
KPI	Key performance indicator
LiFi	Light fidelity
LoS	Line-of-sight
LP	Linear programming
MAC	Medium access control
MBN	Multi-band network
MEMS	Microelectromechanical system
MIMO	Multiple-input multiple-output
MINLP	Mixed-integer nonlinear programming
ML	Machine learning
MM	Majorization-minimization
mmWave	Millimeter wave
MOOP	Multi-objective optimization problem

Acronym	Meaning
NEMS	Nanoelectromechanical system
NFC	Near-field communication
NLoS	Non-line-of-sight
NOMA	Non-orthogonal multiple access
NP-hard	Nondeterministic-polynomial-time hard
NR	New Radio
OFDM	Orthogonal frequency-division multiplexing
OPEX	Operational expenditure
PHY	Physical layer
QCQP	Quadratically constrained quadratic program
QoS	Quality-of-service
RBS	RF base station
RF	Radio frequency
RIS	Reconfigurable intelligent surface
RSMA	Rate-splitting multiple access
RSRP	Reference signal received power
RZF	Regularized zero-forcing
SA	Stand-alone
SCA	Successive convex approximation
SDMA	Spatial division multiple access
SE	Spectral efficiency
SINR	Signal-to-interference-plus-noise ratio
SISO	Single-input single-output
TBS	THz base station

Acronym	Meaning
Tbps	Terabits per second
THz	Terahertz
TUM	Totally unimodular matrix
UAV	Unmanned aerial vehicle
UBS	Upper mid-band base station
UE	User equipment
ULA	Uniform linear array
UMB	Upper mid-band
UPA	Uniform planar array
URLLC	Ultra-reliable and low-latency communication
VLC	Visible light communication
WiFi	Wireless fidelity
XR	Extended reality
ZF	Zero-forcing

Chapter 1

Introduction

1.1 What will 6G be?

The proliferation of smart products, intelligent applications, and interactive services has grown substantially in recent years, a trend that is expected to continue. Recognizing this evolving landscape and rising demands, the limitations of current 5G infrastructure have prompted both academia and industry to focus on developing sixth-generation networks (6G). According to ITU-R Report M.2516-0 (2023),"IMT-2030" is the framework for the future development of International Mobile Telecommunications for 2030 and beyond [1]. It is intended to provide immersive, ubiquitous, and sustainable connectivity by extending and complementing IMT-2020 capabilities. IMT-2030 aims to deliver significantly improved performance in data rates, latency, reliability, density, and energy efficiency, while supporting new usage scenarios such as immersive communication, massive twinning, and AI-native communication.

These goals are commonly described through three connected aspects: the KPIs that define the target performance, the technologies that can enable these targets, and the applications that motivate the evolution toward 6G.

1.1.1 Key Performance Indicators

Across various KPIs, 6G is anticipated to deliver benchmarks far exceeding those of 5G. It is envisioned to achieve peak data rates of up to 1 Tbps, enabling data-intensive applications like holographic communications and real-time digital twins. The user-experienced data rate is expected to reach 1 Gbps, ensuring consistently high performance even in crowded urban environments or during high-demand periods. To support applications requiring extremely low latency, such as autonomous vehicles and remote surgery, 6G targets end-to-end latency as low as 1 ms, coupled with reliability levels as high as 99.999999 percent.

Energy efficiency is another critical goal, with 6G systems aiming for 1 Tb/J, reflecting the need to balance high performance with environmental sustainability. Furthermore, 6G is designed to support seamless connectivity for devices traveling at speeds up to 1000 km/h, addressing requirements for high-speed trains and aircraft. The network is also planned to accommodate up to 10^7 devices per square kilometer, enabling dense IoT ecosystems and smart cities. These KPIs highlight the transformative potential of 6G, but they also show that future networks must jointly support high data rates, reliable connectivity, low latency, dense access, and mobility under diverse propagation conditions.

1.1.2 Key Technologies

To meet these ambitious KPIs, several techniques and technologies have been proposed and explored. Embracing Extremely High Frequency (EHF) bands, including Terahertz (THz), millimeter-wave (mmWave), and optical communications like

Visible Light Communication (VLC), opens up larger bandwidths to support increased peak data rates. Integrating different frequency bands in multi-band networks (MBNs)[1] can improve overall network capacity, enhance mobility support, and enable cost-effective deployment. Ultra-massive Multiple Input Multiple Output (MIMO) systems and cooperative systems leverage spatial diversity to provide extensive connectivity and higher data rates. The incorporation of Artificial Intelligence (AI) and Machine Learning (ML) will enable intelligent, fast decision-making, predictive analytics, and adaptive resource allocation in 6G networks.

Advanced multiple access methods such as Spatial Division Multiple Access (SDMA), Non-Orthogonal Multiple Access (NOMA), Rate Splitting Multiple Access (RSMA), and hybrid RSMA-NOMA can facilitate massive connectivity with limited resources. Reconfigurable Intelligent Surfaces (RISs), with their ability to create smart communication environments through passive reflections and refractions, will also play a key role in meeting these KPIs. Additionally, integrating satellite communications with terrestrial networks will contribute to achieving 6G performance targets. Together, these technologies show that 6G will rely on the joint use of multiple spectrum bands, large antenna systems, intelligent propagation control, and adaptive resource management. The technologies that are directly considered in this dissertation are introduced in the next section.

[1]In this dissertation, the term *multi-band* is used when the network operates over distinct spectrum bands, such as RF/THz, UMB/THz, or sub-6 GHz/UMB bands. The term *multi-frequency* is used in a broader sense to refer to operation over multiple carrier frequencies, either within the same band or across different bands. Thus, multi-band operation is a special case of multi-frequency operation. In this dissertation, *multi-band* is mainly used for network architecture and resource allocation, while *multi-frequency* is mainly used when discussing frequency-dependent beamforming, RIS configuration, and hardware response.

1.1.3 Applications

By achieving the stated KPIs through these technologies, 6G is expected to unlock a wide range of advanced applications. These include holographic communication, enabling real-time observation of 3D objects for virtual meetings, immersive education, and entertainment. Additionally, 6G will enable the tactile internet, combining ultra-low latency with high availability, reliability, and security to support applications such as remote surgery and haptic systems. Advancements in sensing and localization will be critical for applications including autonomous vehicles, smart cities, eHealth, and industrial automation. Moreover, 6G aims to provide global connectivity, bridging the digital divide by delivering reliable communication to remote and underserved regions worldwide.

Supporting these applications requires more than increasing the available bandwidth. The network must also manage how users, spectrum, power, beams, and infrastructure are selected and configured under frequency-dependent propagation, blockage, mobility, and hardware constraints.

1.2 Network Resource Management

Enhanced network planning and resource management solutions continue to play a pivotal role in the evolution of wireless network design. The effectiveness of these solutions is essential for the systematic allocation of network resources, meeting quality-of-service (QoS) requirements, and reducing both capital and operational costs. In broad terms, resource management refers to the allocation and configuration of available communication resources in order to improve performance metrics such as energy efficiency, spectral efficiency, data rate, reliability, and other KPIs.

In conventional wireless networks, resource management mainly involves power allocation, spectrum allocation, user association, and beamforming. In next-generation networks, however, these tasks become more tightly coupled due to the use of wider bandwidths, higher carrier frequencies, dense deployments, mobility, and multiple coexisting frequency bands. Therefore, the design of resource management solutions must account not only for the amount of available resources, but also for the propagation characteristics, hardware constraints, and mobility behavior associated with different parts of the network. The fundamental network resource allocation methods include:

- *Power Allocation:* The transmit power of a signal holds paramount importance, representing the quantity of power, measured in Watts, assigned by a transmitter to the communication link. Additionally, the overall transmitting power from a Base Station (BS) is constrained by a predefined value known as the power budget. This power budget can vary across different frequency bands.

- *Spectrum Allocation:* Bandwidth allocation in wireless communication networks pertains to the distribution and assignment of available frequency ranges (bandwidth) to different users or services. This process is crucial for optimizing data transfer rates, minimizing interference, and ensuring efficient utilization of the radio frequency spectrum.

- *User Association:* User association in wireless communication networks involves the assignment of individual users or devices to specific access points or base stations within the network. This allocation is typically based on factors such as signal strength, data traffic load, and overall network performance.

- *Beamforming:* Beamforming in wireless communication networks is a technique used to focus radio frequency signals in specific directions, improving the efficiency and reliability of data transmission. By adjusting the phase and amplitude of signals from multiple antennas, beamforming concentrates the signal energy towards a targeted receiver or device, enhancing signal strength and quality.

1.3 Key 6G Technologies Considered in This Dissertation

In this section, we introduce the main technologies considered in this dissertation: multi-band networks, THz and upper mid-band communications, massive antenna systems, cooperative transmission, and reconfigurable intelligent surfaces. These technologies are relevant to 6G because they can increase the available bandwidth, improve coverage, support mobility, and provide additional control over the wireless propagation environment.

1.3.1 Multi-band 6G Networks

Over the years, wireless networks have evolved from homogeneous cellular networks to heterogeneous networks with multiple access technologies and different cell types. However, most existing wireless systems still rely mainly on RF, sub-6 GHz, and mmWave bands. These bands alone cannot satisfy the massive connectivity, high data rate, low latency, and broad coverage requirements expected in next-generation wireless networks [2].

Extremely high frequency bands, such as THz and optical bands, can provide much wider transmission bandwidths and very high data rates. However, these bands

also suffer from severe propagation limitations and smaller coverage regions. Lower frequency bands provide more reliable coverage and better mobility support, but they offer less bandwidth. Therefore, there is no single frequency band that can satisfy all 6G requirements.

Multi-band networks (MBNs) address this limitation by allowing different frequency bands to coexist and complement each other. Unlike conventional heterogeneous networks, MBNs involve highly diverse frequency bands with different propagation characteristics, hardware constraints, coverage regions, and bandwidth availability. This makes MBNs a central framework for the resource allocation and mobility management problems considered in this dissertation.

1.3.2 THz and Upper Mid-Band Communications

THz communication is one of the main candidate technologies for supporting extremely high data rates in 6G. The THz and sub-THz spectrum can provide very large bandwidths compared with conventional RF and mmWave bands. This makes THz communication attractive for bandwidth-hungry applications such as extended reality, wireless backhauling, digital twins, and high-speed short-range access. However, THz signals also experience severe propagation losses, blockage, molecular absorption, and sensitivity to beam alignment.

The upper mid-band (UMB), typically referring to frequencies above the conventional sub-6 GHz range and below mmWave, has recently received attention as a promising frequency range for 6G. In particular, the UMB can provide more spectrum than sub-6 GHz while maintaining better coverage and propagation conditions than mmWave and THz bands [3, 4]. Therefore, UMB can serve as a useful complement to

THz in MBNs. In this dissertation, the UMB and THz bands are considered together to study how a lower and more stable band can support mobility and coverage, while the THz band provides much larger bandwidth.

1.3.3 Massive Antenna Systems

Massive antenna systems are expected to play an important role in 6G networks, especially at high frequencies. In the UMB and THz ranges, large antenna arrays can provide high beamforming gains to compensate for propagation losses. The smaller wavelengths at these frequencies also make it possible to deploy a large number of antenna elements in a compact physical space.

However, massive antenna systems require careful design. Fully digital beamforming may be impractical for very large arrays because each antenna element requires a separate RF chain, which increases cost and power consumption. Hybrid beamforming and antenna selection are therefore commonly used to reduce the number of RF chains. Moreover, extremely large antenna arrays (ELAAs) can extend the near-field region, where the conventional far-field channel approximation may no longer be accurate. In this case, the channel depends not only on the angle of the user but also on the distance between the user and the antenna array.

1.3.4 Cooperative Transmission

Cooperative transmission is another important technique for improving coverage, reliability, and data rate in dense wireless networks. In coordinated multi-point transmission, multiple BSs cooperate by sharing data and channel state information to jointly serve users. This cooperation can improve the received signal power, mitigate

interference, and provide spatial diversity, especially for users near cell edges or in blockage-prone environments.

In high-frequency networks, cooperation is particularly useful because links are more sensitive to blockage and coverage limitations. By allowing a user to be served by multiple BSs, cooperative transmission can reduce the risk of link interruption. However, cooperation also increases coordination overhead and can affect mobility management, since a mobile user may experience handovers with respect to multiple serving BSs.

1.3.5 Reconfigurable Intelligent Surfaces

Reconfigurable intelligent surfaces (RISs) are engineered structures composed of thin, semi-passive elements that can control incident electromagnetic waves through adjustable phase shifts and, in some designs, amplitude variations. By properly configuring these elements, RISs can reshape the wireless propagation environment to improve coverage, enhance received signal strength, suppress interference, and improve communication performance.

Due to their low power consumption and ease of deployment on existing infrastructure such as walls, ceilings, facades, and poles, RISs have emerged as a promising technology for dense urban networks, transportation hubs, industrial environments, and other coverage-limited scenarios. Compared with active relays, RISs introduce little additional noise and can provide a more energy-efficient and cost-effective way to enhance wireless connectivity.

In multi-frequency systems, RISs are especially relevant because the same surface

may need to support signals at different carrier frequencies. This makes RIS configuration a frequency-dependent design problem, as discussed in the next section.

1.4 Challenges in Multi-band Resource and Mobility Management

The technologies introduced above create new resource and mobility management challenges in 6G networks. In conventional systems, resource management mainly involves power allocation, spectrum allocation, user association, and beamforming within a limited set of frequency bands. In multi-band and multi-frequency networks, these decisions become more tightly coupled with mobility, blockage, beam alignment, handover, cooperative transmission, and hardware constraints.

In this section, we discuss the main challenges that motivate this dissertation. We first focus on THz resource allocation, where molecular absorption, blockage, and beam squint affect spectrum and power allocation. We then discuss multi-band network integration, where different bands must be coordinated to balance rate, coverage, cost, and mobility support. Next, we explain mobility-aware cooperative UMB/THz networking. Finally, we discuss RIS-assisted multi-frequency operation.

1.4.1 Terahertz Resource Allocation

The THz and sub-THz spectrums, occupying the frequency range from 0.1 THz to 10 THz (wavelength 0.1 mm to 30 μm) and 0.1 to 0.3 THz, respectively, lie between microwave and infrared (IR) regions in the electromagnetic (EM) spectrum. Although THz communication can provide very large bandwidth, its propagation characteristics make resource allocation fundamentally different from lower-frequency systems. THz propagation suffers from severe free-space path loss, which reduces the received signal

power as the transmitter-receiver distance increases. As a result, highly directional antennas and dense BS deployments are often needed to compensate for this loss.

Another characteristic of THz band channels is molecular absorption, which is negligible in lower frequencies used in traditional cellular networks. Even in clear air in terrestrial communications, oxygen and water molecules cause absorption loss. Different environmental factors, such as weather conditions, pressure, temperature, humidity, pollution, and altitude, the abundance of certain gas molecules, all affect THz transmission. More specifically, molecular absorption occurs when traveling EM waves interact with gas molecules, leading to vibrations and rotations in the polar molecules. This process exhibits large absorption peaks at various frequencies [5]. The frequency ranges between these peaks where the molecular absorption coefficient is minimized are selected as transmission windows (TW) as can be seen in figure 1.1. In addition to reducing desired received power, molecular absorption in THz signals generates molecular absorption noise, dependent on transmitting power. This noise, combined with significant thermal noise resulting from the wide THz bandwidth, poses challenges in THz communication due to its cumulative effect on signal quality.

THz signals are also highly sensitive to blockage because their wavelength is very small compared with common objects in the environment. Buildings, vehicles, walls, and the human body can strongly attenuate THz links, making THz communication highly dependent on the availability of line-of-sight paths [6]. Moreover, due to the ultra-wide bandwidth of THz transmission, different frequency components of a signal may point toward different physical directions when a beamformer is designed at a fixed carrier frequency. This beam squint effect reduces the effective array gain and must be considered when determining the number and bandwidth of THz sub-bands.

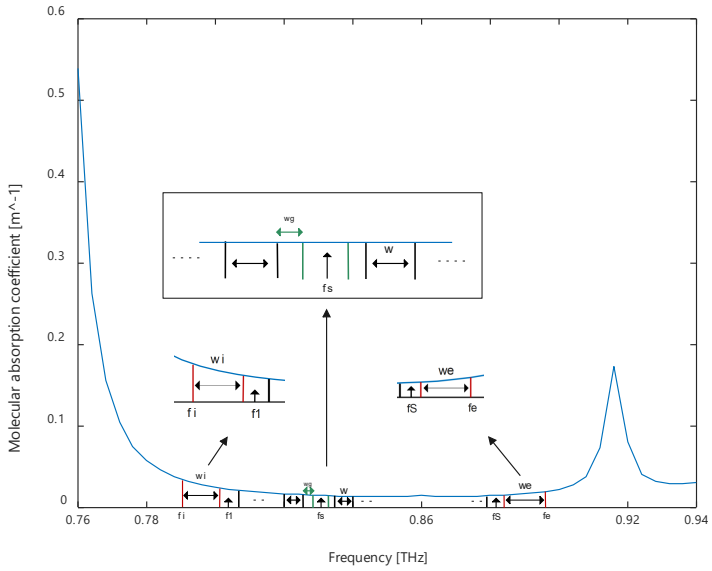


Figure 1.1: Transmission window based on molecular absorption coefficient

Therefore, THz resource allocation cannot be performed by simply dividing the available bandwidth into arbitrary sub-bands. The molecular absorption profile of the selected transmission window should be considered, the edge regions of the transmission window may need to be avoided, and the sub-band bandwidth should be limited to reduce beam-squint-induced gain loss. Moreover, user association, sub-band assignment, and power allocation should account for blockage, molecular absorption noise, interference, dense BS deployment, and multi-connectivity. These challenges motivate the low-complexity THz resource allocation framework studied in this dissertation.

1.4.2 Multi-band Network Integration

A single frequency band cannot simultaneously provide very large bandwidth, wide coverage, reliable mobility support, and low deployment cost. THz links provide

large bandwidth, but they are sensitive to propagation loss, blockage, and mobility. Lower frequency bands provide wider coverage and more stable connectivity, but they cannot provide enough bandwidth for many data-intensive services. Therefore, future networks are expected to integrate multiple frequency bands and use their complementary characteristics.

This integration introduces new resource management challenges. Each band has a different propagation behavior, coverage range, blockage sensitivity, available bandwidth, and hardware requirement. Therefore, user association and traffic offloading cannot be based only on received signal strength. They should also consider service requirements, user mobility, molecular absorption, beam alignment, and the reliability of each band. In addition, the network must decide how different bands should be deployed and coordinated, since stand-alone and integrated MBN architectures lead to different trade-offs in data rate, coverage, handover probability, deployment cost, and backhauling.

As a result, resource allocation in MBNs is more complex than in single-band systems. Power allocation, bandwidth allocation, user association, beamforming, and backhauling must be designed while accounting for the different characteristics of each frequency band. Mobility further complicates this design because the best band for a user may change with location, blockage, velocity, and handover cost. Thus, MBNs require resource management methods that can exploit high-frequency bandwidth when available, while relying on more stable lower-frequency links when mobility or blockage makes high-frequency transmission unreliable. These challenges motivate the study of MBN architectures and resource allocation considerations in this dissertation.

1.4.3 Mobility-aware Cooperative UMB/THz Networking

Mobility is a major challenge in high-frequency networks. In THz systems, narrow beams are needed to compensate for large path loss. However, narrow beams are sensitive to user movement, blockage, and beam misalignment. A mobile user may quickly leave the coverage region of a narrow beam or a small THz cell, leading to intermittent connectivity and frequent handovers.

In UMB/THz MBNs, mobility management is coupled with user association, band selection, cooperative transmission, and beamforming. THz links can provide high data rates but are vulnerable to blockage and mobility, while UMB links can provide better coverage and more stable connectivity. A cooperative UMB/THz MBN can therefore exploit the strengths of both bands: the THz band provides large bandwidth, whereas the UMB improves coverage and mobility support.

Cooperative transmission can further improve reliability by allowing a user to be served by multiple BSs. This can increase the received power, provide spatial diversity, and reduce the impact of blockage. However, cooperation also introduces a new trade-off. A larger cooperative cluster may improve the instantaneous rate, but it can also increase the number of handovers when users move. Therefore, user association, cluster size, and beamforming must be designed jointly.

Another challenge comes from the use of extremely large antenna arrays at high frequencies. These arrays extend the near-field region, making conventional far-field channel models less accurate. As a result, mobility-aware cooperative UMB/THz MBNs require accurate near-field and velocity-aware channel modeling, together with hybrid beamforming and handover-aware resource allocation.

1.4.4 RIS-assisted Multi-frequency Beam and Configuration Management

In RIS-assisted systems, resource management includes not only the allocation of spectrum, power, and BSs, but also the configuration of the RIS reflection response. In single-frequency operation, an RIS phase profile can be designed to reflect the incident signal toward a desired direction. In multi-frequency operation, however, a phase profile designed at one carrier frequency may not generate the same reflected beam at another carrier frequency.

This frequency dependence appears at two different levels. The first one is the array-level effect caused by the fixed physical spacing of the RIS elements. Since the electrical spacing changes with frequency, the dominant reflected beam may deviate from the intended direction, known as beam squint, or several dominant lobes may appear, known as beam split. These effects become more important for large-aperture RISs and high-frequency systems, where narrow beams are highly sensitive to frequency mismatch.

This array-level effect can be studied even when the RIS elements are modeled as ideal frequency-independent phase shifters. This assumption is useful because it isolates the frequency mismatch caused by the fixed RIS aperture and element spacing. At the same time, practical finite-resolution phase control can still be included through quantized RIS phase profiles. Quantization is a practical constraint because RIS elements usually support only a finite number of phase states, and it can introduce structured distortions and additional lobes in the reflected beampattern.

The second level is the element-level frequency dependence of practical RIS elements. Practical RIS elements do not behave as ideal frequency-independent phase shifters. Their reflection coefficients, achievable phase responses, and phase-shift

ranges depend on the incident frequency, circuit parameters, and feasible tuning range. Therefore, even when the desired reflection direction is known, the RIS hardware may not be able to realize the required phase profile uniformly across different frequencies or bands.

RIS-assisted multi-frequency systems therefore require both array-level and element-level understanding. At the array level, it is necessary to characterize beam squint, beam split, and finite-resolution phase effects. At the element level, it is necessary to understand how circuit parameters and the feasible capacitance window determine the supported phase-shift range and the achievable multi-frequency array gain. These two viewpoints are complementary and motivate the RIS-assisted studies in this dissertation.

1.5 Scope of the Dissertation

In this section, we summarize the work completed and presented in this dissertation.

1.5.1 Molecular Absorption-Aware Resource Allocation in Dense THz Networks with Multi-Connectivity

In Chapter 3 of this dissertation, in order to overcome the challenges mentioned in the previous section for the THz communication systems, we have developed a unified framework to maximize the network sum-rate in a multi-user, multi-BS downlink THz network by optimizing user associations, number and bandwidth of sub-bands in a THz TW, bandwidth of leading and trailing edge-bands in a TW, sub-band assignment, and power allocations. The proposed framework incorporates multi-connectivity and captures the impact of molecular absorption coefficient variations in

a TW, beam-squint, molecular absorption noise, and link blockages.

To make the problem tractable, we first propose a convex approximation of the molecular absorption coefficient using curve fitting in a TW, determine the feasible bandwidths of the leading and trailing edge-bands, and then derive a closed-form optimal solution for the number of sub-bands considering beam-squint constraints. We then decompose joint user associations, sub-band assignment, and power allocation problems into two sub-problems, i.e., **(i)** joint user association and sub-band assignment, and **(ii)** power allocation. To solve the former problem, we analytically prove the unimodularity of the constraint matrix which enables us to relax the integer constraint without loss of optimality. To solve the power allocation sub-problem, an FP-based centralized solution as well as an ADMM-based lightweight distributed solution is proposed. The overall problem is then solved using alternating optimization until convergence. Complexity analysis of the algorithms and numerical convergence are presented. Numerical findings validate the effectiveness of the proposed algorithms and extract useful insights about the interplay of the density of BSs, Average order of multi-connectivity, molecular absorption, hardware impairment, imperfect CSI, and link blockages.

1.5.2 Characterization of Multi-band Networks

Chapter 4 of this dissertation presents the vision of MBN communication networks in 6G, where different frequency bands will coexist with the conventional RF spectrum. We will first pinpoint the fundamental challenges in MBN architectures at the PHY and MAC layers, such as unique channel propagation and estimation issues, user offloading and resource allocation, multi-band transceiver design and antenna

systems, mobility and handoff management, backhauling, etc. We then perform a quantitative performance assessment of the two fundamental MBN architectures, i.e., stand-alone MBN and integrated MBN, considering critical factors like achievable rate, and capital/operational deployment cost, and so on. Our results show that stand-alone deployment is prone to higher capital and operational expenses for a predefined data rate requirement. Stand-alone deployment, however, offers flexibility and enables controlling the number of access points in different transmission bands. In addition, we propose a molecular absorption-aware user offloading metric for MBNs and demonstrate its performance gains over conventional user offloading schemes.

1.5.3 Resource Allocation in Cooperative MBNs in the Presence of Mobility

Chapter 5 of this dissertation develops a comprehensive framework to investigate and optimize the downlink performance of cooperative *MBNs* operating on upper mid-band and THz frequencies, where BSs in each band cooperatively serve users. The framework captures sophisticated features such as near-field channel modeling, fully and partially connected antenna architectures, and users' mobility. First, we consider joint user association and hybrid beamforming optimization to maximize the system sum-rate, subject to power constraints, maximum cluster size of cooperating BSs, and users' QoS constraints. By leveraging fractional programming FP and majorization-minimization techniques, an iterative algorithm is proposed to solve the non-convex optimization problem. We then consider handover (HO)-aware resource allocation for *moving* users in a cooperative UMB/THz MBN. Two HO-aware resource allocation

methods are proposed. The first method focuses on maximizing the HO-aware system sum-rate subject to HO-aware QoS constraints. Using Jensen's inequality and properties of logarithmic functions, the non-convex optimization problem is tightly approximated with a convex one and solved. The second method addresses a multi-objective optimization problem to maximize the system sum-rate, while minimizing the total number of HOs. Numerical results demonstrate the efficacy of the proposed algorithms, cooperative UMB/THz MBN over stand-alone THz networks, as well as the critical importance of accurate near-field modeling in extremely large antenna arrays. Moreover, the proposed HO-aware resource allocation methods effectively mitigate the impact of HOs, enhancing performance in the considered system.

1.5.4 Characterization of Beam-Squint and Beam-Split Effects in RIS-assisted Multi-Frequency Networks

Chapter 6 of this dissertation develops a unified analytical framework to rigorously characterize these effects for uniform planar array (UPA)-based RISs under both continuous and practical finite-resolution phase control. We formalize beam-squint as the deviation of the dominant lobe from the desired elevation–azimuth angle pair and beam-split as the emergence of additional dominant lobes under frequency mismatch. For continuous-phase RISs, we derive necessary and sufficient peak conditions, obtain closed-form elevation–azimuth peak families, and establish explicit feasibility conditions as functions of the frequency ratio, incidence and configuration angles, and element spacing. Extending the analysis to practical b-bit phase quantization, we express the quantized beampattern as a superposition of harmonic array responses

via Fourier expansion of the quantization-error phasor, leading to tractable closed-form per-harmonic peak families, dominance ordering, and feasibility conditions. To capture peak displacement induced by harmonic superposition, we further develop a perturbation-based correction leveraging dominant-harmonic curvature and residual-harmonic gradients, significantly improving peak-location accuracy without exhaustive two-dimensional searches. Numerical simulations validate the analysis, quantify peak deviations and beam-split probabilities, and demonstrate tangible sum-rate gains enabled by beam-split-aware codeword selection in multi-RIS, multi-user, multi-band networks.

1.5.5 Circuit-aware Performance Optimization in RIS-Assisted Sub-6 GHz and Mid-Band Multi-Band Networks

Chapter 7 of this dissertation investigates performance optimization in RIS-assisted multi-band networks by explicitly accounting for the practical frequency-dependent response of RIS elements. Focusing on sub-6 GHz and upper mid-band operation, each RIS element is modeled by a circuit consisting of two inductors, a resistor, and a tunable varactor. Based on this model, we analytically characterize the frequency-dependent phase-shift range induced by the feasible capacitance window. We then formulate the RIS capacitance design as a multi-frequency sum array-gain maximization problem, where a single capacitance vector jointly serves multiple frequencies within the desired band. To solve the resulting non-convex problem, we develop a projected adaptive moment estimation (ADAM)-based algorithm with an adaptive zero-phase-based initialization that exploits the circuit response. Numerical results show that, in the considered native-band optimization cases, the proposed method

achieves higher gain than random-initialized ADAM and a genetic-algorithm baseline. They further show that the achievable gain is strongly governed by the overlap between the feasible capacitance window and the frequency-dependent phase-shift range, and that the element spacing affects the robustness of native-band multi-frequency RIS operation.

1.5.6 Conclusion and Future Work

Chapter 8 concludes the dissertation by summarizing the main findings and outlining potential future research directions.

1.6 Publications List

The author's publications in the PhD are listed below:

1. **M. A. Saeidi**, H. Tabassum, "Characterization of Beam-Squint and Beam-Split Effects in RIS-assisted Multi-Frequency Networks," Submitted to IEEE Transactions on Wireless Communications, 2026

2. **M. A. Saeidi** and H. Tabassum, "Resource Allocation in Cooperative Mid-Band/THz Networks in the Presence of Mobility," in IEEE Transactions on Wireless Communications, vol. 25, pp. 5046-5062, 2026, doi: 10.1109/TWC.2025.3615916.

3. **M. A. Saeidi**, H. Tabassum and M. Alizadeh, "Molecular Absorption-Aware User Assignment, Spectrum, and Power Allocation in Dense THz Networks With Multi-Connectivity," in IEEE Transactions on Wireless Communications, vol. 23, no. 11, pp. 16404-16420, Nov. 2024, doi: 10.1109/TWC.2024.3440888.

4. **M. A. Saeidi**, H. Tabassum and M. -S. Alouini, "Multi-Band Wireless Networks: Architectures, Challenges, and Comparative Analysis," in IEEE Communications Magazine, vol. 62, no. 1, pp. 80-86, January 2024, doi: 10.1109/MCOM.003.2300078.

5. **M. A. Saeidi**, H. Shoaib and H. Tabassum, "A Tractable Handoff-Aware Rate Outage Approximation with Applications to THz-Enabled Vehicular Network Optimization," GLOBECOM 2023 - 2023 IEEE Global Communications Conference, Kuala Lumpur, Malaysia, 2023, pp. 5092-5097.

6. **M. A. Saeidi** and H. Tabassum, "Resource Allocation and Performance Analysis of Hybrid RSMA-NOMA in the Downlink," 2023 IEEE 34th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Toronto, ON, Canada, 2023, pp. 1-6.

7. S. Aboagye, **M. A. Saeidi**, H. Tabassum, Y. Tayyar, E. Hossain, H.- C. Yang, and M.-S. Alouini, "Multi-band wireless communication networks: Fundamentals, challenges, and resource allocation," IEEE Transactions on Communications, pp. 1–1, 2024.

Chapter 2

Literature Review and Preliminaries

In this chapter, we review the literature on THz communications, multi-band networks, massive antenna and cooperative systems, mobility management, and RIS-assisted multi-frequency networks. We then provide a brief discussion of the mathematical preliminaries used in the optimization frameworks developed throughout this dissertation.

2.1 THz Communication

A handful of research works investigated the *spectrum and power allocation in single-cell THz* wireless networks without interference. In [7], a distance-aware and bandwidth-adaptive resource allocation was proposed to maximize the communication distance. The system considered a single BS and interference from adjacent sub-bands is modeled as Gaussian noise. Specifically, sub-band assignment using distance-ranking and power control using convex optimization were proposed. In [8], the minimization of total uplink and downlink delay in a UAV-based THz transmission was considered. The authors optimized bandwidth, users' power allocation, and the UAV's location by decomposing the joint problem into three convex sub-problems and solved them via

alternating optimization. The work in [9] studies how fast and far bits can be transmitted in the THz band by maximizing the product of rate and distance. The paper first obtains the sub-window allocation utilizing the Hungarian algorithm and then alternatively optimizes power allocation and transmission distance. Moreover, the authors in [10] considered sum-rate maximization in a single-cell THz downlink system by optimizing power allocation, sub-band assignment as well as their respective bandwidth allocation. The authors approximated the molecular absorption coefficient as an exponential function of frequency and relax integer constraints by adding a non-convex constraint. Then, they applied successive convex approximation (SCA) to solve the problem. Following a similar approach in [10], the authors considered maximizing the minimum rate in the multi-cell uplink THz network by optimizing power allocation, user association, sub-band assignment as well as their respective bandwidth allocation [11]. Nonetheless, a sub-band is assigned to only one user at a time and the total number of sub-bands is greater than or equal to the number of users. Thus, there is no interference from other users or BSs.

None of the aforementioned research considered resource allocation in large-scale THz networks considering interference, blockages, molecular absorption noise, and/or multi-connectivity. Specifically, the adaptive bandwidth allocation used in [11] and [10] makes the interference formulation challenging in multi-cell systems when considering frequency reuse. In the case of variable bandwidth, interference can be non-uniform due to non-uniform bandwidth divisions across different BSs and varying transmit power on those bandwidth divisions. Moreover, adaptive bandwidth allocation increases the complexity of resource allocation, as in [10] for even a single-cell set-up. Additionally, the impact of molecular absorption noise was not considered.

Recently, a handful of research works touched on the resource allocation in large-scale THz networks though without considering the propagation specifics of THz channel. The work in [12] maximizes energy efficiency in a downlink THz system. Matching theory (switch-matching) is used for sub-channel assignment and power allocation is performed using ADMM. In [13], sub-bands and power allocation are optimized to maximize EE in cache-based THz vehicular networks. This work, without taking TWs and molecular absorption noise into account, first obtains a simplified expression for the rate and then relaxes the binary variables for sub-band assignment. The proposed solution is based on Dinkelbach and Lagrangian dual methods. Authors in [14] investigate multiple single antenna APs transmitting the signal of all users on the same frequency. In terms of molecular absorption windows, the maximum value of molecular absorption noise is considered, which leads to a lower bound on the performance.

To our best knowledge, none of the aforementioned research works addressed the problem of power allocation, joint user association, and sub-band partition and assignment in a large-scale THz network for a given THz TW, while accounting for multi-connectivity, blockage-aware rate, molecular absorption noise, efficient spectrum reuse, beam squint, and hardware impairment. This gap motivates the framework in Chapter 3, where THz transmission windows, edge-band selection, beam-squint-aware sub-band partitioning, low-complexity joint user association and sub-band assignment, and centralized/distributed power allocation are studied jointly.

2.2 Multi-band Networks

Several resource allocation optimization problem formulations and solutions have been proposed for existing MBNs (e.g., RF/THz, RF/VLC/mmWave, RF/VLC, etc.). In [15], the authors considered EE optimization of a cache-enabled stand-alone (SA)-RF/THz NOMA-based MIMO MBN by optimizing user clustering, hybrid precoding, and power allocation under fronthaul capacity constraint and power budget. An alternating optimization (AO) scheme that uses an enhanced K-means machine learning scheme for user clustering, a zero-forcing algorithm for hybrid precoding, and an ADMM for power allocation was proposed. Two user association algorithms based on unsupervised learning were proposed in [16] to minimize the standard deviation of network traffic load in an SA-RF/THz MBN while considering users' QoS requirements. In [2], the authors proposed a deep reinforcement learning-based joint network selection and subchannel allocation for a SA-RF/THz MBN, while considering QoS and reliability constraints. A reinforcement learning-based algorithm to jointly optimize vehicle-to-infrastructure network selection and autonomous driving policies under constraints on number of vehicles each BS can serve was designed in [17].

An iterative power and time slot allocation approach to maximize the sum rate of an SA-RF/VLC MBN subject to constraints on QoS requirements, power budget, and maximum slot allocation was proposed in [18]. The proposed approach involved the integer relaxation and rounding technique for the slot allocation subproblem and a Lagrangian decomposition-based power allocation scheme. The authors in [19] proposed a Dinkelbach-type algorithm to optimize power and bandwidth allocation in an SA-RF/VLC MBN under QoS requirements, power, and bandwidth constraints. In [20], the authors considered an SA-RF/VLC/power line communication MBN and

proposed a Dinkelbach-style energy-efficient joint power and backhaul flow optimization algorithm under power, backhaul capacity, and QoS constraints.

In [21], the authors proposed a genetic algorithm-based and a heuristic-based user association, power, and subchannel allocation algorithms to maximize the sum rate of an SA-RF/mmWave MBN under single connectivity and power constraints. In [22], the authors developed a user association and downlink-uplink power allocation scheme to maximize the sum rate of an SA-RF/mmWave MBN, in which the macro BS and small cell BSs operate on the RF and mmWave bands, respectively. The proposed scheme considered single connectivity constraints, power budgets, and QoS requirements and involved using non-convex solvers to tackle the user association and power allocation subproblems alternatively.

Considering FSO communication in a relay-enabled integrated (Int)-RF/FSO MBN, the authors in [23] proposed a power and subcarrier allocation scheme based on dual decomposition to maximize the sum rate under transmit power and subcarrier allocation constraints. In [24], the authors considered an Int-RF/FSO MBN consisting of an RF/FSO-assisted central processor and a network of connected RF/FSO-enabled radio units. While the central processor uses RF/FSO MBN for fronthaul transmissions, the radio units utilize the RF band to serve users in the access network. A sum rate maximization problem that optimizes the covariance and precoding matrices under fronthaul capacity and transmit power constraints was formulated and solved via AO and semidefinite programming.

Overall, the existing MBN literature mainly focuses on specific combinations of frequency bands and particular optimization problems, such as user association,

power allocation, sub-channel assignment, or backhaul/fronthaul optimization. However, the broader architectural trade-offs of MBNs, including stand-alone versus integrated deployment, deployment cost, mobility, handover behavior, and molecular-absorption- aware user offloading, require further characterization. This motivates Chapter 4, where MBN architectures, resource allocation challenges, and representative case studies are investigated from a unified perspective.

2.3 Massive Antenna and Cooperative Systems

The utilization of ELAA systems at higher frequencies is essential for addressing the challenge of severe path loss. However, due to the larger array aperture of these antenna systems, it is advantageous to adopt NFC channel modeling based on the spherical wave model [25]. The impact of NFC on the sum-rate maximization problem in multi-user systems is examined in [26]. Compared to conventional far-field models, incorporating NFC introduces an additional degree of freedom by enabling beam focusing in both the angular and distance domains. ELAA and NFC have also demonstrated improved performance in various applications, including modular arrays, simultaneous wireless information and power transfer, and physical layer security [25–28].

The work in [29] considers optimizing hybrid beamformers for broadcast and unicast scenarios in a cooperative mmWave system using a minimum variance distortionless response method. The authors also proposed a Bayesian learning-based method for jointly optimizing analog and digital beamformers, demonstrating its effectiveness in enhancing spectral efficiency. In [30], sum power minimization in cooperative mmWave networks with hybrid beamforming is investigated, where analog

beamformers are derived using equal-gain transmission, and digital beamformers are optimized through relaxed semidefinite programming, considering per-user QoS requirements. The study in [31] proposed a cell-free THz network architecture using spatial-spectral coupling to improve link reliability and transmission rates through frequency-dependent antenna selection schemes. Authors in [32] introduced a hybrid FSO/THz backhaul system for cell-free infrastructure-to-vehicle communication in smart cities, employing a novel switching-combining method to boost coverage and reduce outage probability.

The above works show the importance of ELAA, NFC modeling, and cooperative transmission at high frequencies. However, these aspects are often studied separately from multi-band operation and user mobility. In cooperative UMB/THz MBNs, the benefits of cooperation, the use of hybrid beamforming, and the accuracy of near-field channel modeling become coupled with user association and mobility management.

2.4 Mobility Management

The work in [33] derived a HO-aware rate outage probability for THz-enabled vehicular communications and optimized BS density and speed to maximize traffic flow, given traffic constraints and an outage threshold. In [34], the issue of HO management in user-centric cell-free massive MIMO was addressed, where the number of HOs was minimized using a noncausal objective function. The work in [35] investigates user-centric cooperative networks within a stochastic geometry framework, accounting for both HO cost and data rate. A hybrid WiFi and Light Fidelity (LiFi) network were explored in [36], where dynamic load balancing with HO was considered, and an enhanced evolutionary game theory-based load balancing method was proposed.

In [37], vertical handovers between VLC and RF networks were investigated using a Markov decision process. The study in [38] analyzed load balancing in a hybrid LiFi/RF MBN by examining the impact of mobility on handover efficiency between consecutive states. A cached-based mmWave/RF MBN was introduced in [39], utilizing device-level caching and dual-mode BSs to mitigate handover failures and reduce inter-frequency measurement energy consumption. This work presented a dynamic matching game between users and BSs to develop a distributed handover mechanism while adhering to constraints on handover failures and limited cache sizes.

The authors of [40] proposed a handover efficiency-based resource allocation for a VLC system, assuming predictable future CSI to calculate users' future rates for resource allocation. The Lagrangian multiplier method was used to solve the resource allocation problem, maximizing the proportional fairness utility function. A delay minimization problem that considers HO interruptions in distributed computing within a high altitude platform station–assisted vehicular network is addressed in [41], where task-splitting ratios, transmit power, bandwidth, and computing resource allocation are optimized. The work in [42] studies mobility-aware user association in an MBN with two frequency bands in the sub-6 GHz range. In this study, user data rates are forecasted using a recurrent neural network, and the results are used to optimize user association, taking into account BS load and HO costs, using the Frank-Wolfe method.

To the best of our knowledge, none of the aforementioned research works investigated the performance benefits of coordinated transmissions in UMB/THz MBNs or THz-only networks with and without user mobility. Moreover, most existing mobility-aware resource allocation works rely on single-antenna transmissions, and therefore do

not jointly optimize hybrid beamforming, user association, cooperative cluster size, and HO-aware objectives under ELAA-based near-field channel modeling. This gap motivates Chapter 5.

2.5 RIS-assisted MBNs

In RIS-assisted systems, frequency dependence can appear at different levels. At the array level, the physical spacing between RIS elements is fixed, and therefore a phase profile designed at one carrier frequency may not generate the same reflected beam at another carrier frequency. At the element level, practical RIS elements may also have frequency-dependent reflection coefficients, since their achievable phase shifts depend on the incident frequency and the tuning parameters. These two effects are related, but they can be studied separately in order to understand their individual impact.

A number of works have studied array-level beam-squint and beam-split effects in RIS-assisted wideband systems, mainly for far-field THz and mmWave communications. In this direction, the work in [43] studied wideband precoding for RIS-aided THz communications and used time-delay units to compensate for the beam-squint effect. The authors in [44] analyzed wideband THz RIS communications and considered distributed RIS deployment to reduce the beam-squint effect caused by large RIS apertures. A distributed RIS architecture with double-layer true-time-delay units was further studied in [45]. In [46], RIS-assisted wideband THz cell-free massive MIMO systems were considered, where the time delays, RIS phase shifts, and digital precoders were jointly designed. The work in [47] investigated wideband beamforming for STAR-RIS-assisted THz communications, where beam-squint can occur on different sides of the STAR-RIS. Related RIS-assisted wideband works have also considered

time-delay-based or distributed architectures to reduce spatial-wideband and beam-squint effects [48, 49]. These works mainly focus on mitigating the loss caused by frequency-dependent array responses, while the RIS elements are commonly modeled with ideal or frequency-independent phase control.

Beam-squint and beam-split effects are also relevant beyond communication beamforming. In RIS-aided wideband mmWave sensing, the work in [50] used beam-squint and beam-split behavior to identify frequency-dependent dominant directions. Related ISAC studies have also shown that frequency-dependent angular behavior can be exploited for sensing and localization [51, 52]. In addition, beam-squint-aware channel estimation has been studied for THz massive MIMO and RIS-assisted THz systems, where the steering direction changes over different subcarriers and must be included in the channel recovery process [53–56]. Although these studies account for frequency-dependent steering mismatch, to the best of our knowledge, they do not provide a unified framework for characterizing both beam-squint and beam-split directions of a UPA-based RIS under continuous and finite-resolution phase profiles. This gap is addressed in Chapter 6, where beam-squint and beam-split effects are characterized for both continuous and quantized RIS phase profiles.

It is worth noting that the assumption of frequency-independent RIS elements does not remove all practical RIS constraints. In many array-level studies, this assumption is used to isolate the effect of the fixed RIS aperture and element spacing from the element-level circuit response. This separation is useful because beam squint and beam split can occur even when each RIS element provides an ideal unit-modulus phase shift. At the same time, practical RISs usually have finite-resolution phase control due to hardware limitations and control-link constraints [57]. Under quantized

phase shifts, the reflected beampattern is not a shifted version of the continuous-phase beampattern, since quantization can introduce structured distortions and additional dominant lobes.

A second line of work considers the practical frequency-dependent response of RIS elements. The work in [58] developed a practical RIS phase-shift model and showed that the reflection amplitude and phase are coupled through the circuit parameters. In [59], the phase and amplitude squint effects of RIS elements were studied for wideband MIMO-OFDM communications. Practical frequency-dependent RIS models were also proposed in [60, 61], where the RIS response varies with the operating frequency and the element tuning parameters. Moreover, [62] modeled the phase-frequency profile of RIS unit cells and optimized it for wideband OFDM systems. These works show that practical RIS elements are not ideal frequency-independent phase shifters.

Despite these advances, the explicit characterization of the circuit-induced phase-shift range as a function of the circuit parameters, incident frequency, and feasible capacitance window remains less explored. Moreover, how this phase-support structure can guide RIS capacitance optimization, especially through circuit-aware initialization, remains insufficiently characterized. The role of the feasible capacitance window in native-band and cross-band RIS operation also remains insufficiently understood. These gaps motivate Chapter 7, where a circuit-based RIS model is used to characterize the supported phase-shift range and optimize a single capacitance vector for RIS-assisted sub-6 GHz and upper mid-band multi-band networks. Therefore, Chapter 7 complements Chapter 6 by relaxing the ideal element-response assumption and focusing on the hardware-level limitations of practical RIS operation.

2.6 Mathematical Preliminaries

This section provides mathematical preliminaries for the algorithms used in this thesis.

2.6.1 Alternating Optimization

Alternating optimization is an iterative method for solving problems where the objective function depends on multiple sets of variables. The method alternates between optimizing one subset of variables while keeping the others fixed, repeating this process until convergence [63]. Consider the optimization problem:

$$\min_{x,y} f(x, y),$$

where $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ are two sets of variables.

2.6.1.1 Algorithm Steps

1. Initialization: Start with an initial guess for one of the variables, say y^0.

2. Fix y, Optimize x: Solve the subproblem:

$$x^{k+1} = \arg\min_{x} f(x, y^k).$$

3. Fix x, Optimize y: Solve the subproblem:

$$y^{k+1} = \arg\min_{y} f(x^{k+1}, y).$$

4. Convergence Check: Repeat the process until a convergence criterion is met, such as:

$$\|x^{k+1} - x^k\| \leq \epsilon \quad \text{and} \quad \|y^{k+1} - y^k\| \leq \epsilon,$$

where ϵ is a small positive threshold.

Alternating optimization is particularly useful when direct optimization over all variables simultaneously is computationally prohibitive. It ensures a decrease in the objective function value at each step, though convergence to a global minimum is not guaranteed, especially for non-convex functions.

2.6.2 Fractional Programming

Fractional Programming addresses optimization problems involving fractional terms of the form $f(x) = \frac{A(x)}{B(x)}$, where $A(x) \geq 0$ and $B(x) > 0$. These problems are prevalent in wireless communication systems, particularly in optimizing Signal-to-Interference-plus-Noise Ratios (SINRs) or energy efficiency metrics. While classical methods like the Charnes-Cooper transform and Dinkelbach's [64] method are effective for single-ratio problems, many real-world problems involve multiple ratios, such as $f(x) = \sum_{m=1}^{M} \frac{A_m(x)}{B_m(x)}$. These multiple-ratio problems are non-convex and challenging to solve optimally due to their inherent NP-hardness.

A novel *quadratic transform* is proposed to address these challenges, enabling the reformulation of fractional terms into forms suitable for iterative optimization [65, 66]. For a single-ratio problem, $f(x) = \frac{A(x)}{B(x)}$, the quadratic transform introduces an auxiliary variable y, leading to a reformulated objective:

$$g(x, y) = 2y\sqrt{A(x)} - y^2 B(x).$$

This decouples $A(x)$ and $B(x)$, allowing their optimization in an iterative manner. At each iteration, the auxiliary variable y is updated as $y^* = \sqrt{A(x)/B(x)}$, ensuring equivalence in the optimal solution and objective value between the original fractional problem and its transformed version. This approach is extended to multi-ratio problems, $f(x) = \sum_{m=1}^{M} \frac{A_m(x)}{B_m(x)}$, by introducing a set of auxiliary variables $\{y_m\}_{m=1}^{M}$:

$$g(x, \{y_m\}) = \sum_{m=1}^{M} \left(2y_m \sqrt{A_m(x)} - y_m^2 B_m(x) \right).$$

The quadratic transform is particularly effective in wireless communication systems for solving resource allocation problems. For example, in *power control* for single-input single-output (SISO) systems, the objective is to maximize the weighted sum rate $\sum_{i=1}^{N} w_i \log(1 + \text{SINR}_i)$, where each SINR term is a ratio of signal power to interference plus noise. The quadratic transform reformulates the problem into a sequence of convex optimization steps, iteratively updating both the power allocation p and auxiliary variables y.

The convergence properties of the quadratic transform ensure that the iterative algorithms reach a stationary point of the original problem, even for non-convex objectives. Moreover, for concave-convex fractional problems (where $A(x)$ is concave, $B(x)$ is convex, and the feasible set X is convex), the quadratic transform guarantees global optimality for single-ratio problems and local optimality for multi-ratio cases. Closed-form solutions are derivable in specific scenarios, significantly reducing computational overhead compared to numerical optimization.

2.6.3 Scaled ADMM

The *Alternating Direction Method of Multipliers (ADMM)* is a widely used optimization algorithm for solving large-scale constrained problems in distributed settings. The *scaled ADMM* variant simplifies the classical ADMM framework by introducing scaled dual variables, which streamline updates and enhance computational efficiency. This method is particularly effective in scenarios where the optimization problem involves separable objectives or constraints, such as distributed convex optimization tasks.

Consider the generic optimization problem:

$$\min_{x,z} \; f(x) + g(z) \quad \text{subject to } Ax + Bz = c,$$

where $f(x)$ and $g(z)$ are convex functions, and A, B, and c are known matrices and vectors. The classical ADMM approach utilizes the augmented Lagrangian method, incorporating a quadratic penalty to enforce the constraint $Ax + Bz = c$. The augmented Lagrangian for this problem is:

$$L_\rho(x, z, y) = f(x) + g(z) + y^T(Ax + Bz - c) + \frac{\rho}{2}\|Ax + Bz - c\|_2^2,$$

where y is the dual variable (Lagrange multiplier) and $\rho > 0$ is a penalty parameter. Scaled ADMM replaces the dual variable y with a scaled variable $u = \frac{1}{\rho}y$, leading to a simplified formulation:

$$L_\rho(x, z, u) = f(x) + g(z) + \frac{\rho}{2}\|Ax + Bz - c + u\|_2^2 - \frac{\rho}{2}\|u\|_2^2.$$

The scaled ADMM algorithm alternates between updating the primal variables x and z, followed by the scaled dual variable u. In each iteration:

1. Update x: Optimize $f(x)$ while keeping z and u fixed:

$$x^{k+1} = \arg\min_x \ f(x) + \frac{\rho}{2}\|Ax + Bz^k - c + u^k\|_2^2.$$

2. Update z: Optimize $g(z)$ while keeping x and u fixed:

$$z^{k+1} = \arg\min_z \ g(z) + \frac{\rho}{2}\|Ax^{k+1} + Bz - c + u^k\|_2^2.$$

3. Update u: Adjust the scaled dual variable:

$$u^{k+1} = u^k + (Ax^{k+1} + Bz^{k+1} - c).$$

This alternating approach decomposes the global problem into smaller, easier-to-solve subproblems. For example, when $f(x)$ or $g(z)$ has a closed-form solution or simple structure, the updates can be highly efficient. Scaled ADMM is well-suited for large-scale optimization due to its distributed nature, allowing x and z updates to be computed independently or in parallel. It is guaranteed to converge for convex problems under appropriate parameter settings, making it robust and reliable for a wide range of applications. In this work, scaled ADMM enables efficient distributed power allocation by decomposing the problem into manageable subproblems, aligning perfectly with the requirements of large-scale, dense THz networks.

2.6.4 Total Unimodularity

Total unimodularity is a useful property of matrices that allows certain integer linear programs to be solved through their linear programming relaxation without losing integrality. A matrix is totally unimodular if the determinant of every square submatrix is equal to 0, 1, or -1. If the constraint matrix of a linear program is totally unimodular and the right-hand side vector is integral, then every vertex of the feasible polytope is integral. Therefore, the corresponding integer constraints can be relaxed while still obtaining an integer optimal solution. This property is useful in resource allocation problems with binary assignment variables.

2.6.5 Successive Convex Approximation

Successive Convex Approximation (SCA) is an iterative optimization technique designed to solve non-convex problems by approximating them with a sequence of convex subproblems [67]. The goal is to leverage the tractability of convex optimization while ensuring the approximations are close to the original problem. Consider the general non-convex optimization problem:

$$\min_{x \in \mathcal{X}} f(x) \quad \text{subject to} \quad g_i(x) \leq 0,\ i = 1, \ldots, m,$$

where $f(x)$ is the objective function, $g_i(x)$ are the constraint functions, and $\mathcal{X} \subseteq \mathbb{R}^n$ is a convex feasible set. Both the objective and the constraints may be non-convex, making the problem challenging to solve directly. To address this, SCA constructs convex approximations $\hat{f}(x; x^k)$ for the objective function and $\hat{g}_i(x; x^k)$ for the constraints around the current iterate x^k.

At each iteration, the algorithm solves a convex subproblem based on these approximations. Specifically, it minimizes the convex approximation of the objective function, subject to the approximated constraints:

$$x^{k+1} = \arg\min_{x \in \mathcal{X}} \hat{f}(x; x^k) \quad \text{subject to} \quad \hat{g}_i(x; x^k) \leq 0, \ \forall i.$$

The solution x^{k+1} becomes the new iterate, and the process is repeated. The convex approximations are updated at each step to reflect the current solution, ensuring the subproblem evolves to represent the original non-convex problem more accurately.

This iterative approach is particularly useful for problems where the non-convexities can be locally approximated by convex functions. The mathematical rigor of SCA lies in its ability to decompose complex problems into manageable convex subproblems, allowing for efficient and systematic optimization. However, the design of the approximations plays a crucial role in its effectiveness, as poorly designed approximations may lead to slow convergence or suboptimal solutions.

2.6.6 Projected ADAM

Adaptive moment estimation (ADAM) is a first-order iterative method that updates the optimization variables using adaptive estimates of the first and second moments of the gradient. For constrained problems, a projected version of ADAM can be used, where each gradient-based update is followed by a projection onto the feasible set. If x^t denotes the variable at iteration t and g^t denotes the gradient, ADAM updates moving averages of the gradient and squared gradient, and then uses them to scale the update direction.

Let m^t and v^t denote the first- and second-moment estimates of the gradient,

respectively. The ADAM updates can be written as

$$m^t = \beta_1 m^{t-1} + (1 - \beta_1) g^t,$$

$$v^t = \beta_2 v^{t-1} + (1 - \beta_2)(g^t \odot g^t),$$

where β_1 and β_2 are momentum parameters, and $\odot$ denotes element-wise multiplication. After bias correction, the variable is updated along the scaled gradient direction. In projected ADAM, this update is followed by a projection step:

$$x^{t+1} = \Pi_{\mathcal{X}} \left(x^t - \alpha \frac{\hat{m}^t}{\sqrt{\hat{v}^t} + \epsilon} \right),$$

where α is the step size, ϵ is a small positive constant, and $\Pi_{\mathcal{X}}(\cdot)$ denotes the projection onto the feasible set $\mathcal{X}$.

Chapter 3

Molecular Absorption-Aware User Assignment, Spectrum, and Power Allocation in Dense THz Networks with Multi-Connectivity

3.1 Overview

The evolution of wireless cellular systems, known as the 6G, will be driven by the need to support revolutionary services such as connected autonomous vehicles, extended reality (XR), digital twins, and immersive remote presence. To meet the requirements of these bandwidth-hungry services, 6G wireless systems must deliver unprecedented data rates [68–72]. THz spectrum has the potential to offer much wider transmission bandwidths with extreme data rates (in the order of multi-Gbps). Nonetheless, THz spectrum comes with its own set of challenges. Different from the traditional radio frequencies ($<$ 6GHz), THz transmissions are susceptible to high propagation losses due to molecular absorption in the environment, molecular absorption noise, blockages, and hardware impairments. The ultra-broad bandwidth contributes to

non-negligible thermal noise. The THz transmissions are also vulnerable to beam squint which causes beams at various frequencies to radiate in different physical directions leading to system performance degradation [73–75]. As a result, dense THz network deployments will be critical to overcome the molecular absorption, blockage, and noise effects.

To enable large-scale THz communication, several critical factors must be considered, including low complexity, efficient and optimal power allocation, user association, and sub-band assignment. It is also essential to account for interference, molecular absorption noise resulting from molecular absorption phenomena at high frequencies, and blockage, while also addressing the beam squint phenomenon. Furthermore, molecular absorption is more pronounced in the THz band compared to the mmWave band. Due to the numerous absorption peaks in the THz band, we need to consider transmission windows (TWs), which are unnecessary in the mmWave band. The high molecular absorption at different frequencies in the THz band makes the channel, and achievable rates frequency-dependent, adding challenges to resource allocation. Therefore, we must take specific steps, such as managing TWs, determining trailing and edge bands to maintain constant molecular absorption, and deciding the total number of sub-bands to avoid beam squint phenomena in the THz band.

In light of these challenges, this chapter presents a novel and low-complexity framework with multi-connectivity that caters to both centralized and distributed power allocation, while tackling the joint user association and sub-band assignment problem and carefully considering blockage and beam squint effects. This work incorporates multiple constraints into the optimization problem to determine the parameters of a given TW and the optimal number of sub-bands, aiming to enhance the overall

system sum-rate.

The contributions of this chapter are summarized as follows:

• We maximize the overall network sum-rate in a downlink THz network with multiple users and multiple BSs by strategically optimizing various parameters. These include user assignments, the count and width of sub-bands within a THz TW, the width of leading and trailing edge-bands within the TW, sub-band, and power allocations. Our approach considers multi-connectivity, factors in molecular absorption coefficient variations, beam-squint, molecular absorption noise, and link blockages. For beam-squint, we employ the fractional bandwidth concept to set an upper limit for each sub-band bandwidth. In particular, compared to the existing works, making uniform partitions of sub-bands over a given TW allows us to concurrently enable frequency reuse by including the interference terms and determine an optimal number of sub-bands in favor of achieving a higher system sum-rate and avoiding beam squint. Additionally, incorporating interference renders the problem highly non-convex, necessitating the development of low-complexity algorithms for solutions.

• To determine the number of sub-bands required, we initially employ curve-fitting techniques to create an analytical function representing the TW of interest. This fitted curve allows us to derive the values of the TW's edge bands, which are the regions where transmission should be avoided. By using the estimated edge band values alongside the beam squint-aware upper bound on the bandwidth, we obtain a lower bound on the total number of sub-bands needed for the system.

• With the number of sub-bands established, we divide the optimization problem into two sub-problems: joint user association and sub-band assignment, and power

allocation. To address the non-convex nature of the association and sub-band assignment sub-problem, we reduce the dimension of the power allocation variables from 3-dimensional (3D) to 2D. Leveraging the structure of the constraints and employing the concept of total unimodularity, we transform the nondeterministic polynomial time (NP-hard) integer programming into a low-complexity linear programming, offering a novel and comprehensive proof in this context.

• We address the power allocation sub-problem through two proposed FP-based approaches. One approach involves centralized FP, while the other introduces a novel and low-complexity scheme based on the ADMM methodology. This distributed scheme reduces signaling and communication overhead significantly. Employing alternating optimization, the overall low-complexity algorithm iteratively solves the joint user association and sub-band assignment, as well as the power allocation sub-problem.

• Numerical results indicate that leveraging multi-connectivity presents performance improvement compared to single-connectivity. Additionally, when evaluating the impact of antennas' directionality, blockage, molecular absorption coefficient, hardware impairment, imperfect CSI, and the number of BSs, the proposed methods outperform the traditional benchmarks. Our proposed framework offers a general solution that can also be applied to lower frequencies. By setting the molecular absorption coefficient to zero and removing TWs, and by using the respective channel model and adjusting the formulations, our framework can be adapted for lower frequencies such as mmWave and sub-6GHz.

The rest of this chapter is organized as follows. Section 3.2 presents the considered downlink multi-cell THz system model, including the transmission model,

beam-squint-aware THz channel model, SINR formulation, and blockage-aware rate model. Section 3.3 formulates the sum-rate maximization problem and discusses the associated constraints, molecular absorption approximation, edge-band selection, sub-band design, and objective-function transformation. Section 3.4 develops the proposed centralized solution, including joint user association, sub-band assignment, and fractional-programming-based power allocation. Section 3.5 presents the low-complexity distributed power allocation approach and analyzes the computational complexity of the proposed algorithms. Section 3.6 provides numerical results and discussions to evaluate the impact of antenna directionality, number of BSs, molecular absorption, blockage density, convergence behavior, available bandwidth, hardware impairments, and imperfect CSI. Finally, Section 3.8 summarizes the chapter.

Notations: This chapter uses boldface lowercase letters to represent vectors and boldface capital letters to represent matrices. $\{0,1\}^{B\times N}$ and $\mathbb{R}_{+}^{B\times N}$ denote the space of $B\times N$ binary-valued and non-negative real-valued matrices, respectively. $\mathbf{I}_N$ stands for the $N\times N$ identity matrix. $|x|$ is used for the absolute value of x, $\|\mathbf{x}\|_2$ denotes the L2 norm of vector $\mathbf{x}$, and $\mathbf{X}^T$ is the transpose of matrix $\mathbf{X}$. $\mathrm{diag}(\mathbf{X}, B)$ is a block diagonal matrix, such that $\mathbf{X}$ is repeated B times diagonally, and $\otimes$ indicates Kronecker product.

3.2 System Model and Assumptions

3.2.1 Downlink Transmission Model

As shown in Fig. 3.1, we consider a downlink THz transmission in a multi-user multi-cell system, where N users are served by B BSs. We define $\mathcal{B} = \{1, 2, \cdots, b, \cdots, B\}$ as the set of all THz BSs deployed in a cellular region and the set of all users in the

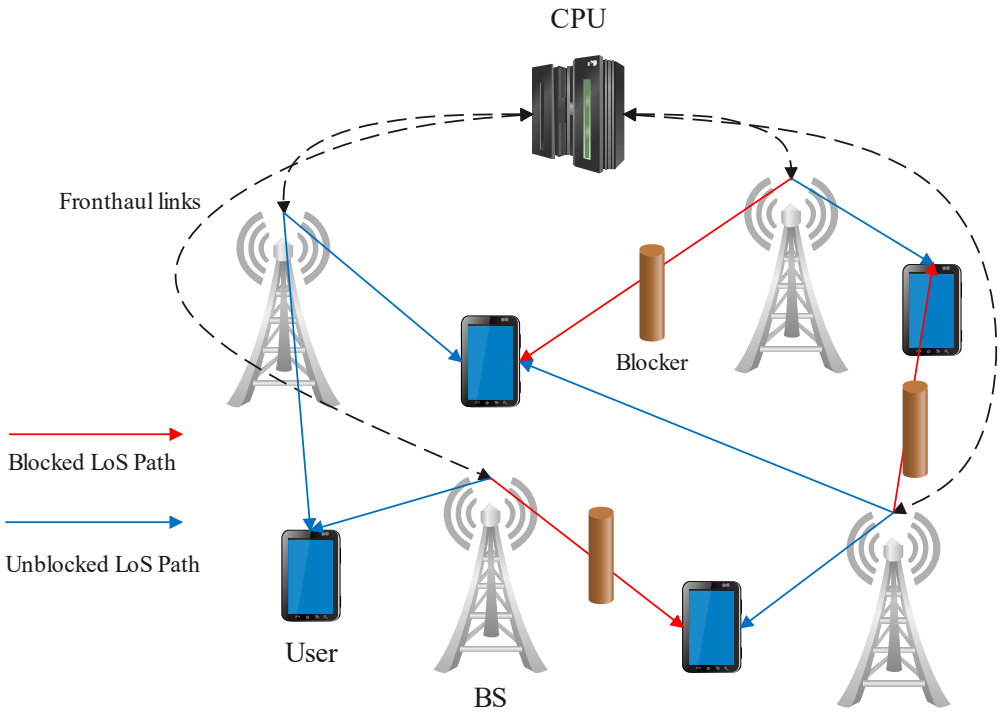


Figure 3.1: System model illustrating BSs and users in the presence of randomly distributed blockers. *Adapted from [76].* © *2024 IEEE.*

cellular region is denoted by $\mathcal{N} = \{1, 2, \cdots, n, \cdots, N\}$.

Each of the THz BS and the user is equipped with a single directional antenna. Due to the existence of path-loss peaks in the THz spectrum, we consider downlink communication in specified transmission windows where the impact of molecular absorption is minimum. We consider dividing a given TW into S orthogonal sub-bands which is determined based on beam-squint and the bandwidth of edge-bands (as will be detailed later). Additionally, guard bands are utilized between each sub-band to prevent inter-band interference. The set of sub-bands is referred

to as $\mathcal{S} = \{1, 2, \ldots, s, \ldots, S\}$. Each sub-band is reused by all BSs, hence intra-band interference exists.

Without loss of generality, we consider $f_1 < f_2 < \cdots < f_S$. Thus, the central frequency of s-th sub-band is given as $f_s = f_I + w_I + (s - 0.5)w + (s - 1)w_G$, where w is the bandwidth of each sub-band, f_I is the initial carrier frequency of the TW of interest, w_I is the edge bandwidth at the beginning of the TW, and w_G is the guard band between two consecutive sub-bands. The total bandwidth of a given TW can thus be given as, $w_T = f_E - f_I$, where f_E is the final frequency of the TW. Thus, spectrum allocation is performed in the spectral range of $[f_I + w_I, \quad f_E - w_E]$, where w_E is the edge bandwidth at the end of the given TW. Both w_I and w_E are designed to avoid transmissions at frequencies with significantly varying molecular absorption. We assume multi-connectivity, i.e., each user can be associated with multiple BSs on a different sub-band. Moreover, each BS assigns a sub-band to one user to avoid intra-band within each cell.

3.2.2 Beam Squint Consideration and THz Channel Model

The ultra-broadband nature of THz transmission can simply increase the ratio between the central frequency, f_s, and its bandwidth, w. This ratio, called fractional bandwidth, can lead to the beam squint effect, in which each frequency component is divided into various directions leading to beam gain degradation [77]. In order to reduce the impact of the beam squint effect, we consider limiting the fractional bandwidth, i.e., B_s of the s-th sub-band within B_{th}.

$$B_s = \frac{w}{f_s} = \frac{w}{f_I + w_I + (s - 0.5)w + (s - 1)w_G} \leq B_{th}. \tag{3.1}$$

The channel power of the THz link from b-th BS to n-th user on sub-band s is modeled as:

$$|h_{b,s,n}|^2 = G_b^{\mathrm{Tx}} G_n^{\mathrm{Rx}} c^2 d_{n,b}^{-2} \exp\left(-k(f_s) d_{n,b}\right) \chi_{b,s,n} (4\pi f_s)^{-2}, \tag{3.2}$$

where G_b^{Tx} and G_n^{Rx} are the antenna gains of b-th BS and n-th user, respectively [78], and $d_{n,b}$ is the distance between b-th BS and n-th user. Also, c is the speed of light, $k(f_s)$ is molecular absorption coefficient at f_s, and N_0 is the additive Gaussian noise spectral density, and $\chi_{n,b,s}$ denotes Nakagami-m[1] fading channel to model line-of-sight (LoS) transmissions [78]. [2] The antenna gain for all BSs and users is a function of the angle from the antenna boresight direction, $\theta^g \in [-\pi, \pi)$, and main lobe beamwidth, Θ^g, where $g \in \{\mathrm{Tx}, \mathrm{Rx}\}$ [79]. Thus, the antenna gain can be modeled as follows:

$$G^g(\theta^g) = \begin{cases} G^g_{\max}, & |\theta^g| \leq \Theta^g \\ G^g_{\min}, & |\theta^g| > \Theta^g, \end{cases} \tag{3.3}$$

where $G^g_{\max}$ and $G^g_{\min}$ are the array gains of the main and side lobes, respectively. When the main lobes of a user and its associated BS coincides, the antenna gain of that link becomes $G^{\mathrm{Tx}}_{\max} G^{\mathrm{Rx}}_{\max}$. Similar to [80], the alignment probability between the main lobes of interfering BSs and a specific user is denoted by $q = \frac{\theta^{\mathrm{Tx}} \theta^{\mathrm{Rx}}}{4\pi^2}$, and the side-lobes are negligible, i.e., $G^g_{\min} \approx 0$. As the antenna's directionality increases, the main lobe's beamwidth proportionally decreases. Consequently, based on (3.3), when the antenna

[1] In the THz spectrum, the non-line of sight (NLoS) paths resulting from scattering and refraction are negligible [13]. Also, the Nakagami-m fading channel model is a general model that can approximate Rician fading for large values of shape parameter ($m \gg 1$).

[2] All channels are assumed to be estimated during the channel training phase and sent to the central processing unit (CPU) via fronthaul links. However, the impact of imperfect channel state information (CSI) on system performance is discussed in Section VI.

exhibits higher directionality (characterized by a smaller Θ^g value), q diminishes because both θ^{Tx} and θ^{Rx} are limited to values smaller than Θ^g. Thus, we consider q as a factor denoting the directionality of the antennas at both users and BSs. Specifically, if $q \approx 0$, it indicates that the interfering BSs have high directionality and that inter-band interference can be disregarded, whereas if $q = 1$, the user captures the entire transmit power of the interfering BSs.

3.2.3 THz SINR Model and Blockage-Aware Rate Model

Let us define the joint user association and sub-band assignment variable as:

$$a_{b,s,n} = \begin{cases} 1, \text{ if user } n \text{ is associated with BS } b \text{ on sub-band } s \\ 0, \quad \text{otherwise .} \end{cases} \tag{3.4}$$

Then, the matrix of joint user association and sub-band assignment is given as $\mathbf{A} \in \{0,1\}^{BS\times N}$, such that $a_{b,s,n}$ is the entry of $\mathbf{A}$ at $((b-1)S+s)$-th row and n-th column, and BS means the multiplication of B and S.

We define the matrix representation of the power allocation as $\mathbf{P} = \{p_{b,s,n}\} \in \mathbb{R}_{+}^{BS\times N},\ \forall b \in \mathcal{B},\ n \in \mathcal{N}, s \in \mathcal{S}$ where $p_{b,s,n}$ is the transmit power from b-th BS to n-th user on sub-band s. Then, the spectral reuse in all BSs, which leads to intra-band interference, is modeled as follows:

$$I_{b,s,n}(\mathbf{P},\mathbf{A}) = \sum_{b'\in\mathcal{B}'} \sum_{n'\in\mathcal{N}'} q a_{b',s,n'} p_{b',s,n'} |h_{b',s,n}|^2, \tag{3.5}$$

such that $\mathcal{N}' \triangleq \mathcal{N}\backslash\{n\}$, $\mathcal{B}' \triangleq \mathcal{B}\backslash\{b\}$.

In addition to thermal noise, it is essential to consider molecular absorption noise,

which is induced by the re-radiation of absorbed signal energy, in the THz band. In particular, as a result of using superconductive materials in the THz frequency, thermal noise can be decreased, therefore molecular absorption noise would be comparable to thermal noise and must be taken into account [81]. Molecular absorption noise for user n associated to BS b over sub-band s is modeled as follows:

$$\Phi_{b,s,n}(\mathbf{A},\mathbf{P}) = a_{b,s,n}p_{b,s,n}|\tilde{h}_{b,s,n}|^2 + \sum_{b'\in\mathcal{B}'}\sum_{n'\in\mathcal{N}'} qa_{b',s,n'}p_{b',s,n'}|\tilde{h}_{b',s,n}|^2, \tag{3.6}$$

where the first term is related to the molecular noise of the desired link and the second term is related to the molecular noise of the interfering links. Also, the channel gain of the molecular absorption noise is given as $|\tilde{h}_{b,s,n}|^2 = Cd_{n,b}^{-2}(1 - \exp(-k(f_s)d_{n,b}))\chi_{b,s,n}{f_s}^{-2}$, such that $C = G_b^{\mathrm{Tx}}G_n^{\mathrm{Rx}}c^2(4\pi)^{-2}\ \forall b,n$. Hence, the cumulative interference and molecular absorption noise from all interfering BSs, which is the summation of $I_{b,s,n}(\mathbf{P},\mathbf{A})$ and the second term in (3.6), is obtained as:

$$\bar{I}_{b,s,n}(\mathbf{P},\mathbf{A}) = \sum_{b'\in\mathcal{B}'}\sum_{n'\in\mathcal{N}'} qa_{b',s,n'}p_{b',s,n'}\left|\bar{h}_{b',s,n}\right|^2, \tag{3.7}$$

where $|\bar{h}_{b,s,n}|^2 = Cd_{n,b}^{-2}\chi_{b,s,n}{f_s}^{-2}$.

Then, the SINR at n-th user associated with b-th BS on sub-band s can be given as [80]:

$$\gamma_{b,s,n}(\mathbf{A},\mathbf{P},w) = \frac{a_{b,s,n}p_{b,s,n}|h_{b,s,n}|^2}{\bar{I}_{b,s,n}(\mathbf{P},\mathbf{A}) + a_{b,s,n}p_{b,s,n}|\tilde{h}_{b,s,n}|^2 + N_0 w}. \tag{3.8}$$

Therefore, the achievable rate at n-th user associated with b-th BS over sub-channel s is stated as follows:

$$\bar{R}_{b,s,n}(\mathbf{A},\mathbf{P},w) = w\log_2(1+\gamma_{b,s,n}(\mathbf{A},\mathbf{P},w)). \tag{3.9}$$

THz transmission is susceptible to blockage due to high penetration losses [73]; thus, we incorporate link blockage probability. We define $\mathbf{\Psi} \in \{0,1\}^{B\times N}$ as the indicator variable for blockage such that $\mathbf{\Psi}_{b,n} = 1$ if there is a blockage between user n and BS b, and otherwise $\mathbf{\Psi}_{b,n} = 0$. Each element of $\mathbf{\Psi}$ follows the following distribution [79, 82]:

$$\mathbf{\Psi}_{b,n} = \begin{cases} 1, & \text{With probability } \ 1 - e^{-\eta d_{b,n}} \\ 0, & \text{With probability } \ e^{-\eta d_{b,n}}, \end{cases} \tag{3.10}$$

where η is the density of blockers in a given area. For a specific η, the likelihood of blocking a link between a user and BS increases exponentially with their distance. The random blockage matrix $\mathbf{\Psi}$ is multiplied by the channel between a user and BS. The blockage-aware rate between a user and BS is thus modeled as follows: If $\mathbf{\Psi}_{b,n} = 0$, the rate is $R_{b,s,n}(\mathbf{A}, \mathbf{P}, w, \mathbf{\Psi}) = \bar{R}_{b,s,n}(\mathbf{A}, \mathbf{P}, w)$, otherwise $R_{b,s,n}(\mathbf{A}, \mathbf{P}, w, \mathbf{\Psi}) = 0$.

3.3 Problem Formulation, Constraints, and Transformations

3.3.1 Problem Formulation

The sum-rate maximization problem in a multi-user, multi-BS downlink THz network is formulated in the following:

$$\mathcal{P}_1: \quad \underset{\mathbf{A},\mathbf{P},w,w_E,w_I,S}{\text{maximize}} \quad \sum_{b\in\mathcal{B}}\sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}} R_{b,s,n} \tag{3.11}$$

$$\text{s.t. } \mathbf{C_1}: \sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}} a_{b,s,n}p_{b,s,n} \le P_b^{\max}, \forall b\in\mathcal{B}, \quad \mathbf{C_2}: \sum_{b\in\mathcal{B}} a_{b,s,n} \le 1, \forall s\in\mathcal{S}, n\in\mathcal{N},$$

$$\mathbf{C_3}: \sum_{n\in\mathcal{N}} a_{b,s,n} = 1, \forall b\in\mathcal{B}, s\in\mathcal{S}, \qquad \mathbf{C_4}: \Gamma_n^L \le \sum_{b\in\mathcal{B}}\sum_{s\in\mathcal{S}} a_{b,s,n} \le S, \forall n\in\mathcal{N},$$

$$\mathbf{C_5}: w_I + Sw + (S-1)w_G + w_E = w_T, \quad \mathbf{C_6}: w \le \frac{f_I B_{\text{th}} + w_I B_{\text{th}} + (s-1)B_{\text{th}}}{1-(0.5-s)B_{\text{th}}} \forall s\in\mathcal{S}$$

$$\mathbf{C_7}: |k(f_E - w_E) - k(f_I + w_I)| \le \epsilon, \qquad \mathbf{C_8}: a_{b,s,n} \in \{0,1\}, \forall n\in\mathcal{N}, b\in\mathcal{B}, s\in\mathcal{S}.$$

In $\mathcal{P}_1$, $\mathbf{C_1}$ constrains the maximum transmit power limit at each BS, where $P_b^{\max}$ is the maximum transmit budget of b-th BS. In order to leverage multi-connectivity, constraint $\mathbf{C_2}: \sum_{b\in\mathcal{B}} a_{b,s,n} \le 1, \forall s\in\mathcal{S}, n\in\mathcal{N}$ is taken into account, which limits a user to be associated with only one BS over the same sub-band, while allowing the association of a user with more than one BS across different sub-bands. $\mathbf{C_3}$ guarantees that each sub-band can only be assigned to one user in a given BS. $\mathbf{C_4}$ considers the fairness of users by imposing Γ_n^L as the respective lower limit on the total number of sub-bands that can be assigned to a user. $\mathbf{C_5}$ constrains the total bandwidth of the TW of interest.

$\mathbf{C_6}$ avoids the beam squint phenomena, and $\mathbf{C_7}$ ensures the molecular absorption coefficient do not vary beyond ϵ within the TW. The edge bands, i.e., w_E and w_I, can guarantee this condition. $\mathbf{C_8}$ denotes that the joint user association and sub-band assignment variables are binary.

The problem $\mathcal{P}_1$ is a mixed-integer non-linear programming (MINLP) problem and is thus non-convex. The nonlinearity of the problem stems from the fact that the

objective function is formulated as a fractional function of user association, sub-band assignment, and power allocation variables, along with their pairwise multiplication. Additionally, the problem is combinatorial due to the involvement of binary variables and non-convex due to the non-convexity of the molecular absorption coefficient in $\mathbf{C_7}$. An additional hurdle in tackling problem $\mathcal{P}_1$ lies in determining the appropriate values for the number of sub-bands and their respective bandwidths. This complexity arises from the objective function, which involves a summation across the number of sub-bands. Consequently, finding the optimal values for S and w is not a straightforward task as the two variables are intertwined. Furthermore, when considering a given S, the optimization of bandwidth cannot be conducted independently, as it is intrinsically tied to the total number of sub-bands. We decompose and solve the problem in two steps. First, we obtain a feasible solution for the bandwidths of the leading and trailing edge spectrum, the number of sub-bands, and their transmission bandwidths by proposing a convexification of the molecular absorption coefficient. Then, joint user association and sub-band assignment is performed in an optimal manner leveraging the uni-modularity of the problem followed by power optimization in an iterative manner using alternating optimization.

3.3.2 Molecular Absorption Coefficient

The molecular absorption coefficient, which is negligible in lower frequencies, is an important aspect of the THz band and can be different based on various environmental conditions, such as pressure, gas molecules, and so on [72]. The molecular absorption coefficient of the isotopologue t of gas g for a molecular volumetric density

at temperature T and pressure of p can be obtained as:

$$k(f) = \sum_{(t,g)} \frac{p^2 T_{\text{sp}} q^{(t,g)} N_A S^{(t,g)} f \tanh\left(\frac{hcf}{2k_b T}\right)}{p_0 T_k V T^2 f_c^{(t,g)} \tanh\left(\frac{hcf_c^{(t,g)}}{2k_b T}\right)} F^{(t,g)}(f), \tag{3.12}$$

where p_0 is the reference pressure at 1 atm, T_{sp} is the temperature at standard pressure, and the mixing ratio of gases is $q^{(t,g)}$. N_A is Avogadro's constant, and V is the gas constant. $S^{(t,g)}$ is the line intensity, showing the strength of the absorption by a specific type of molecule. f denotes the frequency of the electromagnetic wave, and $f_c^{(t,g)}$ represents the resonant frequency of gas g. h and k_b denote the Planck's and Boltzmann's constants, respectively. For the frequency f, the Van Vleck-Weisskopf asymmetric line shape is given by:

$$F^{(t,g)}(f) = \frac{100c\alpha^{(t,g)} f}{\pi f_c^{(t,g)}} \left(\frac{1}{G^2 + \left(\alpha^{(t,g)}\right)^2} + \frac{1}{H^2 + \left(\alpha^{(t,g)}\right)^2} \right), \tag{3.13}$$

such that $G = f + f_c^{(t,g)}$, $H = f - f_c^{(t,g)}$, and the Lorentz half-width is obtained by:

$$\alpha^{(t,g)} = \left(\left(1 - q^{(t,g)}\right) \alpha_{\text{air}}^{(t,g)} + q^{(t,g)} \alpha_0^{(t,g)} \right) \left(\frac{p}{p_0}\right) \left(\frac{T_0}{T}\right)^l, \tag{3.14}$$

where the air broadened half-widths, $\alpha_{\text{air}}^{(t,g)}$, self-broadened half-widths, $\alpha_0^{(t,g)}$, and temperature broadening coefficient l can be obtained directly from the high-resolution transmission molecular absorption (HITRAN) database [83].

3.3.3 Convexifying the Molecular Absorption Coefficient

After obtaining the molecular absorption coefficient based on the previous section calculations and HITRAN database, we need to find an analytical expression for the

molecular absorption coefficient as a function of frequency to enable us to derive the widths of the leading and trailing edge sub-bands, namely w_I and w_E. Additionally, using a convex function that spans the entire TW would be advantageous, allowing the use of numerical techniques like interior point methods to effectively address the feasibility problem in determining w_I and w_E. Therefore, we propose the following convex approximation of the molecular absorption coefficient using curve-fitting techniques.

$$\bar{k}(f) = t_1 \exp\left(\frac{-1}{(t_2 f + t_3)^2}\right) + t_4, \tag{3.15}$$

where $t_1, t_4 > 0$, and t_2, t_3 are the fitted curve parameters. Table II presents values for curve-fitting parameters utilizing the suggested function for a variety of THz transmission windows. Note that $\bar{k}(f)$ is a convex function when $\frac{-\sqrt{6}-3t_3}{3t_2} \leq f \leq \frac{\sqrt{6}-3t_3}{3t_2}$, which can be satisfied in the TW of interest by selecting appropriate f_I and f_E. Since the proposed function is twice differentiable as shown below, the convexity of the approximated molecular absorption coefficient can be verified.

$$\frac{d^2\bar{k}(f)}{df^2} = \frac{-2t_1 t_2^2 \exp\left(\frac{-1}{(t_2 f + t_3)^2}\right)\left(3(t_2 f + t_3)^2 - 2\right)}{(t_2 f + t_3)^6}.$$

3.3.4 Leading and Trailing Edge Bands of the TW

For a given TW, f_I, f_E, and curve-fitting parameters of the respective TW obtained from Table II, we can use the convexity of $\bar{k}(f)$ and solve the following set of inequalities derived from constraint $\mathbf{C_7}$ to obtain the feasible values of w_E and w_I.

Table 3.1: Curve-fitting parameters based on the approximated function of molecular absorption coefficient, $\bar{k}(f)$. *Adapted from [76]. © 2024 IEEE.*

TW Label	**TW Range (THz)**	t_1	t_2	t_3	t_4
TW_1	0.448 - 0.531	1.1	-14.5233	7.1063	0.0173
TW_2	0.624 - 0.722	0.8	11.3600	-7.6442	0.0139
TW_3	0.78 - 0.915	0.5	9.6221	-8.1526	0.0139
TW_4	0.997 - 1.063	1.2	-21.6372	22.28	0.0882

$$\bar{k}(f_E - w_E) - \bar{k}(f_I + w_I) \leq \epsilon \tag{3.16a}$$

$$\bar{k}(f_I + w_I) - \bar{k}(f_E - w_E) \leq \epsilon, \tag{3.16b}$$

Considering $f_I + w_I < f_e - w_E$, by using the interior-point method for solving (3.16), the feasible values of leading and trailing edge bands, i.e., $\hat{w}_I$ and $\hat{w}_E$ are obtained. After having $\hat{w}_I$ and $\hat{w}_E$ obtained, we determine the total number of sub-bands and their respective bandwidth as discussed in the next subsection.

3.3.5 Bound on the Number of THz Sub-bands with Beam-Squint

Acquiring the optimal S and w presents a challenge as they are interwined and due to the inclusion of a summation over the variable S in the objective function of problem $\mathcal{P}_1$. Hence, for the given $\hat{w}_I$ and $\hat{w}_E$, a feasible lower bound on the number of sub-bands required to overcome beam-squint can be obtained using constraint $\mathbf{C_6}$. We first recast constraint $\mathbf{C_6}$ as $\frac{\frac{w}{B_{\text{th}}} - f_I - \hat{w}_I + 0.5w + w_G}{w + w_G} \leq s, \forall s \in \mathcal{S}$. When $\mathbf{C_6}$ is satisfied for $\min(\mathcal{S}) = 1$, it avoids the beam-squint at all sub-bands, thus by substituting $s = 1$, we get:

$$\frac{w}{B_{\text{th}}} - f_I - \hat{w}_I - 0.5w \leq 0. \tag{3.17}$$

By substituting $w = \frac{w_T - \hat{w}_E - \hat{w}_I - (S-1)w_G}{S}$ from $\mathbf{C_5}$ into (3.17), the lower bound on S is given by:

$$S_{\text{LB}} \geq \left\lceil \frac{2(w_T - \hat{w}_I + w_G - \hat{w}_E) + B_{\text{th}}(-w_T + \hat{w}_I - w_G + \hat{w}_E)}{2w_G + 2B_{\text{th}}(f_I + \hat{w}_I) + B_{\text{th}} w_G} \right\rceil. \tag{3.18}$$

Then, using S_{LB}, the respective bandwidth is given as follows:

$$w = \frac{w_T - \hat{w}_E - \hat{w}_I - (S_{\text{LB}} - 1)w_G}{S_{\text{LB}}}. \tag{3.19}$$

Lemma 3.1. *For a given power allocation and user association, the objective function in $\mathcal{P}_1$, i.e., $\sum_{b\in\mathcal{B}} \sum_{n\in\mathcal{N}} \sum_{s\in\mathcal{S}} R_{b,s,n}$ is a decreasing function of the total number of sub-bands, S. Therefore, the lower bound on $S_{\text{LB}} = S^*$ is optimal.*

Proof. Let us assume that $\lceil \alpha S \rceil \leq S$ sub-bands are assigned to a given user, such that $\alpha \in [0, 1]$. Then, the achievable rate of the user is given as follows:

$$\begin{aligned} R_n &= \sum_{b\in\mathcal{B}} \sum_{s=1}^{\lceil \alpha S \rceil} \frac{\bar{w} - (S-1)w_G}{S} \log\left(1 + \frac{\frac{BP_b^{\max}}{\lceil \alpha S \rceil}}{\frac{BP_b^{\max}}{\lceil \beta S \rceil} + N_0(\frac{\bar{w} - (S-1)w_G}{S})}\right) \\ &= B\lceil \alpha S \rceil \frac{\bar{w} - (S-1)w_G}{S} \log\left(1 + \frac{\frac{BP_b^{\max}}{\alpha}}{\frac{BP_b^{\max}}{\beta} + N_0(\bar{w} - (S-1)w_G)}\right) \\ &= f_W(S) f_R(S), \end{aligned} \tag{3.20}$$

where $\bar{w} = w_T - \hat{w}_E - \hat{w}_I$. α and β determine the portion of BSs and sub-bands assigned to a given user. It can be seen that $f_W(S)$ is a decreasing function of S, while $f_R(S)$ is an increasing function of S. However, due to the near-to-zero value of N_0, $N_0(\bar{w} - (S-1)w_G)$ is negligible compared to the first term in the denominator, even for large values of S. Also, the absolute value of $f_W(S)$ is always greater than that of $f_R(S)$, i.e., $|f_W(S)| >> |f_R(S)|$ due to the large bandwidths in the THz band.

Therefore, the product of these two functions is monotonically decreasing with respect to S. Based on the similar discussion for the rest of the users, we can conclude that the objective function $\sum_{b\in\mathcal{B}}\sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}} R_{b,s,n}$ is a decreasing function of S. □

Based on the result of **Lemma 3.1**, we can select S_{LB} as the optimal value for the total number of sub-bands and their respective bandwidth using (3.19).

3.3.6 Transformation of the Objective Function

The optimization problem $\mathcal{P}_1$ is MINLP and the non-linearity comes from the objective function since $\gamma_{b,s,n}(\mathbf{A},\mathbf{P})$ is a non-convex function of the joint power allocation, user association, and sub-band assignment variables. To handle the non-convexity of the objective function, we transform the objective function using the following Lemma.

Lemma 3.2. *By reducing the dimension of the power allocation variables from $B\times S\times N$ to $B\times S$ using $\sum_{n\in\mathcal{N}} a_{b,s,n}p_{b,s,n}=p_{b,s}$, the objective function in $\mathcal{P}_1$ can be written as $f(\mathbf{A},\mathbf{P})=\sum_{b\in\mathcal{B}}\sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}} a_{b,s,n}w\log(1+\gamma_{b,s,n}(\mathbf{P}))$.*

Proof. Considering the fact that each BS transmits to only one user over a given sub-band, i.e., equation in $\mathbf{C_3}$, we have $\sum_{n\in\mathcal{N}} a_{b,s,n}p_{b,s,n}=p_{b,s}$. Also, $\sum_{n'\in\mathcal{N}'} a_{b',s,n'}p_{b',s,n'}=p_{b',s}-a_{b',s,n}p_{b',s,n}$. The cumulative interference can thus be written as follows:

$$\bar{I}_{b,s,n}(\mathbf{P},\mathbf{A})=\sum_{b'\in\mathcal{B}'} qp_{b',s}\left|\bar{h}_{b',s,n}\right|^2-\sum_{b'\in\mathcal{B}'} qa_{b',s,n}p_{b',s,n}\left|\bar{h}_{b',s,n}\right|^2. \tag{3.21}$$

Since each user can only associate with one BS over a given sub-band, i.e., equation in $\mathbf{C_2}$, the second term in (3.21) is zero if the association in the numerator of $\gamma_{b,s,n}$ is one, due to having $a_{b,s,n}+\sum_{b'\in\mathcal{B}'} a_{b',s,n}=1$. Otherwise, the SINR and the rate

are zero. A similar analogy can be used for the numerator, hence we can take the binary variable out of the logarithm and both the numerator and denominator while considering the 2D power allocation. Thus, the objective function can be replaced with $f(\mathbf{A},\mathbf{P}) = \sum_{b\in\mathcal{B}}\sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}} a_{b,s,n} w \log(1+\gamma_{b,s,n}(\mathbf{P}))$. □

3.4 Proposed Algorithm for WSR Maximization in a Multi-Cell THz Network

In this section, we present a two-stage algorithm, where the initial stage determines the edge bands, total number of sub-bands, and bandwidth as described in the previous section. The second stage iteratively optimizes the joint user association and sub-band assignment, and power allocation variables.

3.4.1 Joint User Association and Sub-band Assignment

Given $\{S^*, w, \hat{w}_I, \hat{w}_E, \mathbf{P}\}$, $\mathcal{P}_1$ can be rewritten as follows:

$$\begin{aligned} \mathcal{P}_2: \quad & \underset{\mathbf{A}}{\text{maximize}} \quad \sum_{b\in\mathcal{B}}\sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}} w\log(1+\gamma_{b,s,n}(\mathbf{A})) \\ & \text{s.t.} \quad \mathbf{C_2},\ \mathbf{C_3},\ \mathbf{C_4},\ \mathbf{C_8}. \end{aligned} \tag{3.22}$$

Similar to the proof of Lemma 3.2, employing constraint $\mathbf{C_3}$ results in $\sum_{n\in\mathcal{N}} a_{b,s,n} p_{b,s,n} = p_{b,s}$, causing the power budget constraint in $\mathbf{C_1}$ to become $\sum_{s\in\mathcal{S}} p_{b,s} \leq P_b^{\max}$, $\forall b$. Consequently, $\mathbf{C_1}$ can be eliminated from the problem of joint user association and sub-band assignment since it no longer contains the binary variable. Due to the unimodular structure of $\mathcal{P}_2$, we can solve the problem *optimally* without utilizing greedy or

matching solutions. To this end, we need to define $\mathbf{a} = [a_1, \ldots, a_i, \ldots, a_{BSN}]^T \in \{0,1\}^{BSN\times 1}$, consisting of all association and assignment variables, such that $a_i = a_{b,s,n}$, and $i = n + N(s-1) + NS(b-1)$. In the following, all constraints are written in matrix format.

$$\mathbf{C} : [\mathbf{I}_{NS\times NS}, \ldots, \mathbf{I}_{NS\times NS}] = \mathbf{1}^T_{B\times 1} \otimes \mathbf{I}_{NS\times NS}, \tag{3.23a}$$

$$\mathbf{D} : \text{diag}(\mathbf{L}, B),\ \mathbf{L} = \mathbf{I}_{S\times S} \otimes \mathbf{1}^T_{N\times 1}, \tag{3.23b}$$

$$\mathbf{E} : [\mathbf{I}_{N\times N}, \ldots, \mathbf{I}_{N\times N}] = \mathbf{1}^T_{BS\times 1} \otimes \mathbf{I}_{N\times N}, \tag{3.23c}$$

$$\mathbf{F} : [-\mathbf{I}_{N\times N}, \ldots, -\mathbf{I}_{N\times N}] = -\mathbf{1}^T_{BS\times 1} \otimes \mathbf{I}_{N\times N} = -\mathbf{E}. \tag{3.23d}$$

In (3.23), $\mathbf{C}$ and $\mathbf{D}$ represent constraints $\mathbf{C_2}$ and $\mathbf{C_3}$, respectively. Matrices $\mathbf{E}$ and $\mathbf{F}$ can be obtained by dividing the inequalities in constraint $\mathbf{C_4}$ into two constraints, such that $\mathbf{E} : \sum_{b\in\mathcal{B}} \sum_{s\in\mathcal{S}} a_{b,s,n} \leq S, \forall n \in \mathcal{N}$, and $\mathbf{F} : -\sum_{b\in\mathcal{B}} \sum_{s\in\mathcal{S}} a_{b,s,n} \leq -\Gamma^L_n, \forall n \in \mathcal{N}$. Moreover, the right-hand side of the above constraints can be reformulated as $\mathbf{k}_1 = \mathbf{1}_{NS\times 1}$, $\mathbf{k}_2 = \mathbf{1}_{BS\times 1}$, $\mathbf{k}_3 = [S, \ldots, S] \in \mathbb{N}^{NB\times 1}$, and $\mathbf{k}_4 = [-\Gamma^L_1, \ldots, -\Gamma^L_N] = \mathbf{\Gamma}^L \in \mathbb{N}^{NB\times 1}$. Finally, using the above-mentioned, the optimization problem $\mathcal{P}_2$ can be reformulated as follows:

$$\begin{aligned} \mathcal{P}_3 :\quad & \underset{\mathbf{a}}{\text{maximize}} \sum_{i=1}^{BSN} \mathbf{q}^T\mathbf{a} \\ \text{s.t. } & \mathbf{C_2} : \mathbf{Ca} \preceq \mathbf{k}_1,\ \mathbf{C_3} : \mathbf{Da} = \mathbf{k}_2,\ \mathbf{C_{4a}} : \mathbf{Ea} \preceq \mathbf{k}_3, \\ & \mathbf{C_{4b}} : \mathbf{Fa} \preceq \mathbf{k_4},\ \mathbf{C_{8a}} : a_i \in \{0,1\}, \forall i \in \{1, \ldots, BSN\}, \end{aligned} \tag{3.24}$$

where $\mathbf{q} = [w \log(1 + \gamma_i(\boldsymbol{p}))]^T, \forall i \in \{1, \ldots, BSN\}$, such that γ_i is the i-th element of the vector form of its respective matrix and i is obtained similar to the index of a_i. In what follows, leveraging the concepts of unimodularity, we prove that the integer linear programming problem $\mathcal{P}_3$ can be converted into a linear programming problem without loss of optimality, and the solutions are integer values. To this end, we first define $\mathbf{T} = [\mathbf{C}^T, \mathbf{D}^T, \mathbf{E}^T, \mathbf{F}^T]^T$ as the constraint matrix of $\mathcal{P}_3$, and $\mathbf{k} = [\mathbf{k}_1^T, \mathbf{k}_2^T, \mathbf{k}_3^T, \mathbf{k}_4^T]^T$.

Definition 3.1 (Totally Unimodular Matrix (TUM))**.** *A matrix* $\mathbf{M}$ *is TUM if each square submatrix has a determinant of 0, 1, or -1, i.e., all entries are either 0 or* $\pm$ *1 [84].*

Theorem 3.1. *If the constraint matrix* $\mathbf{T}$ *is totally unimodular, and* $\mathbf{k}$ *is integer, problem* $\mathcal{P}_3$ *can be relaxed by removing the integer constraint, and the resulting problem, which is linear programming, can be solved via the dual-simplex method without loss of optimality.*

Proof. See Theorem (Unimodularity implies integrality) in [85]. □

In the following, we first perform some operations on the matrix ($\mathbf{T}$) that preserve TUM (as described below) leading to an equivalent matrix $\mathbf{T}^{(3)}$. We then show that $\mathbf{T}^{(3)}$ is TUM which in turn implies that $\mathbf{T}$ is TUM.

Definition 3.2 (Operations that Preserve Unimodularity)**.** *Total unimodularity is preserved under transposition, permutation of rows or columns, adding or removing repetitive rows or columns, scaling rows or columns by* -1*, and pivoting over any non-zero element [84].*

Consider the constraint matrix $\mathbf{T} = [\mathbf{C}^T, \mathbf{D}^T, \mathbf{E}^T, \mathbf{F}^T]^T$. From the definition of matrix $\mathbf{E}$ and $\mathbf{F}$, we know that $\mathbf{F} = -\mathbf{E}$ (see (3.23d)). Thus, $\mathbf{T}$ can be written as $\mathbf{T} = [\mathbf{C}^T, \mathbf{D}^T, \mathbf{E}^T, -\mathbf{E}^T]^T$. Using **Definition 2**, we can multiply all rows of matrix $\mathbf{F}$ by -1, resulting in $\mathbf{T}^{(1)} = [\mathbf{C}^T, \mathbf{D}^T, \mathbf{E}^T, \mathbf{E}^T]^T$. Also, since matrix $\mathbf{E}$ appeared twice, we can eliminate it, which results in the matrix $\mathbf{T}^{(2)} = [\mathbf{C}^T, \mathbf{D}^T, \mathbf{E}^T]^T$ (as shown in the figure). Next, we apply the pivoting operation (defined below) to the non-zero elements of the first $\mathbf{I}$ matrix of $\mathbf{C}$.

Definition 3.3. *Pivoting Operation [84]: Consider the following $m \times n$ matrix $\mathbf{A}$, the pivoting operation over a non-zero element, like e, results in a matrix $\mathbf{B}$ that has the form:*

$$A = \begin{bmatrix} e & \mathbf{c}^T \\ \mathbf{b} & \mathbf{D} \end{bmatrix} \longrightarrow B = \begin{bmatrix} -e & e\mathbf{c}^T \\ e\mathbf{b} & \mathbf{D} - e\mathbf{b}\mathbf{c}^T \end{bmatrix} \tag{3.25}$$

where $\mathbf{b}$ and $\mathbf{c}$ are $(m-1)$ and $(n-1)$ dimensional column vectors, and $\mathbf{D}$ is a $(m-1) \times (n-1)$ matrix. Note that pivoting over the element $-e$ of matrix $\mathbf{B}$ results in the original matrix $\mathbf{A}$. The pivoting operation is critical in establishing the total unimodularity of matrix T.

After pivoting, the resulting matrix, $\mathbf{T}^{(3)}$, has the structure (as shown in Fig. 3.2). Since the transformation from $\mathbf{T}$ to $\mathbf{T}^{(3)}$ uses the operations mentioned in **Definition 2**, and they preserve total unimodularity, it suffices to show that $\mathbf{T}^{(3)}$ is TUM, which in turn proves that $\mathbf{T}$ is TUM.

Lemma 3.3. *Matrix $\mathbf{T_3}$ and in turn $\mathbf{T}$ is TUM and $\mathbf{k}$ is integer. Thus, $a_i \in \{0, 1\}$ becomes $a_i \geq 0$ and the resulting linear program has the same integer optimum point as $\mathcal{P}_3$.*

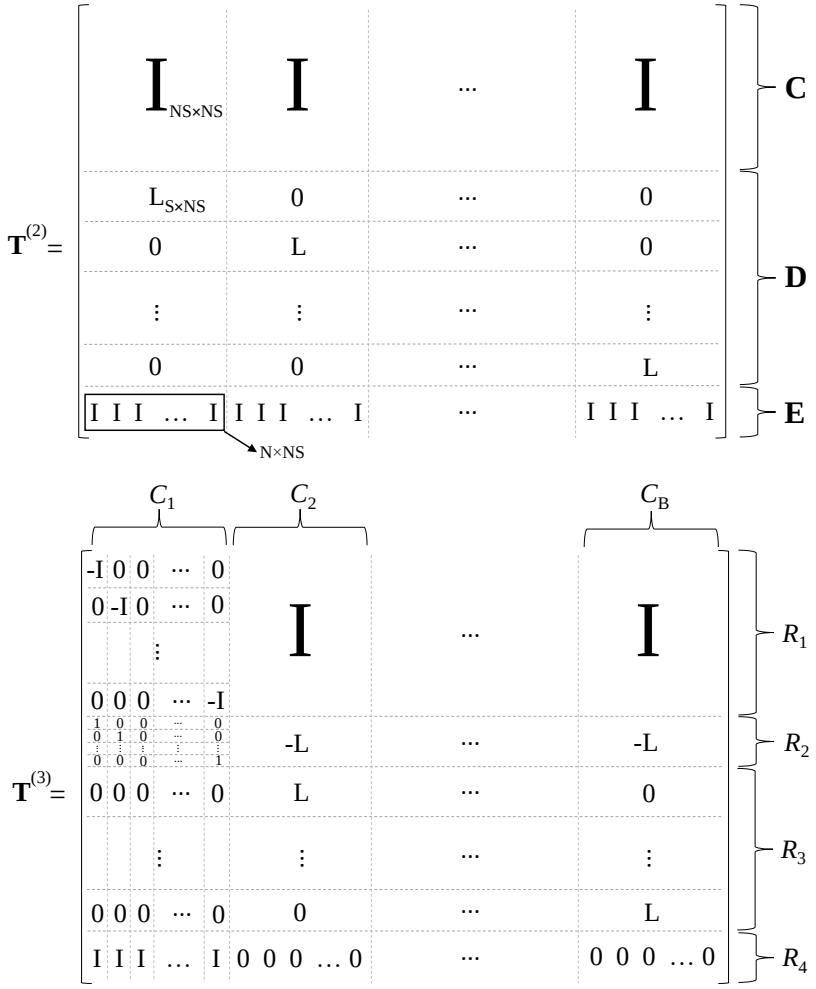


Figure 3.2: Graphical illustration of $\mathbf{T}^{(2)}$ and $\mathbf{T}^{(3)}$. *Adapted from [76]. © 2024 IEEE.*

Proof. See Section 3.7.1. □

According to **Theorem 1** and **Lemma 3**, the integer programming in $\mathcal{P}_3$ can be efficiently solved in polynomial time using the dual-simplex method, and the joint user association and sub-band assignment matrix $\mathbf{A}$ can be reconstructed using vector $\mathbf{a}$. Table-3.2 compares the time complexity of the proposed linear programming (LP) and binary solution (Branch-and-bound method using Mosek solver [86]).

Table 3.2: Comparison of time complexity between the proposed LP and binary solution of $\mathcal{P}_3$ *Adapted from [76]. © 2024 IEEE.*

Problem Size	$B=2,S=2,N=2$	$B=4,S=8,N=6$	$B=6,S=10,N=12$	$B=15,S=20,N=30$
Proposed LP	0.0090 [Sec]	0.0109 [Sec]	0.0129 [Sec]	0.1268 [Sec]
Binary Solution	0.1794 [Sec]	0.2882 [Sec]	0.4124 [Sec]	2.4470 [Sec]

3.4.2 Centralized Power Allocation with Fractional Programming and Bisection Search

Given $\{S^*, w, \hat{w}_I, \hat{w}_E, \mathbf{A}\}$, $\mathcal{P}_1$ must be solved for $\mathbf{P}$, which is a non-convex problem due to the variables in both numerator and denominator of the objective function. Solving a such problem is NP-hard and finding a globally optimum solution is computationally prohibitive. Applying the change of variables $\bar{p}_{b,s} = \sqrt{p_{b,s}}, \forall n \in \mathcal{N}$, the optimization problem can be formulated as:

$$\begin{aligned}
\mathcal{P}_4 : \underset{\bar{\mathbf{P}},\gamma_{b,s,n}}{\text{maximize}}\ & f(\bar{\mathbf{P}},\gamma_{b,s,n}) = \sum_{b\in\mathcal{B}}\sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}} a_{b,s,n} w \log(1+\gamma_{b,s,n}) \\
\text{s.t.}\quad & \mathbf{C_1} : \sum_{s\in\mathcal{S}} \bar{p}_{b,s}^2 \le P_b^{\max}, \forall b \in \mathcal{B}, \\
& \mathbf{C_9} : \gamma_{b,s,n} = \frac{\bar{p}_{b,s}^2 |h_{b,s,n}|^2}{\sum\limits_{b'\in\mathcal{B}'} q\bar{p}_{b',s}^2 \left|\bar{h}_{b',s,n}\right|^2 + \bar{p}_{b,s}^2 \left|\tilde{h}_{b,s,n}\right|^2 + N_0 w}.
\end{aligned} \tag{3.26}$$

Note that the summations over the users and the binary variables in $\mathbf{C_1}$ are omitted according to **Lemma 2**. We first apply the Lagrangian dual transform to relocate the SINR expression within $\gamma_{b,s,n}$ to the outside of the logarithm. By doing so, we are able to apply the quadratic transformation and solve the problem iteratively with closed-form expressions. Subsequently, we employ the quadratic transformation to convert the SINR into a concave quadratic function.

In order to apply the Lagrangian dual transform, for fixed $\bar{\mathbf{P}}$, the Lagrangian function of $\mathcal{P}_4$ with respect to $\mathbf{C_9}$ can be written as follows:

$$\mathcal{L}(\bar{\mathbf{P}},\boldsymbol{\gamma},\boldsymbol{\lambda}) = \sum_{b\in\mathcal{B}}\sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}}\Big[a_{b,s,n}w\log(1+\gamma_{b,s,n}) \\ -\lambda_{b,s,n}\left(\gamma_{b,s,n} - \frac{\bar{p}_{b,s}^2|h_{b,s,n}|^2}{\sum\limits_{b'\in\mathcal{B}'} q\bar{p}_{b',s}^2|\bar{h}_{b',s,n}|^2+\bar{p}_{b,s}^2|\tilde{h}_{b,s,n}|^2+N_0w}\right)\Big], \tag{3.27}$$

where $\boldsymbol{\gamma}\in\mathbb{R}_+^{BS\times N}$ is the matrix of all $\gamma_{b,s,n}$ and $\boldsymbol{\lambda}\in\mathbb{R}_+^{BS\times N}$ is the matrix of all Lagrange multipliers $\lambda_{b,s,n}$. Given $\bar{\mathbf{P}}$, the first-order condition is satisfied with respect to $\gamma_{b,s,n}$ when $\frac{\partial\mathcal{L}(\bar{\mathbf{P}},\boldsymbol{\gamma},\boldsymbol{\lambda})}{\partial\gamma_{b,s,n}}=0$, which leads to

$$\lambda_{b,s,n}^* = \frac{a_{b,s,n}w\left(\sum\limits_{b'\in\mathcal{B}'} q\bar{p}_{b',s}^2|\bar{h}_{b',s,n}|^2+\bar{p}_{b,s}^2|h_{b,s,n}|^2+N_0w\right)}{\sum\limits_{b'\in\mathcal{B}'} q\bar{p}_{b',s}^2|\bar{h}_{b',s,n}|^2+\bar{p}_{b,s}^2(|h_{b,s,n}|^2+|\tilde{h}_{b,s,n}|^2)+N_0w} \tag{3.28}$$

and

$$\gamma_{b,s,n}^* = \frac{\bar{p}_{b,s}^2|h_{b,s,n}|^2}{\sum\limits_{b'\in\mathcal{B}'} q\bar{p}_{b',s}^2|\bar{h}_{b',s,n}|^2+\bar{p}_{b,s}^2|\tilde{h}_{b,s,n}|^2+N_0w} \tag{3.29}$$

Next, by substituting $\lambda_{b,s,n}^*$ into the Lagrange function in (3.27), we have:

$$f_1(\bar{\mathbf{P}},\boldsymbol{\gamma}) = \sum_{b\in\mathcal{B}}\sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}}\Big[a_{b,s,n}w\log(1+\gamma_{b,s,n}) - a_{b,s,n}w\gamma_{b,s,n} \\ + \frac{a_{b,s,n}w(1+\gamma_{b,s,n})\bar{p}_{b,s}^2|h_{b,s,n}|^2}{\sum\limits_{b'\in\mathcal{B}'} q\bar{p}_{b',s}^2|\bar{h}_{b',s,n}|^2+\bar{p}_{b,s}^2(|h_{b,s,n}|^2+|\tilde{h}_{b,s,n}|^2)+N_0w}\Big]. \tag{3.30}$$

According to [65, 66], strong duality holds for the objective function in (3.26). Thus, the transformed objective function $f_1(\bar{\mathbf{P}},\boldsymbol{\gamma})$ can be maximized instead of $f(\bar{\mathbf{P}},\boldsymbol{\gamma})$.

The transformed objective function $f_1(\bar{\mathbf{P}}, \boldsymbol{\gamma})$ is still a non-convex function of $\bar{\mathbf{P}}$ due to the sum-of-ratios. Hence, the following quadratic transformation is applied to handle the non-convexity.

$$\begin{aligned} f_2(\bar{\mathbf{P}}, \boldsymbol{\gamma}, \boldsymbol{Y}) = & \sum_{b\in\mathcal{B}} \sum_{n\in\mathcal{N}} \sum_{s\in\mathcal{S}} \big[a_{b,s,n} w [\log(1+\gamma_{b,s,n}) - \gamma_{b,s,n}] \\ & + 2 y_{b,s,n} \sqrt{a_{b,s,n} w (1+\gamma_{b,s,n}) \bar{p}_{b,s}^2 |h_{b,s,n}|^2} \\ & - y_{b,s,n}^2 \big(\sum_{b'\in\mathcal{B}'} q \bar{p}_{b',s}^2 \big| \bar{h}_{b',s,n} \big|^2 + \bar{p}_{b,s}^2 (|h_{b,s,n}|^2 + |\tilde{h}_{b,s,n}|^2) + N_0 w \big) \big], \end{aligned} \tag{3.31}$$

where $\mathbf{Y} \in \mathbb{R}_+^{BS\times N}$ is the matrix of auxiliary variables $y_{b,s,n}$, and $f_2(\bar{\mathbf{P}}, \boldsymbol{\gamma}, \boldsymbol{Y})$ is a concave function of $\boldsymbol{Y}$.

Remark 3.1. *The optimal solution of* $\max_{\boldsymbol{Y}} f_2(\bar{\mathbf{P}}, \boldsymbol{\gamma}, \boldsymbol{Y})$*, i.e.,* $\boldsymbol{Y}^*$*, results in* $f_1(\bar{\mathbf{P}}, \boldsymbol{\gamma})$*. The reason is that by setting the partial derivative of* $f_2(\bar{\mathbf{P}}, \boldsymbol{\gamma}, \boldsymbol{Y})$ *with respect to* $\boldsymbol{Y}$*, and substituting the resulting* $\boldsymbol{Y}$ *into* (3.31)*, the objective function* $f_1(\bar{\mathbf{P}}, \boldsymbol{\gamma})$ *can be obtained.*

Based on *Remark 1*, for fixed $\bar{\mathbf{P}}$ and $\boldsymbol{\gamma}$, $\boldsymbol{Y}^*$ can be found by setting $\frac{\partial f_2(\bar{\mathbf{P}}, \boldsymbol{\gamma}, \boldsymbol{Y})}{\partial y_{b,s,n}} = 0$, i.e.,

$$y_{b,s,n}^* = \frac{\sqrt{a_{b,s,n} w (1+\gamma_{b,s,n}) \bar{p}_{b,s}^2 |h_{b,s,n}|^2}}{\sum_{b'\in\mathcal{B}'} q \bar{p}_{b',s}^2 \big| \bar{h}_{b',s,n} \big|^2 + \bar{p}_{b,s}^2 (|h_{b,s,n}|^2 + |\tilde{h}_{b,s,n}|^2) + N_0 w}. \tag{3.32}$$

Then, to obtain optimal power allocation given $\boldsymbol{\gamma}, \boldsymbol{Y}$, we rewrite $\mathcal{P}_4$ as:

$$\begin{aligned} \mathcal{P}_5: \quad & \underset{\bar{\mathbf{P}}}{\text{maximize}} \quad f_2(\bar{\mathbf{P}}, \boldsymbol{\gamma}, \boldsymbol{Y}) \\ \text{s.t.} \quad & \mathbf{C_1}: \sum_{s\in\mathcal{S}} \bar{p}_{b,s}^2 \le P_b^{\max}, \forall b \in \mathcal{B}. \end{aligned} \tag{3.33}$$

The optimization problem $\mathcal{P}_5$ is a convex problem and a closed-form expression for $\bar{p}_{b,s}$ can be found. By introducing Lagrange multiplier μ_b for $\mathbf{C_1}$, the Lagrange function is given by

$$\mathcal{L}_2(\bar{\mathbf{P}}, \boldsymbol{\mu}) = f_2(\bar{\mathbf{P}}, \boldsymbol{\gamma}, \boldsymbol{Y}) - \sum_{b\in\mathcal{B}} \mu_b (\sum_{s\in\mathcal{S}} \bar{p}_{b,s}^2 - P_b^{\max}), \tag{3.34}$$

where $\boldsymbol{\mu} = \{\mu_1, \ldots, \mu_B\}$. The first-order optimality condition yields $\frac{\partial \mathcal{L}_2(\bar{\mathbf{P}}, \boldsymbol{\mu})}{\partial \bar{p}_{b,s}} = 0$, so we have

$$\bar{p}_{b,s}^* = \frac{\sum\limits_{n\in\mathcal{N}} y_{b,s,n} \sqrt{a_{b,s,n} w (1+\gamma_{b,s,n}) |h_{b,s,n}|^2}}{\sum\limits_{b'\in\mathcal{B}'} \sum\limits_{n\in\mathcal{N}} q y_{b',s,n}^2 |\bar{h}_{b',s,n}|^2 + \sum\limits_{n\in\mathcal{N}} y_{b,s,n}^2 (|h_{b,s,n}|^2 + |\tilde{h}_{b,s,n}|^2) + \mu_b}. \tag{3.35}$$

The Lagrange multiplier $\boldsymbol{\mu}$ satisfies the complementary slackness for the constraint $\mathbf{C_1}$, which ensures μ_b is either zero or a positive value that ensures the power budget constraint. It can be seen that $\bar{p}_{b,s}^*$ is a decreasing function of μ_b. Thus, the values of $\boldsymbol{\mu}$ can be found by solving $J(\mu_b) = \sum\limits_{s\in\mathcal{S}} \left(\bar{p}_{b,s}(\mu_b)\right)^2 - P_b^{\max} = 0$ using bisection search.

The overall procedure to obtain optimal power allocation is summarized in **Algorithm-1**.

3.5 Low-Complexity Distributed Solution and Complexity Analysis

In this section, we provide a light-weight distributed power control solutions for time-sensitive wireless applications and analyze the worst-case complexity of the proposed algorithms.

3.5.1 Low-Complexity Distributed Solution

The optimization problem $\mathcal{P}_5$ is a quadratically constrained quadratic program (QCQP) that can be efficiently solved using the ADMM method in a distributed manner [87]. We first reformulate $\mathcal{P}_5$ by introducing first and second blocks of primal variables. We then proceed to employ the scaled ADMM approach to achieve power allocation in a distributed manner with reduced complexity.

By introducing the auxiliary variables $\boldsymbol{\Delta} = [\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_B]^T \in \mathbb{R}_+^{B\times S}$ and $\boldsymbol{\delta_b} = [\delta_{b,1}, \ldots, \delta_{b,S}]^T$, such that $\delta_{b,s}$ is equivalent to $\bar{p}_{b,s}$, problem $\mathcal{P}_5$ is reformulated as follows:

$$\begin{aligned} \mathcal{P}_6: \quad & \underset{\bar{\mathbf{P}},\boldsymbol{\Delta}}{\text{minimize}} \quad -f_2(\bar{\mathbf{P}},\boldsymbol{\gamma},\boldsymbol{Y}) + \mathbb{1}_{\mathbf{C_1}}(\boldsymbol{\Delta}) \\ & \text{s.t.} \quad \mathbf{C_{12}}: \boldsymbol{\Delta} = \bar{\mathbf{P}}, \end{aligned} \tag{3.36}$$

In order to take the constraint $\mathbf{C_1}$ into account, for fixed $\boldsymbol{Y}$, and $\boldsymbol{\gamma}$, we define the

Algorithm 1 Centralized power allocation via fractional programming

Input: Initializing $\bar{\mathbf{P}}$, $\boldsymbol{\gamma}$ with feasible values, maximum number of iterations $L_{\max}$, stopping accuracy ϵ_1, bisection search parameters.

for $t = 0, 1, \ldots$ **do**
  Update $\boldsymbol{Y}^{(t+1)}$ using (3.32)
  for $\forall b \in \mathcal{B}$ **do**
    if $J(\mu_b = 0) \leq 0$ **then**
      Update $\bar{p}_{b,s}^{(t+1)}$ using (3.35) when $\mu_b = 0$
    else
      Find μ_b^* using bisection search and update $\bar{p}_{b,s}^{(t+1)}$ using (3.35)
    end if
  end for
  Update $\boldsymbol{\gamma}^{(t+1)}$ using (3.29)
  Until $|f(\bar{\mathbf{P}},\boldsymbol{Y},\boldsymbol{\gamma})^{(t+1)} - f(\bar{\mathbf{P}},\boldsymbol{Y},\boldsymbol{\gamma})^{(t)}| < \epsilon_1$ or $t = L_{\max}$
end for

Output: The optimal power allocation: $\mathbf{P}^* = (\bar{\mathbf{P}}^{(t+1)})^2$.

following indicator function, which is used in the objective function of $\mathcal{P}_6$.

$$\mathbb{1}_{\mathbf{C_1}}(\boldsymbol{\Delta}) = \begin{cases} 0 & \sum\limits_{s\in\mathcal{S}} \delta_{b,s}^2 \leq P_b^{\max}, \forall b \in \mathcal{B} \\ \infty & \text{o/w .} \end{cases} \tag{3.37}$$

According to the scaled ADMM method, we consider $\bar{\mathbf{P}}$ and $\boldsymbol{\Delta}$ as the first and second blocks of primal variables [87]. Then, we introduce $\boldsymbol{Z} = [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_B]^T \in \mathbb{R}_+^{B\times S}$, and $\boldsymbol{z_b} = [z_{b,1}, \ldots, z_{b,S}]^T$ as the dual variables, accounting for the equality constraint in $\mathbf{C_{12}}$. The augmented Lagrangian function for optimization problem $\mathcal{P}_6$ is obtained as follows:

$$\mathcal{L}_D(\bar{\mathbf{P}}, \boldsymbol{\Delta}, \boldsymbol{Z}) = -f_2(\bar{\mathbf{P}}, \boldsymbol{\gamma}, \boldsymbol{Y}) + \mathbb{1}_{\mathbf{C_1}}(\boldsymbol{\Delta}) + \frac{\rho}{2}\sum_{b\in\mathcal{B}} \|\boldsymbol{\delta}_b - \bar{\boldsymbol{p}}_b + \boldsymbol{z}_b\|_2^2, \tag{3.38}$$

where $\bar{\boldsymbol{p}}_b = [\bar{p}_{b,1}, \ldots, \bar{p}_{b,S}]^T$, and $\rho > 0$ is the penalty factor of the augmented Lagrangian. The augmented Lagrangian maximization and multiplier update can be conducted iteratively through the ADMM approach.

In what follows, we first explain how to obtain $\bar{\mathbf{P}}$, $\boldsymbol{\Delta}$, and $\boldsymbol{Z}$, and then describe how the process is distributed over all BSs.

3.5.1.1 Obtaining the first block of primal variables, $\bar{\mathbf{P}}$

The power allocation matrix, $\bar{\mathbf{P}}$ is derived by solving the following optimization problem:

$$\bar{\mathbf{P}} \leftarrow \arg\min_{\bar{\mathbf{P}}} \mathcal{L}_D(\bar{\mathbf{P}}, \boldsymbol{\Delta}, \boldsymbol{Z}) = \arg\min_{\bar{\mathbf{P}}} - f_2(\bar{\mathbf{P}}, \boldsymbol{\gamma}, \boldsymbol{Y}) + \frac{\rho}{2}\sum_{b\in\mathcal{B}} \|\boldsymbol{\delta}_b - \bar{\boldsymbol{p}}_b + \boldsymbol{z}_b\|_2^2. \tag{3.39}$$

The closed-form solution to the problem in (3.39) is derived by setting $\frac{\partial \mathcal{L}_D(\bar{\mathbf{P}},\boldsymbol{\Delta},\boldsymbol{Z})}{\partial \bar{p}_{b,s}} = 0$.

$$\bar{p}_{b,s} = \frac{\sum\limits_{n\in\mathcal{N}} y_{b,s,n}\sqrt{a_{b,s,n}w(1+\gamma_{b,s,n})\left|h_{b,s,n}\right|^2} + \frac{\rho}{2}(\delta_{b,s}+z_{b,s})}{\sum\limits_{b'\in\mathcal{B}'}\sum\limits_{n\in\mathcal{N}} q y_{b',s,n}^2 \left|\bar{h}_{b',s,n}\right|^2 + \sum\limits_{n\in\mathcal{N}} y_{b,s,n}^2(\left|h_{b,s,n}\right|^2 + \left|\tilde{h}_{b,s,n}\right|^2) + \frac{\rho}{2}}. \tag{3.40}$$

3.5.1.2 Obtaining the second block of primal variables, $\boldsymbol{\Delta}$

In order to obtain the optimal second block of primal variables, i.e., $\boldsymbol{\Delta}$, we need to solve the following optimization problem:

$$\boldsymbol{\Delta} \leftarrow \arg\min_{\boldsymbol{\Delta}} \mathcal{L}_D(\bar{\mathbf{P}},\boldsymbol{\Delta},\boldsymbol{Z}) = \arg\min_{\boldsymbol{\Delta}} \mathbb{1}_{\mathbf{C_1}}(\boldsymbol{\Delta}) + \frac{\rho}{2}\sum_{b\in\mathcal{B}} \|\boldsymbol{\delta}_b - \bar{\boldsymbol{p}}_b + \boldsymbol{z}_b\|_2^2. \tag{3.41}$$

The optimization problem in (3.41) can be decomposed into B sub-problems that can be solved in parallel. The b-th sub-problem is given as:

$$\begin{aligned} \mathcal{P}_7: \quad & \underset{\boldsymbol{\delta}_b}{\text{minimize}} \quad -\frac{\rho}{2}\sum_{s\in\mathcal{S}} (\delta_{b,s} - \bar{p}_{b,s} + z_{b,s})^2 \\ \text{s.t.} \quad & \mathbf{C_{13}}: \sum_{s\in\mathcal{S}} \delta_{b,s}^2 \leq P_b^{\max}, \end{aligned} \tag{3.42}$$

Each sub-problem in $\mathcal{P}_7$ is a projection onto $l2$ ball (a convex problem) that can be solved by utilizing stationarity and complementary slackness conditions [88]. Therefore, the optimal solution of the second block of primal variables is obtained as:

$$\boldsymbol{\delta}_b = \boldsymbol{\xi}_b \times \min\left\{\frac{\sqrt{P_b^{\max}}}{\|\boldsymbol{\xi}_b\|_2}, 1\right\}, \ \forall b \in \mathcal{B}, \tag{3.43}$$

where $\boldsymbol{\xi}_b = [\xi_{b,1}, \ldots, \xi_{b,S}]^T = [\bar{p}_{b,1} - z_{b,1}, \ldots, \bar{p}_{b,S} - z_{b,S}]^T$. Similarly, the dual variable $\boldsymbol{Z}$ can be obtained by $\bar{\mathbf{P}} - \boldsymbol{\Delta}$. In **Algorithm-2**, the overall process of power allocation with ADMM combined with FP is summarized. First, the central processor unit (CPU) computes the auxiliary variables $\boldsymbol{Y}$, $\boldsymbol{\gamma}$ and then the first block of primal variables, i.e., $\bar{\mathbf{P}}$. Then, the CPU transmits $\bar{\mathbf{P}}$ to all BSs, and each BS obtains the optimal value of the second bock of primal variables, i.e., $\boldsymbol{\Delta}$ and sends them back to the CPU. This process continues until convergence or reaches the maximum number of iterations. It is noteworthy that there is a trade-off between the number of iterations in the ADMM, which is the inner algorithm in **Algorithm-2**, and the optimality of the solution. Therefore, for faster convergence, we need to adjust the penalty factor ρ. In **Algorithm-2**, BSs are not required to possess the CSI of every user associated with other BSs. They need only possess a copy of $\boldsymbol{\xi}_b$. If we wish to implement **Algorithm-1** in the same distributed manner, each BS must have the CSI of all users, thereby increasing the communication overhead. In addition, **Algorithm-1** is computationally more expensive when obtaining $\bar{\mathbf{P}}$ at each BS.

3.5.2 Complexity Analysis of the Proposed Algorithms

In what follows, we provide a worst-case complexity analysis of the proposed algorithms.

3.5.2.1 Algorithm-1

The complexity of Algorithm 1 is calculated as follows. Step-2 and Step-10 of **Algorithm-1**, each has the complexity of $\mathcal{O}(BSN)$. The bisection search in Step-7 requires $\mathcal{O}(B \log(\frac{\mu_{\max}}{\epsilon_b}))$, where $\mu_{\max}$ is the upper value of the Lagrange multiplier,

μ_b, and ϵ_b is the error tolerance of bisection search. Step-7 also has $\mathcal{O}(BS)$ for calculating the power allocation. Therefore, the overall complexity of **Algorithm-1** is $\mathcal{O}(I_P(2BSN + BS + B\log(\frac{\mu_{\max}}{\epsilon_b})))$, such that I_P is the number of iterations

Algorithm 2 ADMM-based distributed power allocation

Input: Initializing $\bar{\mathbf{P}}$, $\boldsymbol{\gamma}$ with feasible values, maximum number of iterations $L_{\max}$ for FP, and $L^A_{\max}$ for ADMM, stopping accuracy ϵ_1 for FP and ϵ_A for ADMM, and ρ.

for $t = 0, 1, ...$ **do**
  Update $\boldsymbol{Y}^{(t+1)}$ using (3.32)
  Initialize $\boldsymbol{\Delta}, \boldsymbol{Z}$
  for $k = 0, 1, \ldots$ **do**
    Update $\bar{\mathbf{P}}^{(k+1)}$ using (3.40) and send it to all BSs
    Each BS updates $\boldsymbol{\delta}_b$ using (3.43) and sends it back to the CPU
    Update $\boldsymbol{Z}^{(k+1)}$ using $\boldsymbol{Z}^{(k+1)} \leftarrow \boldsymbol{Z}^{(k)} + \bar{\mathbf{P}}^{(k)} - \boldsymbol{\Delta}^{(t)}$
    Until $\frac{\|\bar{\mathbf{P}} - \boldsymbol{\Delta}\|_2}{\|\boldsymbol{\Delta}\|_2} < \epsilon_A$ or $k = L^A_{\max}$
  end for
  Update $\boldsymbol{\gamma}^{(t+1)}$ with $\boldsymbol{\Delta}^{(k+1)}$ using (3.29)
  Until $|f(\bar{\mathbf{P}}, \boldsymbol{Y}, \boldsymbol{\gamma})^{(t+1)} - f(\bar{\mathbf{P}}, \boldsymbol{Y}, \boldsymbol{\gamma})^{(t)}| < \epsilon_1$ or $t = L_{\max}$
end for

Output: The optimal power allocation: $\mathbf{P}^* = (\bar{\mathbf{P}}^{(t+1)})^2 = (\boldsymbol{\Delta}^{(t+1)})^2$.

Algorithm 3 Overall two-stage algorithm for solving $\mathcal{P}_1$

Input: Upper bound for fractional bandwidth, i.e., B_{th}, start and end frequencies of the TW of interest, f_I and f_E, tolerance value of the approximated window, ϵ.

Stage 1:

Obtain the feasible values of $\hat{w}_I$ and $\hat{w}_E$ using interior-point method.
Obtain S_{LB} using (3.18)

Stage 2:

Initialize a feasible $\mathbf{A}^{(0)}$ heuristically. Using $\mathbf{A}^{(0)}$, initialize $\mathbf{P}$ with equal power allocation and a scale factor between 0 and 1.
for $i = 0, 1, ...$ **do**
  Solve $\mathcal{P}_3$ using the dual-simplex method and obtain $\mathbf{A}^{(i)}$.
  Solve power allocation sub-problem using either **Algorithm-1** or **Algorithm-2** and obtain $\mathbf{P}^{(i)}$
  Until $|f(\mathbf{P}, \mathbf{A})^{(i+1)} - f(\mathbf{P}, \mathbf{A})^{(i)}| < \epsilon_3$
end for

Output: The optimal power allocation, $\mathbf{P}^*$, Joint user association and sub-band assignment, $\mathbf{A}^*$, Number of sub-bands $\hat{S}$, and starting and end edge bands, $\hat{w}_I, \hat{w}_E$.

Algorithm-1 requires to converge based on ϵ_1.

3.5.2.2 Algorithm-2

For the inner algorithm (i.e., ADMM), Step-5 and Step-6 require the complexity of $\mathcal{O}(BS)$ together. Similar to **Algorithm-1**, the complexity of Step-2 and Step-10 is $\mathcal{O}(BSN)$. Hence, the overall complexity of Algorithm-2 is obtained as $\mathcal{O}(I_{FP}(2BSN+I_{AD}2BS))$, where I_{FP} and I_{AD} is the number of iterations required for convergence in the outer and inner ADMM algorithm, respectively. Based on our experiments, we observed that by selecting proper penalty factor for ADMM, i.e., ρ, the ADMM algorithm can converge in a few iterations. Therefore, the overall complexity of **Algorithm-2** is less than that of centralized power allocation in Algorithm 1.

3.5.2.3 Algorithm-3

The complexity of Stage 1 in **Algorithm-3** is $\mathcal{O}(8I_I)$, such that I_I is the number of iterations for the interior-point method. In Stage 2, the order of complexity in the joint user association and sub-band assignment when solving LP in $\mathcal{P}_3$ is $\mathcal{O}((BSN)^3)$. Solving $\mathcal{P}_3$ without using Theorem 1 results in the complexity of order $\mathcal{O}(2^{BSN})$. As a result, the overall complexity of solving $\mathcal{P}_1$ with the solution presented in **Algorithm-3** is $\mathcal{O}(I_M(8I_I+(BSN)^3+I_P(2BSN+BS+B\log(\frac{\mu_{\max}}{\epsilon_b}))))$ if **Algorithm-1** is used for power allocation and is $\mathcal{O}(I_M(8I_I+(BSN)^3+I_{FP}(2BSN+I_{AD}2BS)))$ if **Algorithm-2** is utilized for solving $\mathcal{P}_4$, where I_M is the total number of iterations required for **Algorithm-3** to converge. The proof of convergence of the alternating optimization employed in the second stage of **Algorithm-3** to a sub-optimal point is provided in Section 3.7.2.

3.6 Numerical Results and Discussions

This section focuses on evaluating the performance of the proposed algorithms and presenting their convergence behaviors. The results extract useful insights related to the impact of the number of BSs, blockages, molecular absorption, antennas' directionality, hardware impairment, multi-antenna BSs, and imperfect CSI on the performance.

3.6.1 Simulation Set-up and Parameters

Unless otherwise stated, we consider that the users and BSs are randomly distributed across a circular region with a radius of 30 [m], the maximum transmit power is set to be $P_b^{\max} = 1$ Watts, the shape parameter of Nakagami-m fading is $m = 20$, fractional bandwidth is $B_{\text{th}} = 1\%$, array gains are $G_{\max}^{\text{Tx}} = G_{\max}^{\text{Rx}} = 25$ [dB]. The coverage of each BS is determined by transmit power and antenna gains. The blockage density is $\eta = 0.005$ [m^{-1}], the penalty factor of ADMM is $\rho = 2.2$, and the accuracy of all three algorithms is set to be $\{\epsilon_1, \epsilon_2, \epsilon_3\} = 10^{-3}$. In this section, the considered gas molecules used to calculate the molecular absorption coefficient and their respective ratios are as follows: N2 (Nitrogen): 76.545% - O2 (Oxygen): 20.946 % - H2O (Water): 1.57% - CO2 (Carbon dioxide): 0.033 % - CH4 (Methane): 0.906%

In the figures of this section, *'Algo3 - Cent. PA'* and *'Algo3 - Dist. PA'* represent the use of Algorithm-3 with **Algorithm-1** and **Algorithm-2**, respectively. *'Algo3 - Cent. PA - SC'* stands for the case of using single connectivity, in which we constrain the users to associate with only one BS. *'Eq. PA + Opt. UASA'* denotes the use of equal power allocation and the proposed joint user association and sub-band

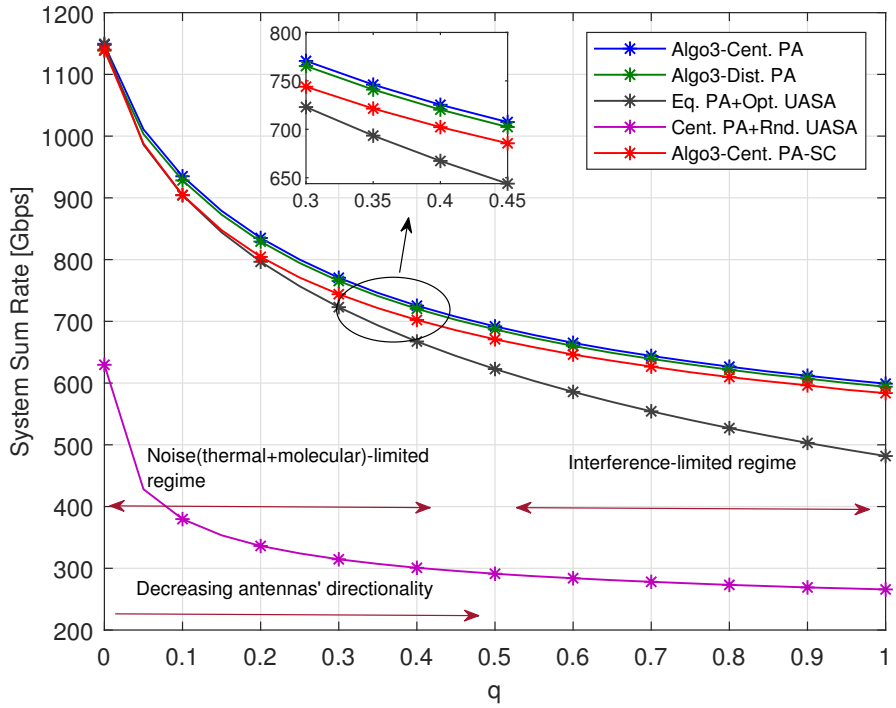


Figure 3.3: Varying antennas' directionality on network sum-rate. $\Gamma_n^L = 1$, $B = 6$, $N = 12$. *Adapted from [76]. © 2024 IEEE.*

assignment, and *'Cent. PA + Rnd. UASA'* shows the use of proposed power allocation with **Algorithm-1** along with a feasible random user association and sub-band assignment. Moreover, *'Average order of multi-connectivity'* (AOM) is obtained by counting the number of BSs to which each user is associated and then averaging out over all users.

3.6.2 Impact of Antennas' Directionality

Figs. 3.3 and 3.4 depict the impact of altering the antennas' directionality q on the network sum-rate and AOM. From Fig. 3.3, we note that the centralized and distributed versions of **Algorithm-3** outperform *'Eq. PA + Opt. UASA'* and *'Cent. PA + Rnd.*

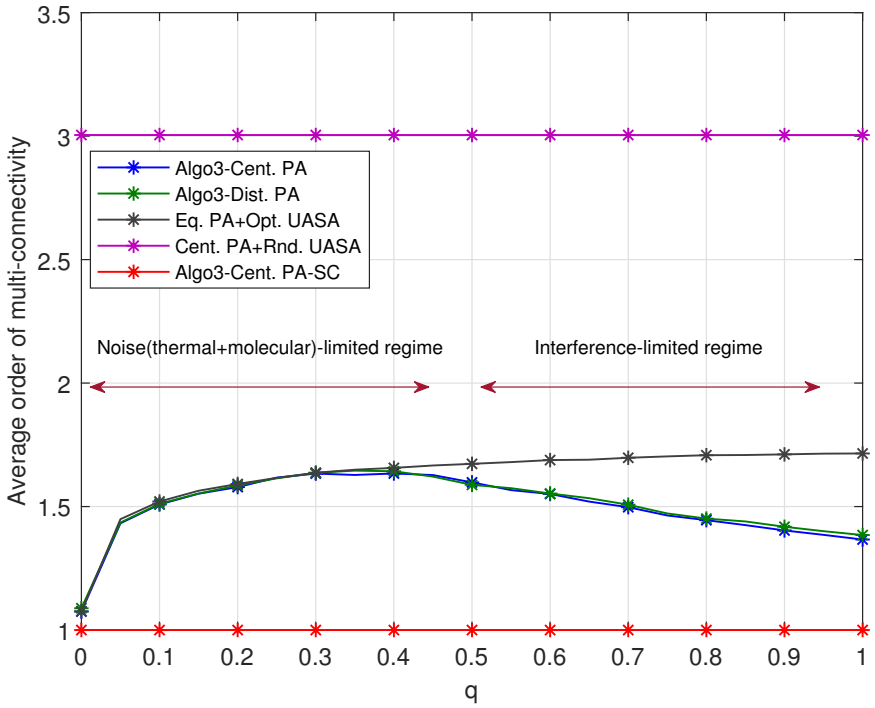


Figure 3.4: Varying antennas' directionality on average multi-connectivity order. $\Gamma_n^L = 1$, $B = 6$, $N = 12$. *Adapted from [76]. © 2024 IEEE.*

UASA', especially for scenarios when q is high, i.e., low directionality resulting in high interference; thereby interference-limited regime. Subsequently, increasing q (lowering antennas' directionality) degrades the performance of all methods. Moreover, it can be observed that multi-connectivity improves the system sum-rate, especially for high directionality. Fig. 3.4 shows that in the noise-limited regime, AOM increases as q increases. However, in the interference-limited regime, as opposed to *'Eq. PA + Opt. UASA'* the AOM of the proposed methods decreases due to optimal allocations decrease the impact of interference on the system sum-rate.

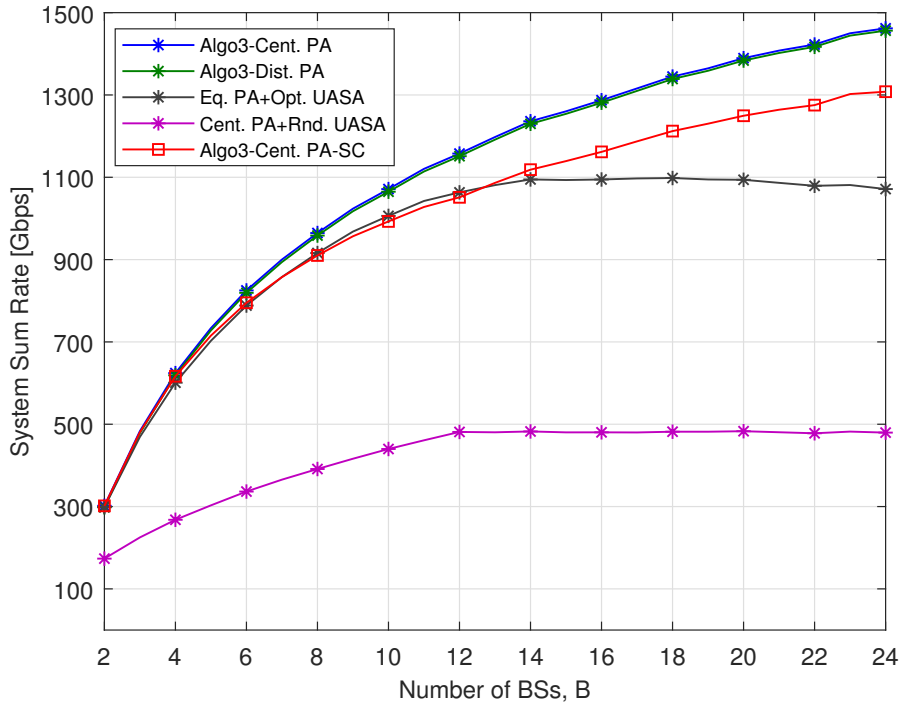


Figure 3.5: System sum-rate vs number of BSs, $\Gamma_n^L = 1$, $q = 0.2$, $N = 12$. *Adapted from [76]. © 2024 IEEE.*

3.6.3 Impact of the Number of BSs

Figs. 3.5 and 3.6 illustrate the influence of increasing the number of BSs on the system performance for both proposed methods and other benchmarks. From Fig. 3.5, it is evident that as the number of BSs increases, both *'Algo3 - Cent. PA'* and *'Algo3 - Dist. PA'* outperform *'Eq. PA - Opt. UASA'* and *'Cent. PA + Rnd. UASA'*. Also, it is evident that the proposed single-connectivity case shows an inferior performance compared to the proposed methods with multi-connectivity, and the gap increases with the number of BSs. The improvement in the sum-rate achieved by random user association and sub-band assignment reaches a saturation point, highlighting the superior performance of the proposed low-complexity joint user association and

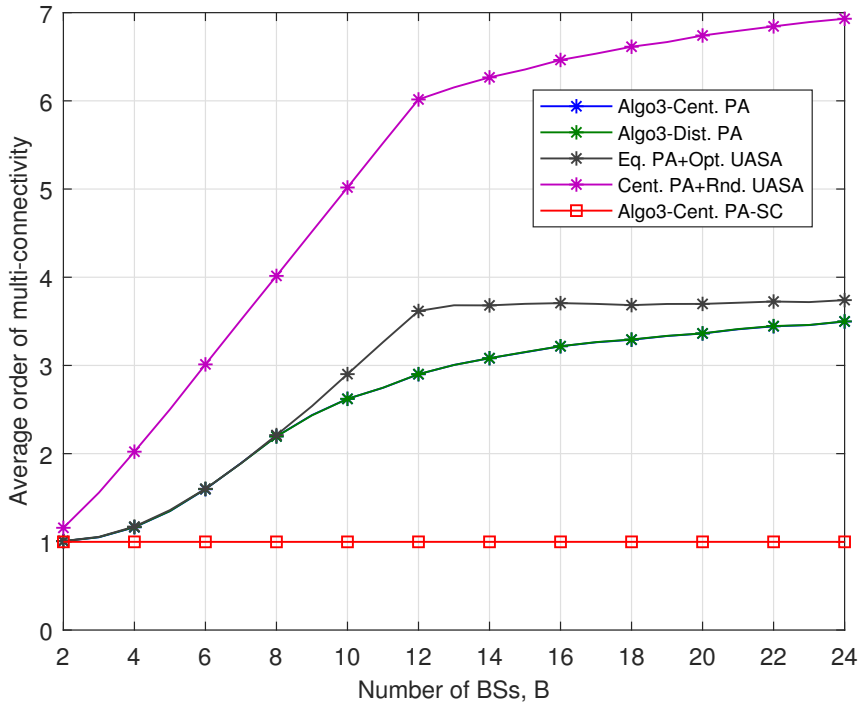


Figure 3.6: Average multi-connectivity order vs number of BSs, $\Gamma_n^L = 1$, $q = 0.2$, $N = 12$. *Adapted from [76]. © 2024 IEEE.*

sub-band assignment. Moreover, the gain exhibited by the proposed methods over *'Eq. PA - Opt. UASA'*, which benefits from the proposed joint user association and sub-band assignment, increases with the number of BSs. This improvement is attributed to the optimal power allocation that compensates for increased interference levels while effectively leveraging multi-connectivity. Furthermore, from Figure 3.6, we can observe that the AOM of the proposed methods increases with B, which reflects the impact of leveraging multi-connectivity in such a system. Although the increment of AOM is seen for *'Eq. PA - Opt. UASA'*, due to the lack of optimal power allocation along with the user association, the increase in AOM cannot compensate for the increased level of interference as the total number of BSs rises.

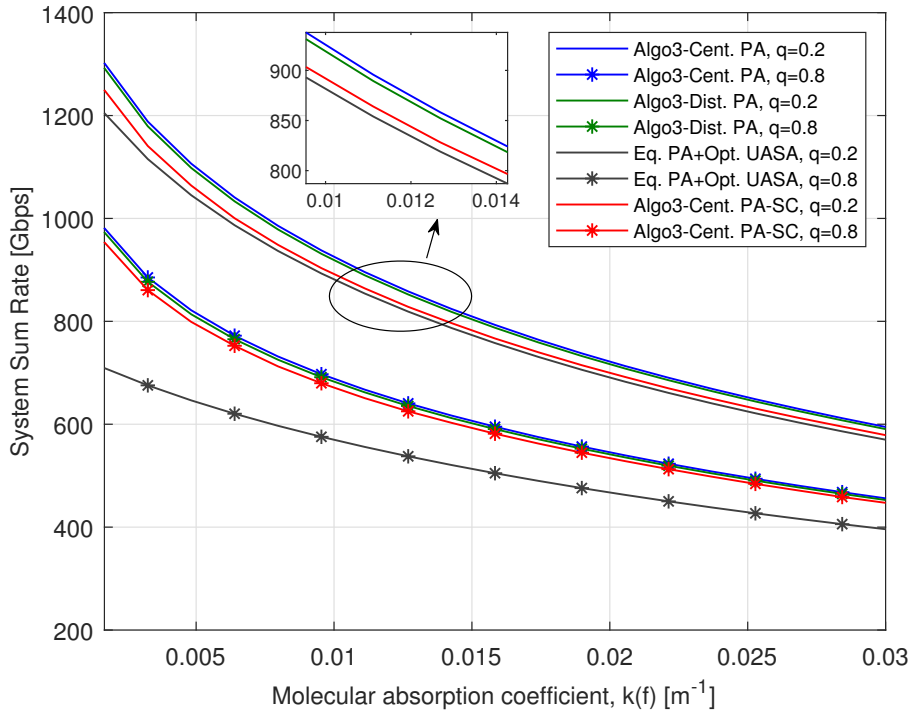


Figure 3.7: Network sum-rate vs molecular absorption coefficient, $\Gamma_n^L = 1$, $B = 6$, $N = 12$. *Adapted from [76]. © 2024 IEEE.*

3.6.4 Effect of Molecular Absorption Coefficient

Varying the molecular absorption coefficient directly, for a given carrier frequency, can reflect different environmental factors, such as pressure, temperature, altitude, and so on [72]. Fig. 3.7 highlights that the system sum-rate decreases when the molecular absorption coefficient $k(f)$ increases, which reflects different environments. The proposed methods, however, outperform the *'Eq. PA - Opt. UASA'*, especially for smaller values of $k(f)$ and reduced antennas' directionality, i.e., $q = 0.8$, as they can optimally reduce the interference for a given user. Moreover, raising $k(f)$ reduces the differences between the proposed methods and *'Eq. PA - Opt. UASA'*, even for $q = 0.8$. The reason is that since $k(f)$ impacts the channels exponentially, the channel

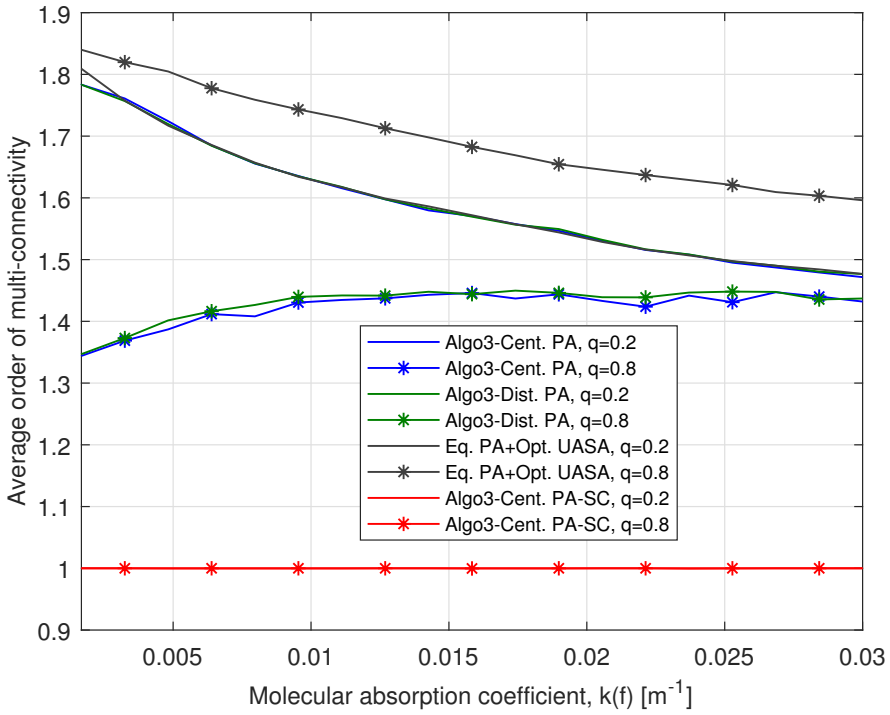


Figure 3.8: Average multi-connectivity order vs molecular absorption coefficient, $\Gamma_n^L = 1$, $B = 6$, $N = 12$. *Adapted from [76].* © *2024 IEEE.*

gain of the signal power in the numerator of the SINR for each user decreases as the molecular absorption coefficient increases. It also increases the impact of molecular absorption noise. Therefore, although still outperforming *'Eq. PA - Opt. UASA'*, the proposed algorithms' performance is impeded by channel gains when having an increased molecular absorption coefficient. Again, Fig. 3.8 depicts that the proposed algorithms strike the balance between the multi-connectivity required with *'Eq. PA - Opt. UASA'* and single-connectivity to optimize the system performance gains.

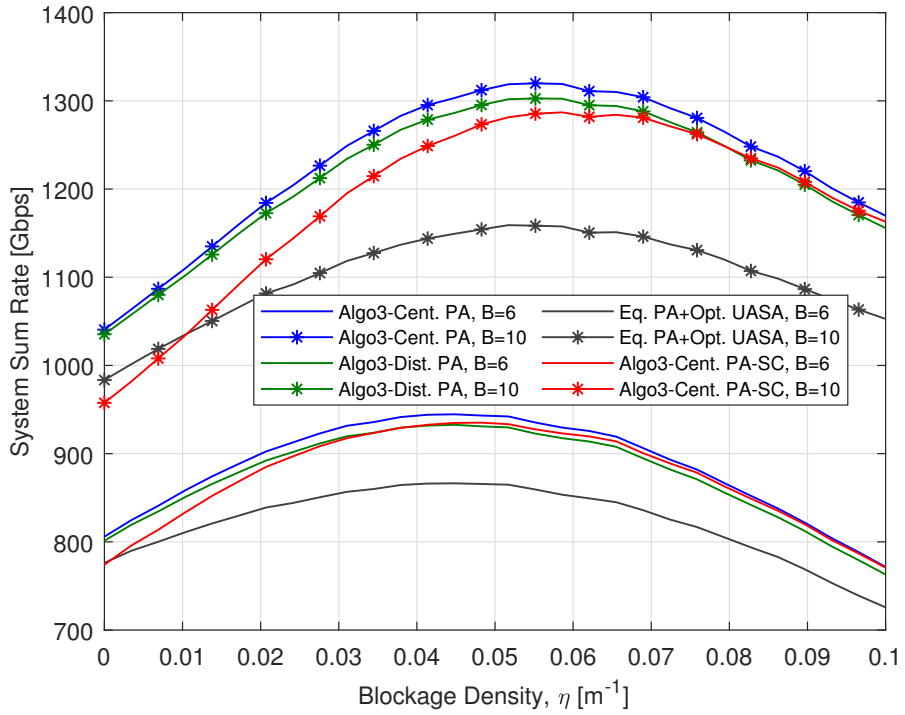


Figure 3.9: System sum-rate vs blockage density, $\Gamma_n^L = 1$, $q = 0.2$, $N = 12$. *Adapted from [76]. © 2024 IEEE.*

3.6.5 Blockage Density

Figs. 3.9 and 3.10 illustrate the impact of increasing blockage density (η) on the system sum-rate and AOM for different numbers of BSs. Observing Fig. 3.9, it becomes evident that as the blockage density increases, both proposed methods outperform *'Eq. PA - Opt. UASA'*. For lower values of η, the gap between multi-connectivity and single-connectivity widens. This is due to the availability of more unblocked links for connecting to multiple BSs. Furthermore, it is noticeable that employing a greater number of BSs enhances performance and increases the performance gap between the proposed methods and other benchmarks.

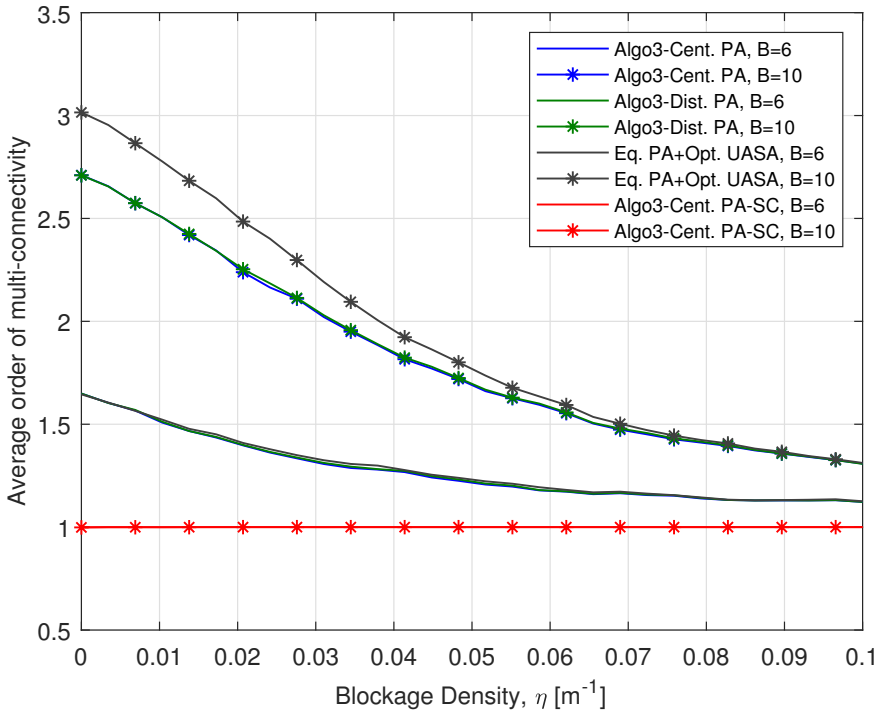


Figure 3.10: Average multi-connectivity order vs blockage density, $\Gamma_n^L = 1$, $q = 0.2$, $N = 12$. *Adapted from [76]. © 2024 IEEE.*

When blockage density increases, the proposed power allocation and user association strategies aim to maximize the system sum-rate. Consequently, more blocked links are observed at the interference of a given user. Therefore, the system sum-rate improves as blockage density rises. However, after reaching a certain blockage density, the overall data rate starts to decline. This decline is a consequence of an increase in η, which exponentially raises the likelihood of blocking the optimal link for a given user. Fig. 3.10 illustrates that the AOM for all cases decreases as η increases, i.e., in strong blockage environments, the proposed solution discourages multi-connectivity. Nonetheless, employing a greater number of BSs encourages the AOM level.

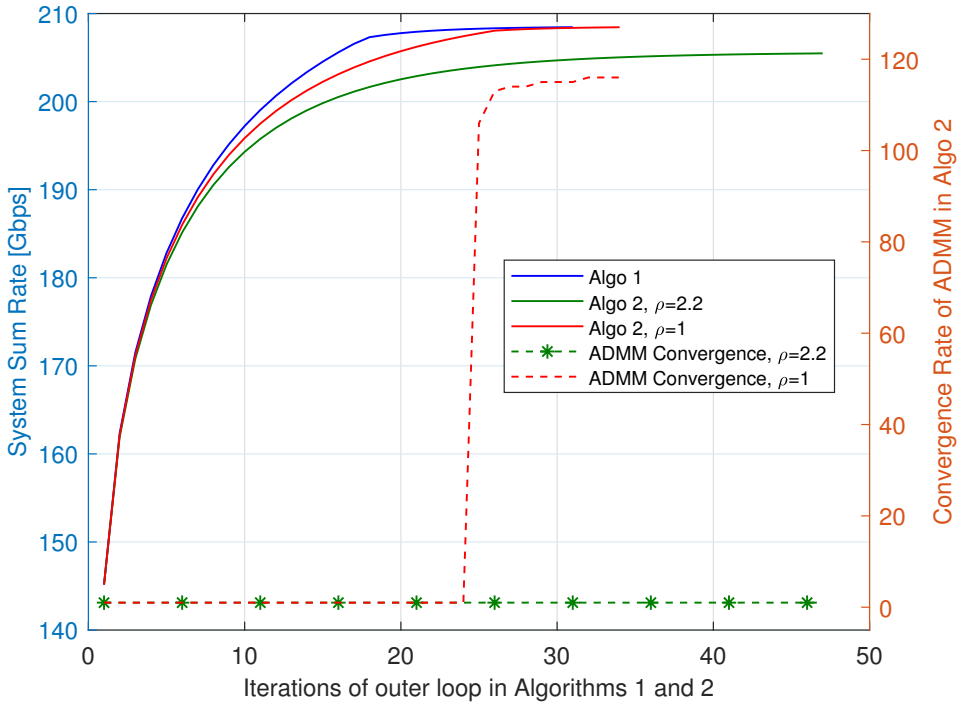


Figure 3.11: Convergence of the proposed algorithms for power allocation, and the impact of varying the penalty factor in Algorithm-2. *Adapted from [76]. © 2024 IEEE.*

3.6.6 Convergence Analysis

Fig. 3.11 illustrates the convergence of Algorithms 1 and 2, which are represented in the figure by *'Algo 1'* and *'Algo 2'*, respectively. It is evident that the proposed algorithms for power allocation converge in a few iterations. Also, the convergence of **Algorithm-2** is depicted for different penalty factors of the ADMM algorithm, i.e., ρ to reflect the impact of the ADMM algorithm convergence on the overall **Algorithm-2**. The number of iterations at which the inner ADMM algorithm converges is displayed on the right y-axis. It can be seen that by reducing ρ, **Algorithm-2** converges faster and approaches **Algorithm-1** at the expense of having a greater convergence

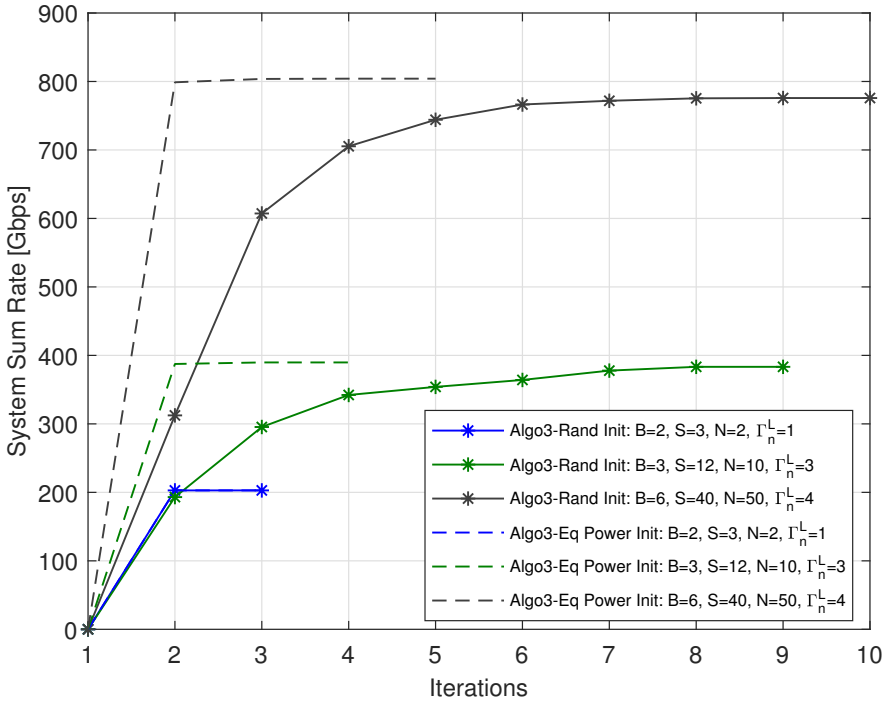


Figure 3.12: Convergence of the proposed algorithm, Algorithm-3, for different setups, minimum required number of sub-bands for each user, i.e., Γ_n^L, and different initializations. *Adapted from [76]. © 2024 IEEE.*

rate for the inner ADMM algorithm. Using $\rho = 2.2$ for the penalty factor lets the inner algorithm converge in only one iteration, which can significantly reduce the communication overhead in the system. It can be deduced that there is a trade-off between the optimality of the solution obtained by **Algorithm-2** and the overall complexity when **Algorithm-2** is utilized in a distributed manner.

Figure 3.12 shows the convergence of the overall proposed **Algorithm-3** for various system configurations, user association and sub-band assignment initialization, and the minimum required number of sub-bands for users, i.e., Γ_n^L. In this figure, *'Algo 3-Rand Init'* stands for the case where the initialization for the user association and

sub-band assignment, i.e., **A** is a random and feasible point. *'Algo 3-Eq Power Init'* is the case when the initial **A** is obtained by calculating the rates using equal power allocation. It is evident that increasing the number of users, BSs, and sub-bands increases the convergence rate of Algorithm-3. The reason is that the total number of optimization variables is increases, leading to expanding the feasible set and search space of the optimization problem that can take more time to converge. Additionally, as shown in the previous section, the computational complexity of **Algorithm-3** increases as a function of optimization variables, which in turn, requires the algorithm a greater number of iterations to converge. It can also be seen that with random initialization for user association and sub-band assignment, Algorithm-3 can achieve a comparable performance compared to the case with equal power allocation-based initialization.

3.6.7 Number of Sub-bands, Available Bandwidth and Molecular Absorption Coefficient

Fig. 3.13 demonstrates the impact of varying the tolerance of variation within each TW, i.e., ϵ, on the total available bandwidth, which is obtained as $\bar{w}_T = f_e - w_e - (f_I + w_I)$, and the optimal number of sub-bands, achieved in (3.18). For TW_1, the total available bandwidth is less than that of TW_3 since the length of the TW is smaller. Moreover, the smaller TW length of TW_1 leads to having higher edge bands compared to TW_3 for a given ϵ, leading to a reduction in the available bandwidth. Although increasing the tolerance ϵ leads to higher available bandwidth, it reduces the edge bands and therefore causes a higher molecular absorption coefficient as shown in Fig. 3.14.

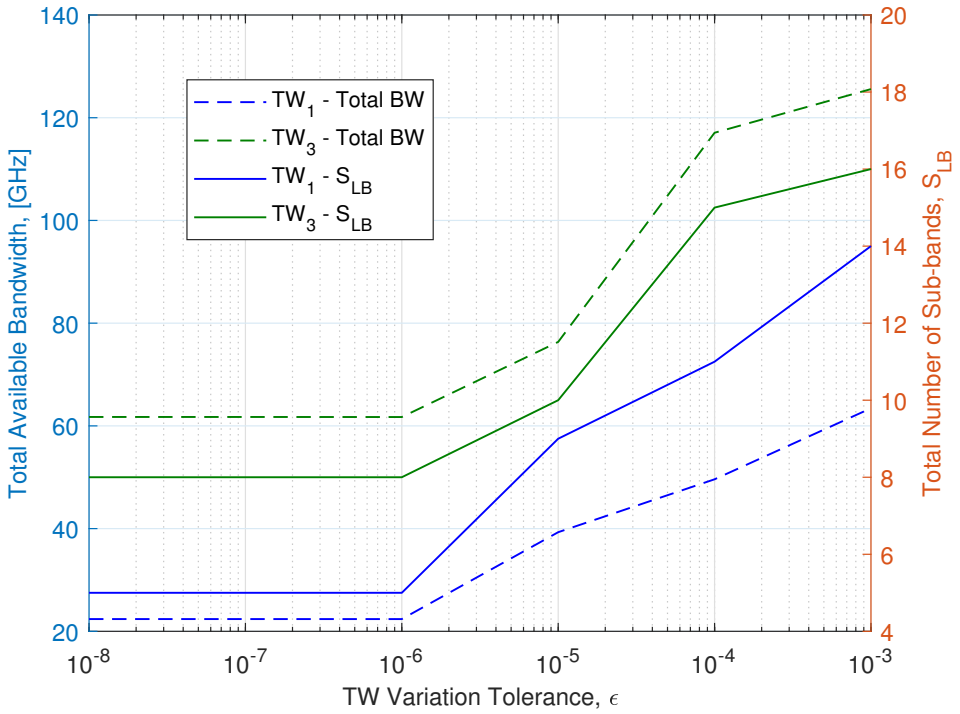


Figure 3.13: Total available bandwidth and optimal number of sub-bands for different TWs versus molecular absorption coefficient variation tolerance, ϵ. *Adapted from [76]. © 2024 IEEE.*

3.6.8 Multi-Antenna Base Stations

In this sub-section, we aim to show the performance of the proposed resource allocation methods under the scenario where each BS is equipped with a uniform linear array comprising M antenna elements. For beamforming, maximum ratio transmission (MRT) beamforming is utilized. Denoting the beamformer at b-th BSs on sub-band s by $\boldsymbol{u}_{b,s} = \boldsymbol{h}_{b,s,n} \in \mathbb{C}^{M\times 1}$ [89], the SINR at user n from BS b on sub-band

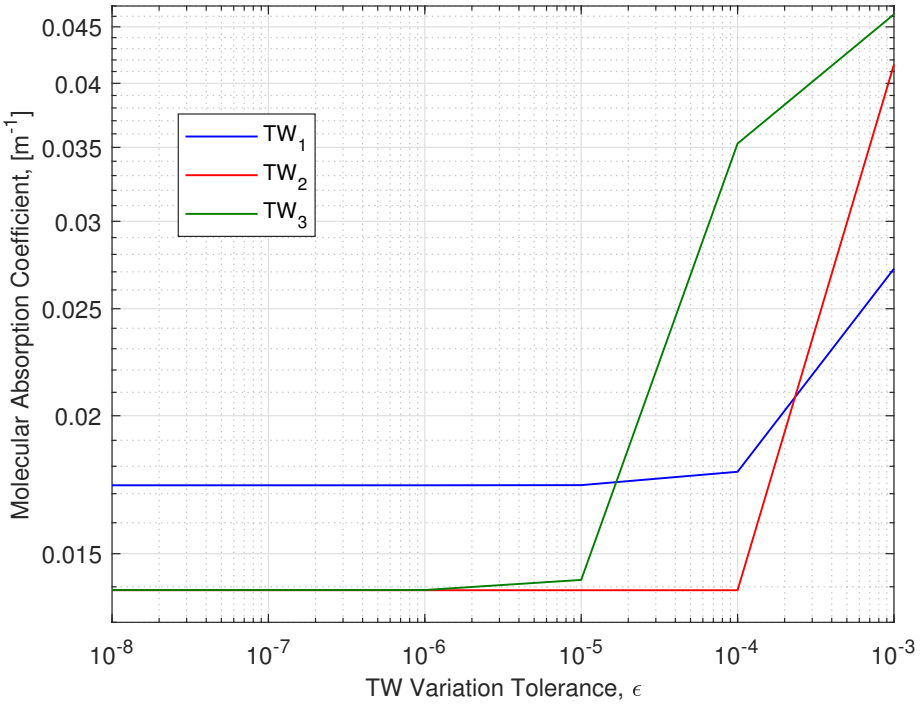


Figure 3.14: Molecular absorption coefficient $k(f)$ for different TWs versus molecular absorption coefficient variation tolerance, ϵ. *Adapted from [76]. © 2024 IEEE.*

s is given by:

$$\gamma_{b,s,n} = \frac{p_{b,s}\left|\boldsymbol{h}_{b,s,n}^{H}\boldsymbol{u}_{b,s}\right|^{2}}{\sum\limits_{b'\in\mathcal{B}'} q p_{b',s}\left|\bar{\boldsymbol{h}}_{b',s,n}^{H}\boldsymbol{u}_{b',s}\right|^{2} + p_{b,s}\left|\tilde{\boldsymbol{h}}_{b,s,n}^{H}\boldsymbol{u}_{b,s}\right|^{2} + N_0 w}, \tag{3.44}$$

where the notation $\boldsymbol{h} \in \mathbb{C}^{M\times 1}$ in boldface represents the equivalent complex vector of their single-antenna channel counterpart. Fig. 3.15 illustrates the effect of varying the number of antennas at each BS. The results indicate that increasing M enhances the proposed solutions, surpassing other methods. Moreover, as the number of antennas increases, the performance of "Algo3-Cent. PA-SC" becomes comparable to that of

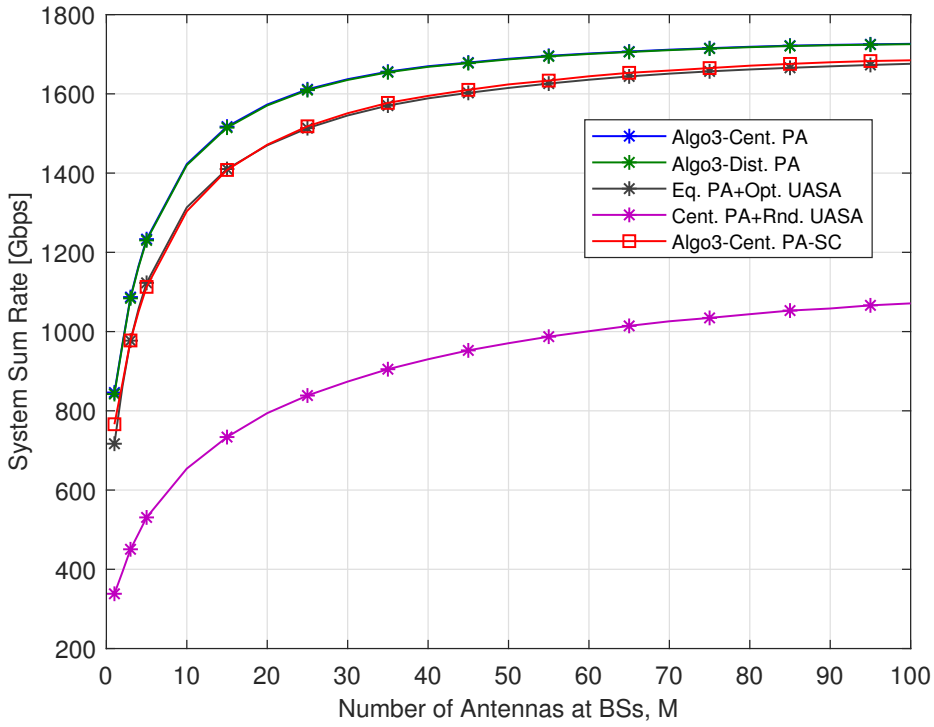


Figure 3.15: System sum-rate vs the number of antennas at BSs, $q = 0.5$, $B = 10$, $N = 12$ *Adapted from [76]. © 2024 IEEE.*

"Eq. PA+Opt. UASA," which benefits from multi-connectivity.

3.6.9 Impact of Hardware Imperfections

In this sub-section, we account for the influence of hardware imperfections (HI) in THz transceivers on the system's performance. Consequently, we incorporate the impact of HI into the SINR formulation. Distortions stemming from HI at both transmitters and receivers can be modeled as additive Gaussian noise [90, 91]. For the received signal at a user $y = h(x + n_t) + n_r$, $n_t \sim \mathcal{CN}(0, k_t^2 p)$ represents the distortion noise of the transmitter, and $n_t \sim \mathcal{CN}(0, k_r^2 |h|^2 p)$ denotes the distortion at the receiver side. Here, k_t and k_r are the HI parameters at the transmitter and receiver, respectively.

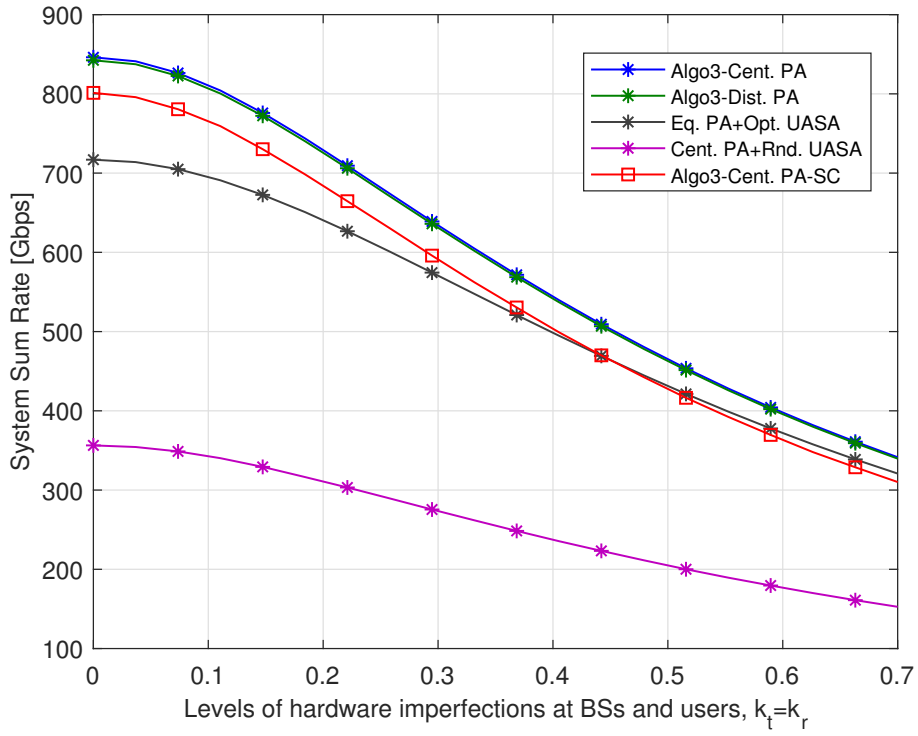


Figure 3.16: System sum-rate vs the level of HI at BSs and users ($k_t = k_r$), $q = 0.5$, $B = 10$, $N = 12$ *Adapted from [76]. © 2024 IEEE.*

Note that the variance of the distortion noise is proportional to the input power of the hardware. Employing this HI modeling, the HI-aware SINR at user n associated with BS b over sub-band s is formulated as follows:

$$\gamma_{b,s,n}^{\mathrm{HI}} = \frac{p_{b,s}|h_{b,s,n}|^2}{\bar{I}_{b,s,n}(\mathbf{P}) + K_{b,s,n}(\mathbf{P}) + p_{b,s}|\tilde{h}_{b,s,n}|^2 + N_0 w}, \tag{3.45}$$

where $\bar{I}_{b,s,n}(\mathbf{P})$ is defined in (3.7), and

$$K_{b,s,n}(\mathbf{P}) = (k_t^2 + k_r^2)p_{b,s}|h_{b,s,n}|^2 + k_r^2\Big(\sum_{b' \in \mathcal{B}'} q p_{b',s}|h_{b',s,n}|^2\Big). \tag{3.46}$$

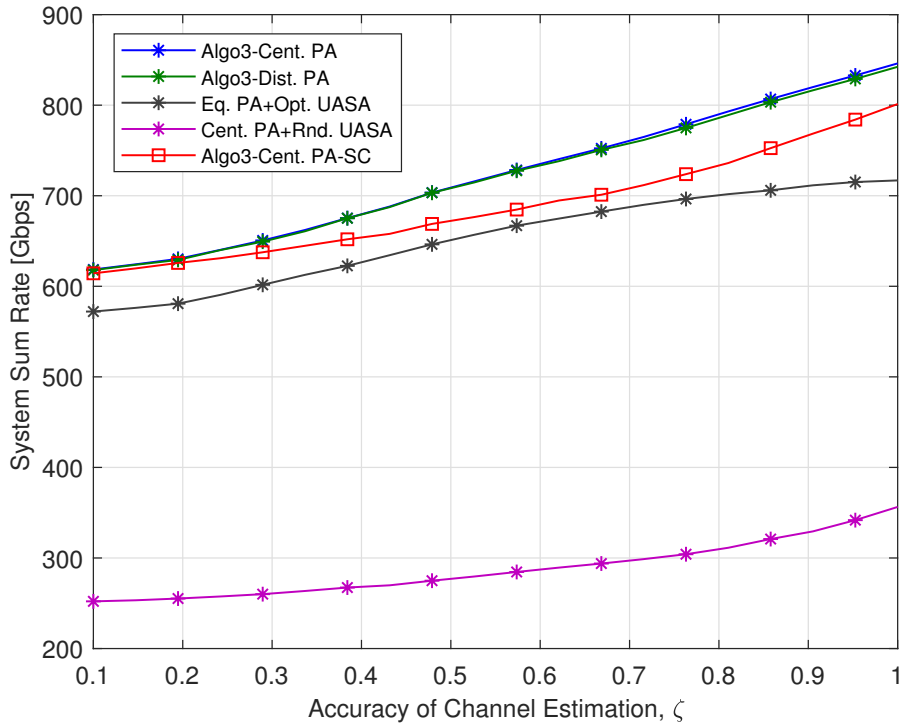


Figure 3.17: System sum-rate vs the accuracy of channel estimation (ζ), $q = 0.5$, $B = 10$, $N = 12$ *Adapted from [76]. © 2024 IEEE.*

Based on $\gamma^{\text{HI}}_{b,s,n}$, all of the formulations used in Algorithms 1,2, and 3 can be modified. The impact of different levels of hardware imperfections on the performance of the proposed solutions is depicted in Fig. 3.16. It is observed that the proposed algorithms outperform other benchmarks while exhibiting a decreasing overall performance as $k_r = k_t$ increases. Additionally, the performance of the single-connectivity case deteriorates more rapidly, approaching the case of "Eq. PA+Opt. UASA" at higher levels of HI.

3.6.10 Impact of Imperfect CSI

For the case of imperfect CSI, we consider the following model $h = \zeta\hat{h} + \sqrt{1-\zeta^2}e$,

where $\hat{h}$ is the channel estimation of h, ζ is the accuracy of channel estimation, and e is the estimation error that follows a complex Gaussian distribution with zero mean and variance of d^{-2}, such that d is the distance of the considered link [92]. Fig. 3.17 demonstrates the impact of imperfect CSI on the proposed algorithms, where increasing ζ leads to approaching perfect CSI. The results show that the performance of both the proposed solutions and other benchmarks improves with higher channel estimation accuracy. However, the improvement of "Eq. PA+Opt. UASA" is less pronounced compared to other approaches due to its lack of optimal power allocation.

3.7 Complete Proof

3.7.1 Proof of Lemma 3.3

To prove $\mathbf{T}^{(3)}$ is TUM, we utilize the following Theorem:

Theorem 3.2. *A $m \times n$ matrix A is TUM **iff** $\forall I \subseteq \{1, \ldots, m\}$, there exists a partition of I into sets F and L such that for all $j \in \{1, \ldots, n\}$, $\left|\sum_{i \in F} A_{ij} - \sum_{i \in L} A_{ij}\right| \leq 1$ holds [84].*

To apply **Theorem 3.2**, we need to show that for any subset of rows of $\mathbf{T}^{(3)}$, we can find a partitioning of the rows into two groups (G_A and G_B) where for each column, the sum of the elements in group A minus the sum of the elements in group B is either 0, 1, or -1. To do this, we first divide rows of $\mathbf{T}^{(3)}$ into four groups denoted by R_1, R_2, R_3 and R_4. The columns are also divided into B groups. In Figure 3.2, a detailed presentation of $\mathbf{T}^{(3)}$ is provided where we can see details of the first matrices in C_1.

Now, let's assume that K rows are selected, i.e. $I = \{r_1, r_2, \ldots, r_K\}$. Considering the groups of rows, we divide I into four sets I_1, I_2, I_3, and I_4, where $I_j \subseteq R_j$ and

$\sum_{j=1}^{4} |I_j| = K$. Based on this, we propose the following partitioning of the rows. All the rows in I_2 and I_3 go to group G_B, and all the rows in I_4 go to group G_A. Rows in I_1 go to either group G_A or G_B based on the following rule. Consider $r_n \in I_1$, we have:

$$\text{Group of } r_n = \begin{cases} G_B & \text{if} \quad \exists r_m \in I_2, \quad n \equiv m \pmod{S} \\ G_A & \text{otherwise} \end{cases} \tag{3.47}$$

This condition comes from two facts about $\mathbf{T}_3$. First, in each column, we have exactly 3 non-zero elements. Second, if we index the rows in R_1 from 1 to NS, and the rows in R_2 from 1 to S, we can see that for each column the rows of the first two non-zero elements belong to R_1 and R_2, and their row index have the same remainder with respect to S.

Now, we show that the *proposed partitioning follows the conditions of* **Theorem 3.2**. Consider a column from C_1, c. Since the maximum number of non-zero elements in each column is three, we'll face one of the following scenarios:

• There isn't any non-zero element in c, which means that $|\sum_{i\in G_A} \mathbf{T}^{(3)}_{i,c} - \sum_{i\in G_B} \mathbf{T}^{(3)}_{i,c}| = 0$

• There is only one 1 or -1 present in c, i.e., the sum of the elements in one of the groups (G_A or G_B) is 0 and the other one is -1 or +1, which makes their difference -1 or +1.

• There are two non-zero elements in c. Let's denote the rows corresponding to the non-zero elements by r_n, r_m $(n < m)$. Since all elements in R_3 are zero, we have three different scenarios. First, $r_n \in R_1$ and $r_m \in R_2$. In this case, they both are in group G_B, and since all other elements are zero, we have $|\sum_{i\in G_A} \mathbf{T}^{(3)}_{i,c} - \sum_{i\in G_B} \mathbf{T}^{(3)}_{i,c}| =$

$|0 - \mathbf{T}^{(3)}_{n,c} - \mathbf{T}^{(3)}_{m,c}| = |0 - (-1) - (+1)| = 0$. Second, $r_n \in R_2$ and $r_m \in R_4$. Since all elements of R_2 are in G_B, and all elements of R_4 are in G_A, we have: $|\sum_{i \in G_A} \mathbf{T}^{(3)}_{i,c} - \sum_{i \in G_B} \mathbf{T}^{(3)}_{i,c}| = |\mathbf{T}^{(3)}_{n,c} - \mathbf{T}^{(3)}_{m,c}| = |(+1) - (+1)| = 0$. Third, $r_n \in R_1$ and $r_m \in R_4$. In this case, since we don't have any non-zero element from R_2, this means that no row from R_2 is selected that has the same remainder with respect to S as n. Thus, $r_n \in G_A$. This means $|\sum_{i \in G_A} \mathbf{T}^{(3)}_{i,c} - \sum_{i \in G_B} \mathbf{T}^{(3)}_{i,c}| = |\mathbf{T}^{(3)}_{n,c} + \mathbf{T}^{(3)}_{m,c} - 0| = |(-1) + (+1) - 0| = 0$.

• There are three non-zero elements in c. Let's denote the rows with r_n, r_m, and r_k, where $n < m < k$. Based on the structure of $\mathbf{T}^{(3)}$, we know that $r_n \in R_1$, $r_m \in R_2$, and $r_k \in R_4$. Thus, $r_n, r_m \in G_B$ and $r_k \in G_A$. Thus, we have $|\sum_{i \in G_A} \mathbf{T}^{(3)}_{i,c} - \sum_{i \in G_B} \mathbf{T}^{(3)}_{i,c}| = |\mathbf{T}^{(3)}_{k,c} - \mathbf{T}^{(3)}_{n,c} - \mathbf{T}^{(3)}_{m,c}| = |(+1) - (-1) - (+1)| = +1$.

Therefore, the proposed partitioning follows the requirements of **Theorem 3.2** in C_1. In a similar manner, one can show that it's also true for other columns $c \in C_k$, where $2 \leq k \leq \mathrm{B}$. Since each column has at most three non-zero elements, we have four similar scenarios as discussed above.

Subsequently, for any column c of $\mathbf{T}^{(3)}$, i.e. $c \in C_k$, where $k \in \{1, 2, \ldots \mathrm{B}\}$, the proposed partitioning of the selected rows into G_A and G_B follows the requirement of Theorem 3.2. This means that $\mathbf{T}^{(3)}$ is TUM, implying that $\mathbf{T}$ is TUM; thus, proving Lemma **??**.

3.7.2 Proof of Convergence for Algorithm-3

For the first stage, based on **Lemma 1**, we showed that the lower bound on S is optimal. In step 3 of the second stage at the i-th iteration, for fixed power allocation $\mathbf{P}^{(i-1)}$, the globally optimal values of user association and sub-band assignment variables, $\mathbf{A}^{(i)}$, are obtained due to the convexity of the transformed linear optimization

problem in $\mathcal{P}_3$.

Then, for fixed $\mathbf{A}^{(i)}$, we need to show that solving $\mathcal{P}_4$ converges to a stationary point for $\mathbf{P}^{(i)}$. Denoting the objective function of $\mathcal{P}_4$ at the t-th iteration of either Algorithm 1 or 2 by $f(\mathbf{P}^{(t)})$, for fixed $\mathbf{A}^{(i)}$, we have:

$$
\begin{aligned}
f(\mathbf{P}^{(t)}) &= f_1(\mathbf{P}^{(t)}, \boldsymbol{\gamma}^{(t)}) && (3.48)\\
&\geq f_1(\mathbf{P}^{(t)}, \boldsymbol{\gamma}^{(t-1)}) && (3.49)\\
&= f_2(\mathbf{P}^{(t)}, \boldsymbol{\gamma}^{(t-1)}, \bar{\boldsymbol{Y}}^{(t-1)}) && (3.50)\\
&\geq f_2(\mathbf{P}^{(t)}, \boldsymbol{\gamma}^{(t-1)}, \boldsymbol{Y}^{(t-1)}) && (3.51)\\
&\geq f_2(\mathbf{P}^{(t-1)}, \boldsymbol{\gamma}^{(t-1)}, \boldsymbol{Y}^{(t-1)}) && (3.52)\\
&= f_1(\mathbf{P}^{(t-1)}, \boldsymbol{\gamma}^{(t-1)}) && (3.53)\\
&= f(\mathbf{P}^{(t-1)}) && (3.54)
\end{aligned}
$$

where (3.48) is due to the fact that by substituting optimal $\boldsymbol{\gamma}$ in f_1, the original objective function can be obtained. (3.49) holds since updating γ maximizes f_1 when other variables are fixed. (3.50) follows **Remark 1** such that $\bar{\boldsymbol{Y}}$ is obtained using equation (3.32). (3.51) holds since updating $\boldsymbol{Y}$ maximizes f_2 while other variables are fixed. (3.52) follows the fact that updating $\mathbf{P}$ maximizes f_2 when other variables are fixed. (3.53) and (3.54) follows the similar reasoning as (3.50) and (3.48), respectively. Therefore, f is monotonically non-decreasing after each iteration. Moreover, since the objective function, f is bounded due to the constraints, Algorithms 1 and 2 converge. At the convergence, a local optimal point of the reformulated objective function f_2 is obtained. The solution is a stationary point of the objective function f in $\mathcal{P}_4$ [66]. Therefore, since the original objective function is non-decreasing after

each sub-problem, the alternating optimization in stage 2 of Algorithm 3 converges to a sub-optimal solution.

3.8 Summary

This chapter considered molecular absorption-aware and blockage-aware resource allocation within a multi-cell THz system, considering multi-connectivity to maximize the system's total sum-rate. To make the problem tractable, we first proposed a convex approximation of the molecular absorption coefficient and derived a closed-form lower bound for the optimal number of sub-bands with consideration to beam-squint effects. The optimization problem is then decomposed into the power allocation, and the joint user association and sub-band assignment sub-problems. Leveraging the concepts from unimodular matrices, various low-complexity centralized and distributed solutions are presented. Numerical findings underscore the proposed solutions' superiority compared to traditional benchmarks, as well as the advantages of multi-connectivity compared to single-connectivity. Also, the impact of varying blockage density along with the antennas' directionality, hardware impairment, and imperfect CSI are explored. Moreover, the proposed framework can be extended to include explicit minimum rate requirements for users by using SCA, and FP, as in [93, 94], to enhance performance from the users' perspective.

Copyright and Co-authorship Statement

This chapter has resulted in the following publication: M. A. Saeidi, H. Tabassum, and M. Alizadeh, "Molecular Absorption-Aware User Assignment, Spectrum, and Power Allocation in Dense THz Networks With Multi-Connectivity," *IEEE Transactions*

on Wireless Communications, vol. 23, no. 11, pp. 16404–16420, Nov. 2024, doi: 10.1109/TWC.2024.3440888 [76]. © 2024 IEEE.

I was the primary author of this work and was responsible for the development of the main idea, problem formulation, system modeling, mathematical derivations, algorithm development, simulations, numerical analysis, and manuscript preparation. Author 2 supervised the work and refined the manuscript. Author 3 contributed to the proof of Lemma 3.3.

Chapter 4

Characterization of Multi-band Networks

4.1 Overview

To date, wireless technology has been developed primarily over the RF, sub-6 GHz, and mmWave bands. However, these bands alone cannot keep up with the massive connectivity, high data rate, low latency, and broad coverage requirements expected from next-generation wireless networks. Extremely high frequency bands, such as optical and THz, can provide much wider transmission bandwidths and very high data rates. Nevertheless, these bands also suffer from unique propagation limitations, including higher spreading loss, atmospheric absorption, diffuse scattering, weather sensitivity, and smaller coverage regions. Therefore, there is no one-size-fits-all spectrum solution for 6G networks.

In this context, multi-band networks (MBNs) are expected to play an important role by allowing different frequency bands to coexist and complement each other. While lower frequency bands can provide broader coverage and more stable connectivity, higher frequency bands can provide much larger bandwidths and highly directional transmissions. Therefore, an MBN can opportunistically exploit the most

suitable frequency band according to the service requirement, user mobility, propagation condition, and deployment scenario. Recent industrial developments also support this direction. For example, an over-the-air multi-band demonstration showcased sub-6 GHz and mmWave end-to-end 5G NR systems using Ericsson's 5G NR pre-commercial BSs and Qualcomm Technologies' 5G NR UE prototypes [95]. In addition, Huawei developed SingleRAN Pro as a 5G-oriented all-in-one site solution that can support sub-6 GHz and mmWave NR [96].

MBNs can be classified into two fundamental architectures:

- *Stand-alone MBN (SA-MBN):* BSs operating at different frequency bands are deployed separately, and each BS supports only one spectrum band at a time. This architecture provides flexibility in deploying different BS densities for different bands in order to compensate for their propagation characteristics.

- *Integrated MBN (Int-MBN):* Each BS can operate over more than one frequency band. Inspired by multi-homing users that can receive data over different bands simultaneously [97], multi-band BSs can enable simultaneous transmission and reception across multiple frequency bands.

The aforementioned MBNs can be realized with multi-homing user devices or stand-alone user devices, as illustrated in Fig. 4.1. Another possible MBN architecture is a cascade configuration, where RF and EHF links are used for access and backhaul transmissions, respectively.

The objective of this chapter is to characterize how different frequency bands can coexist, what challenges arise from this coexistence, and how SA-MBN and Int-MBN architectures compare under practical performance and deployment metrics.

The main contributions of this chapter can be summarized as follows:

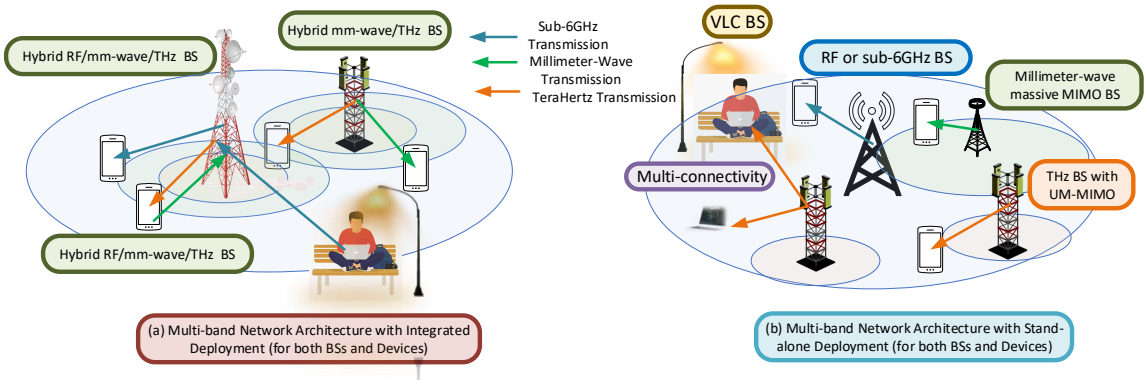


Figure 4.1: Vision of (a) Int-MBN architecture and (b) SA-MBN architecture in next generation wireless networks. *Adapted from [98]. © 2024 IEEE.*

• We discuss the main design challenges of MBNs, including multi-band transceiver design, multi-band antenna systems, channel estimation and modeling, deployment and planning, user offloading, mobility management, resource allocation, and back-hauling.

• We propose a molecular-absorption-aware user offloading metric for RF/THz MBNs and compare it with standard association rules such as maximum data rate, RSRP, and SINR-based association.

• We compare SA-MBN and Int-MBN architectures through case studies involving deployment density, THz association probability, spectral efficiency, achievable rate, deployment cost efficiency, HO probability, outage probability, and average network data rate.

The rest of this chapter is organized as follows. Section 4.2 discusses the main implementation and design challenges in MBNs. Section 4.3 presents potential resource allocation objectives and constraints. Section 4.4 studies molecular-absorption-aware user offloading in an RF/THz MBN. Section 4.5 compares SA-MBN and Int-MBN

architectures through several performance metrics. Finally, Section 4.6 summarizes the main findings of the chapter.

4.2 Fundamental Challenges in MBNs

In this section, we highlight the fundamental MBN challenges from the perspective of multi-band transceiver and antenna design as well as PHY and MAC layer design issues.

4.2.1 Multi-band Transceiver Design

Designing multi-band transceivers is crucial in Int-MBNs. There are various methods that can be used for transceiver design, and each approach has its own set of challenges.

4.2.1.1 Electronic up-conversion

EHF signals are typically produced by using frequency multiplication in cascade and up-conversion techniques [99]. Additionally, in a multi-band transceiver, frequency splitters can be applied to generate an intermediate frequency as one of the desired output frequencies. However, operating at EHF bands can cause the CMOS technology components to exceed their optimal operating range, which limits the output power.

4.2.1.2 Optical down-conversion

Multi-band transceivers can be realized by down-converting photonic signals to produce EHF signals. For instance, a Mach-Zehnder modulator generates signals at lower

frequencies, while also maintaining the input photonic signals as the output. While the design of photonic-based transceivers is relatively simpler, they are sensitive to environmental conditions, such as temperature and humidity, which can affect their performance.

4.2.1.3 Opportunistic frequency switching for Int-MBN

A BS equipped with separate EHF and RF transceivers can employ both hard and soft switching to opportunistically exploit both links. Hard switching refers to the process of switching frequencies using mechanical or electronic switches. This method is simple to implement, but lead to high power consumption and signal distortion caused by the switch transients. Soft switching, on the other hand, uses a combination of analog and digital signal processing techniques. This method can result in lower power consumption and less signal distortion, but it can also be more complex to implement and may have higher latency.

Hard switching reduces the lifetime of the devices [100]; whereas soft switching can outperform hard switching by providing lower outage probability. Note that multi-band transceivers and switching mechanisms are critical in the Int-MBN deployment; whereas SA-MBN can operate with the users and BSs transmitting or receiving on a single predefined frequency. The synchronization of time and frequency is also a challenge for successful decoding of messages transmitted on multi-band streams.

4.2.2 Multi-band Antenna Systems

To enable high gain directional antennas, EHF allows to build dense arrays of antennas. There are several approaches to realize the arrays of tunable multi-band

antennas as discussed below.

4.2.2.1 Physically adjustable antennas

based on micro or nano-electromechanical systems (MEMS/NEMS) can physically change the size of the radiating elements and achieve different resonant frequencies [101]. The main challenge is to overcome the delay introduced by the control of the MEMS systems, particularly for fast data rates.

4.2.2.2 Electronically tunable antennas

Another approach to building multi-band antenna systems is by using graphene-based 2D materials to construct software-defined nano-antenna arrays. The resonant frequency of these antennas can be adjusted by varying the graphene Fermi energy without any physical modifications. However, graphene is primarily useful for the THz band, and other materials are needed to cover lower frequency bands. It is challenging to design a material that can operate at all frequencies.

4.2.2.3 Distributed multi-band antenna systems

To overcome the limited transmit power of existing EHF transceivers and antenna systems operating at EHF bands, one solution is to leverage distributed antenna system. All BSs can cooperate to transmit the same data to a user with reduced transmission powers. At the same time, the spatial diversity increases, leading to achieving higher data rates. However, the communication overhead and control signals increases compared to centralized systems. Therefore, faster, scalable, and robust algorithms are required to enable the use of cooperative distributed multi-band antenna systems.

4.2.3 Channel Estimation and Modeling

Since the channel state information (CSI) must be accessible at various nodes for resource management, faster, robust, and unique channel estimation solutions are critical that can capture the unique channel characteristics of each frequency band. Also, the channel coherence time (determined from the Doppler frequency shift) varies significantly as a function of users' velocity and carrier frequencies. Thus, the channel estimation solutions should be faster to cope up with the reduce channel coherence time at EHF. In this context, deep learning can be applied for faster channel estimation and resource management as compared to traditional computationally exhaustive solutions.

The absence of precise channel models for EHF bands in various environments is also a fundamental challenge. The reason is that, when frequency increases, signals become more prone to atmospheric absorption losses caused by various gas molecules. Every band in the EHF spectrum has a transmission window (TW) where the molecular absorption is minimal. To accurately analyze and optimize the performance, unique channel models are required for different TWs [102]. Curve-fitting techniques can be used to create precise models of absorption loss for each TW.

4.2.4 Network Deployment and Planning

Optimal network deployment and planning is critical due to additional degrees of freedom offered by various transmission frequencies. For instance, increasing the number of RF BSs increases interference; whereas, EHF needs more BSs to overcome propagation losses. Along another note, SA-MBN enables deployment flexibility with

higher capital, operational, and backhauling expenses, whereas Int-MBN requires sophisticated multi-band transceivers. Based on the existing reports, one can postulate that the deployment and operational cost of hybrid BSs in Int-MBN is greater than that of SA-MBN [96]. To be precise, using a BS with three sectors, each enabled with a separate frequency band, and utilizing a switching mechanism to exploit all three bands can be more expensive than a BS with just one sector that uses a single frequency band. Thus, it is critical to investigate the trade-off between the deployment cost and performance of different MBN architectures.

4.2.5 User Offloading and Mobility Management

Since multiple frequency bands experience unique propagation characteristics and transmission bandwidths, user offloading methods need to adapt accordingly. Also, EHF transmissions are prone to intermittent connectivity caused by narrower beams, beam misalignment, users' mobility, and small coverage zones, resulting in more handoffs and disconnections. However, as EHF bands can deliver higher data rates, all of the required data can be transferred almost spontaneously as soon as a connection is established. In this context, MBNs are also crucial to opportunistically exploit more stable RF/mmWave networks to maintain the desired QoS of moving users. New traffic offloading mechanisms in MBNs are required that takes into account molecular absorption, beam misalignment and users' mobility. In addition, the ability of a user to connect to two BSs in an SA-MBN deployment can result in additional overhead due to coordination between distant BSs. This issue can be resolved by using Int-MBN where a user can associate with a single hybrid BS equipped with multi-band transceiver.

4.3 Resource Allocation for MBNs

Resource allocation in MBNs involves several coupled decisions, including power allocation, bandwidth allocation, sub-channel assignment, and user association. These decisions become challenging because MBNs combine narrow-band and wide-band transmissions while also depending on environment-specific propagation conditions and transceiver-specific limitations. In the following, we discuss representative objective functions and design constraints that can arise in resource allocation problems for MBNs.

4.3.1 Objective Functions

In addition to commonly used objectives such as data rate, energy efficiency, and fairness, resource allocation in MBNs can also be designed based on the following performance metrics.

- **DCE (bps/Hz/$):** This metric measures the cost-effectiveness of MBNs by relating the achievable network spectral efficiency to the total cost of deploying and operating the network infrastructure. This cost includes capital expenditures, such as BS deployment and backhauling links, as well as operational expenditures, such as maintenance, energy consumption, and site rental. The metric can also be extended to include the energy cost of communication and sensing services. Optimizing DCE is useful for reducing the overall cost of 6G network deployment, improving the return on investment, supporting scalable network expansion, and promoting more sustainable and energy-efficient operation.

- **HO-aware data rate (bps):** Analyzing the HO-aware data rate is important in MBNs because different frequency bands provide different coverage regions. This can lead to frequent or unnecessary HOs, especially in 6G and beyond networks that are expected to support high-mobility applications, such as high-speed railway systems. Mathematically, the HO-aware data rate can be defined as $R_{HA} = R(1-H_c) = R(1-\min\{H_d, 1\})$, where R denotes the achievable data rate of the user, $H_c = \min\{H_d, 1\}$ is the normalized HO cost, and $H_d = h_d H$. Here, h_d represents the HO delay and H denotes the HO rate [69, 103]. Therefore, HO-aware resource allocation in MBNs should account for HO initiation criteria, HO thresholds, and HO execution time in order to reduce service interruptions and data rate degradation.

- **Data Rate in Finite Block Length Regime [104]:** For a channel with Shannon capacity C, the maximum achievable transmission rate in the finite block length regime depends on the block-length N, the error probability ϵ, and the channel dispersion $V(\gamma) = \frac{\gamma(2+\gamma)}{(1+\gamma)^2}\log_2^2 e$, i.e.,

$$R_{\text{FBL}} = C - \sqrt{\frac{V(\gamma)}{N}}Q^{-1}(\epsilon),$$

  where γ is the SINR, and $Q(\cdot)$ is the inverse of Gaussian-Q-function. Optimizing the finite block length data rate in MBNs is important for supporting URLLC services in 6G and beyond networks.

- **QoS Deviation** [105]: Minimizing the weighted sum of the deviation between the QoS observed by a given network application and its corresponding QoS requirement can provide useful insight into whether the available resources are

sufficient to satisfy users' QoS demands under the existing service level agreement.

4.3.2 Design Constraints

The following constraints should be considered along with the minimum data rate requirements for access and backhaul links, backhaul flow constraints, latency requirements, and transmit power budgets.

- **Multi-connectivity:** Existing wireless networks and their resource allocation schemes typically assign each user to a single BS, which corresponds to a single-connectivity constraint. However, this constraint should be relaxed as intelligent user devices with multi-homing capabilities become more common. For a given MBN with I BSs, the multi-connectivity constraint for a single user can be given by:

$$\sum_{i=1}^{I} x_i \leq I, \tag{4.1}$$

  where $x_i \in \{0, 1\}$ is connectivity of a given user to BS i. Implementing multi-connectivity requires careful interference management.

- **Backhaul-Aware Data Rate:** Backhaul capacity constraints are more involved in MBNs than in conventional single-band networks because of the co-existence of different bandwidths and the aggregation of traffic across multiple bands. For example, BSs in Int-MBNs require ultra-high-capacity backhaul links and more sophisticated backhauling strategies to support simultaneous transmissions over multiple bands. In contrast, SA-MBNs use single-band BSs,

but they may require a larger number of backhaul links since each band is supported by a separate BS deployment. Therefore, the backhaul architecture must be able to handle heterogeneous traffic with different QoS requirements while providing sufficient capacity. Mathematically, the backhaul capacity constraint for a BS in any type of MBN can be expressed as

$$\sum_{s=1}^{S}\sum_{k=1}^{K} R_{s,k} \leq C_{\mathrm{BH}}^{i}, \tag{4.2}$$

where S and K denote the number of multiple transmission links operating in different frequency bands and users, respectively, $R_{s,k}$ is the access network traffic for user k on band s, and C_{BH}^{i} is the backhaul link's capacity for the BS i. In addition, each user's backhaul-aware rate constraint

$$\min\left(C_{\mathrm{BH}}^{i}, \sum_{s=1}^{S} R_{s,k}\right) \geq R_{\mathrm{th}}^{k}, \tag{4.3}$$

must be guaranteed, where R_{th}^{k} denotes the QoS requirement of user k.

- **Next Generation Multiple Access Schemes:** The selection of the multiple access scheme, such as NOMA, space division multiple access, or RSMA, can influence the structure of the constraints in MBN resource allocation problems. For example, RSMA has recently attracted attention as an effective multiple access technique. In RSMA, each user's message is divided into two parts: a common part that can be decoded by all users and a private part intended only for that user [106]. For an RSMA-based MBN with I BSs and K users, the corresponding optimization constraints can include the following:

– *Power budget:* The downlink transmit power constraints for SA-MBN and Int-MBN can be expressed as $\sum_{k=0}^{K} P_k \leq P_{\max}$ and $\sum_{i=1}^{I}\sum_{k=0}^{K} P_{i,k} \leq P_{\max}$, respectively, where P_0, P_k for $k \geq 1$, $P_{i,0}$, $P_{i,k}$ for $k \geq 1$ and $P_{\max}$ are the transmit power of the common message for SA-MBN, the transmit power of the private message for user k in SA-MBN, the transmit power of the common message for BS i in Int-MBN, the transmit power of the private message of user k from BS i in Int-MBN, and the maximum allowable transmit power.

– *Common message rate:* To ensure that the common stream can be successfully decoded by all users in an Int-MBN, the constraint

$$\sum_{i=1}^{I}\sum_{k=1}^{K} C_{i,k} \leq \sum_{i=1}^{I} R_0\left(P_{i,0}, P_{i,k}, H\left(f_i\right)\right), k \geq 1,$$

with $C_{i,k}$, R_0 and $H\left(f_i\right)$ being the portion of the common rate of user k from the BS i, a function for the common rate of user k, and the channel power gain of the BS i, respectively, is required.

– *QoS constraint:* To guarantee that the achievable rate of user k satisfies the minimum rate requirement R_{th} in an Int-MBN, the constraint

$$\sum_{i=1}^{I} C_{i,k} + \sum_{i=1}^{I}\sum_{k=1}^{K} R_{i,k}\left(P_{i,k}, H\left(f_i\right)\right) \geq R_{\text{th}},$$

where $R_{i,k}$ is a function for the private data rate of user k from BS i, is needed.

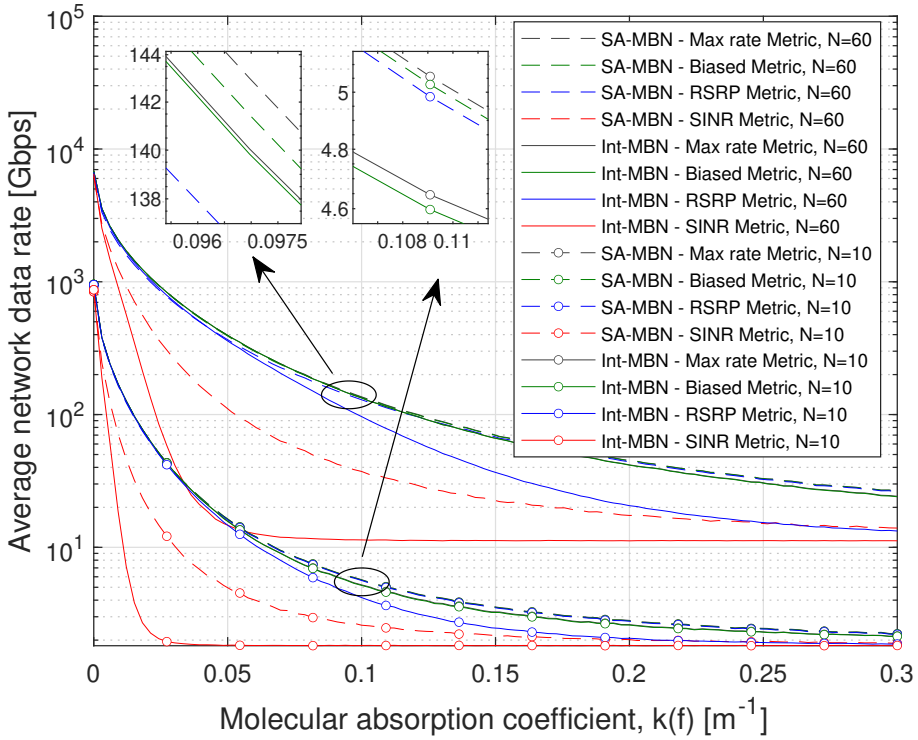


Figure 4.2: Network data rate versus molecular absorption coefficient for different number of BSs and UE-BS association metrics *Adapted from [98]. © 2024 IEEE.*

4.4 User Offloading in RF/THz MBN: A Case Study

We also propose a novel simplified metric to perform efficient molecular absorption-aware offloading in MBNs and demonstrate its superior performance compared to the standard solutions.

For SA-MBN deployment, we consider N_T single antenna THz BSs (TBSs) and N_R single antenna RF BSs (RBSs). For Int-MBN deployment, we consider N_{Hyb} hybrid BSs operating on both the THz and RF frequencies. We investigate the performance of a typical single antenna user. The RF and THz channels are modeled according to [80]. The simulation parameters are as follows. Antenna gain is 25 [dB]

for THz transceivers and 0 [dB] for RF transceivers. The carrier frequency of THz is 1 [THz] and is 2 [GHz] for RF transmission. The probability of beam alignment with interfering TBSs is set to be 0.006. RF bandwidth is 40 [MHz], and the transmit power of RBSs and TBSs are 2 and 0.2 [Watts], respectively.

We propose a molecular absorption-aware bias for RBS and TBS as:

$$\text{Bias}_{\text{R}} = B_R p_R^{Rx}, \quad \text{Bias}_{\text{T}} = B_T p_T^{Rx} \exp(-K(f_T)d), \tag{4.4}$$

where B_R and B_T denote the RF and THz transmission channel bandwidths, respectively. Also, p_R^{Rx}, p_T^{Rx}, $\exp(K(f_T))$, and d denote the RF and THz received powers obtained using (1) and (3) in [80], molecular absorption coefficient at carrier frequency f_T, and the distance between BS and the typical user, respectively.

Fig. 4.2 compares various user offloading metrics in terms of network average data rate for both SA-MBN and Int-MBN. The maximum data rate metric based association outperforms others. However, the proposed bias-based association reduces CSI overhead and computational complexity, while attains a performance comparable to the maximum data rate for both SA-MBN and Int-MBN deployments. The proposed scheme is followed by the reference signal received power (RSRP) metric and the SINR metric. The inferior performance of the SINR metric is due to not considering the impact of transmission bandwidths. Hence, when using SINR as the association metric, RF is selected most of the time, and even by increasing the number of BSs, we cannot attain the high data rate provided by TBSs.

Considering different values of the molecular absorption coefficient, i.e., $K(f_T)$, reflects a variety of environments with varying humidity, temperature, and weather. By increasing the molecular absorption, the Int-MBN performance degrades more

sharply as user association probability to TBS is higher in Int-MBN. By increasing the number of BSs from $N = 10$ to $N = 60$, the average network data rate improves due to the higher association to TBS in SA-MBN and THz mode of hybrid BSs in Int-MBN.

4.5 SA-MBN vs Int-MBN: A Case Study

This section provides a qualitative and quantitative comparison of the two MBN architectures in terms of BS deployment density, user offloading, spectral efficiency, and deployment cost efficiency.

4.5.1 Deployment Density

Fig. 4.3 demonstrates the number of required BSs to reach the target rate requirement considering both SA-MBN and Int-MBN for different THz transmission bandwidths, i.e., B_T. We first consider that the number of BSs in SA-MBN increases equally as Int-MBN, i.e., $N_T = N_R = N_{\text{Hyb}}$. It can be seen that as we increase the target rate, SA-MBN requires a higher number of BSs compared to Int-MBN. Then, we investigate the case of SA-MBN with flexible number of BSs (*SA-MBN-FN*), i.e., selecting the number of RBSs and TBSs unequally. It can be seen that SA-MBN-FN can approach Int-MBN in terms of deployment density. That is, high interference in RF transmissions and low communication distance of THz transmissions can be compensated by raising the number of TBSs and lowering the RBSs (while minimizing the total number of BSs to a level similar to Int-MBN). Nonetheless, the deployment cost can vary for the three configurations.

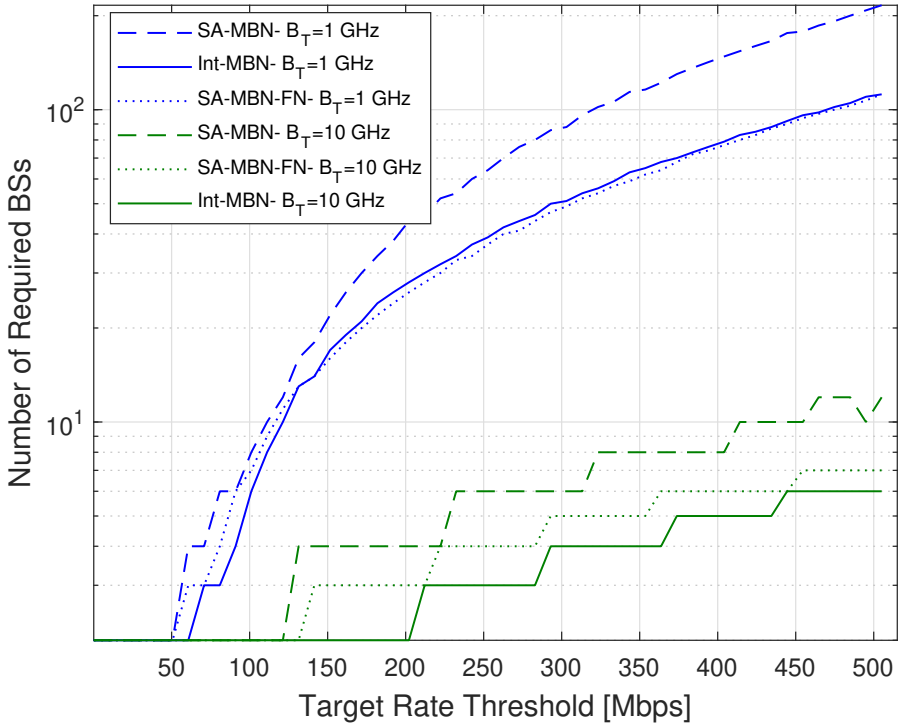


Figure 4.3: The required number of BSs versus the target rate threshold. $R_{\max} = 400[m]$, $P_R^{tx} = 3$[Watts] *Adapted from [98]. © 2024 IEEE.*

Along another note, we observe that increased THz bandwidth improves the system performance, as it lets the typical randomly selected user achieves the required data rate with less number of BSs.

4.5.2 Association to TBSs

Fig. 4.4 shows the association probability of a user with TBSs as a function of the number of BSs. Note that RF association can simply be calculated by subtracting the association probability with TBSs from one. We note that a user in Int-MBN selects TBS with a higher probability than RBS. The reason is that both frequency bands are available at each deployed hybrid BS. The user, thus, would select the one

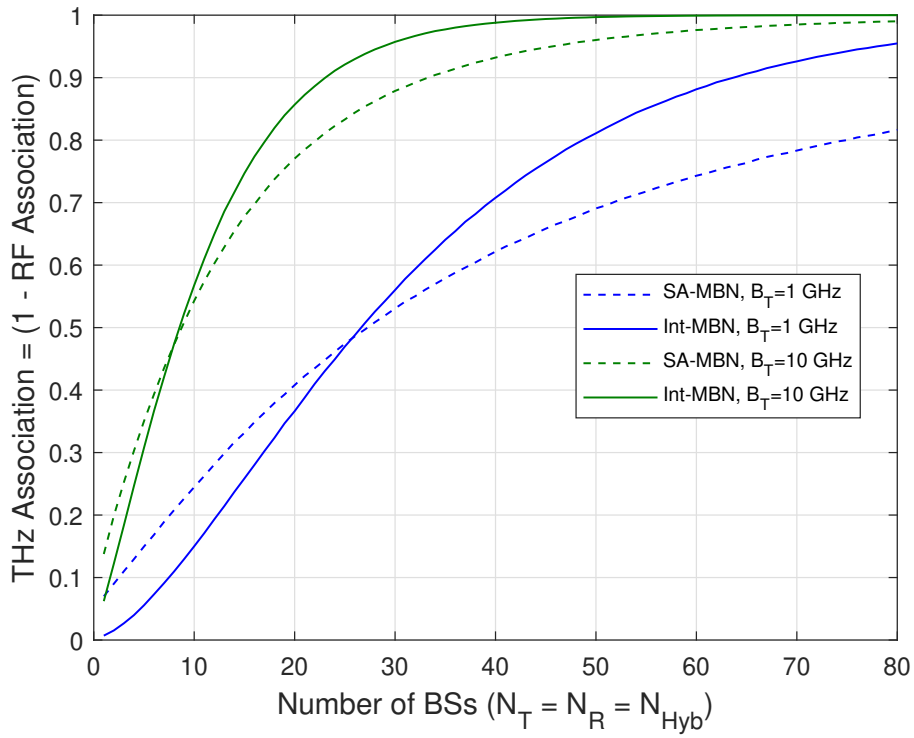


Figure 4.4: Impact of increasing the number of BSs at various THz bandwidth on THz association probability *Adapted from [98]. © 2024 IEEE.*

that provides a higher data rate, which is the THz. On the other hand, in SA-MBN, the user may not have access to closer TBS thus may associate to a nearby RBS. The association of users with TBSs increases exponentially when the THz transmission bandwidth is large.

4.5.3 Spectral Efficiency and Achievable Rate

In Fig. 4.5, the SE and data rate of each deployment for different THz bandwidths is demonstrated. The value of SE in the MBN is not expected to be huge as SE in THz band is typically smaller than that of RF because of the higher spreading losses and the existence of molecular absorption loss. However, the THz performance gains

are achieved mainly due to large THz bandwidth, which results in high achievable data rates. Subsequently, the larger the THz transmission bandwidth the higher is the achievable data rate (in the order of Gbps) per user as is shown in Fig. 4(b). Also, we note that the SE of SA-MBN is higher than that of Int-MBN. The reason is that in SA-MBN, BSs are more distributed in a given area. Therefore, the likelihood of finding a nearby BS is greater in SA-MBN.

In addition, we note that SE improves at first and then degrades as the number of BSs increases. When the number of BSs are less and more users are associating to RF BSs, the SE first increases due to increase in received signal power and then reduces due to increasing RF interference. This uni-modal shape is visible in the region where the association to RBS is higher than TBS.

4.5.4 Deployment Cost Efficiency

Deployment Cost Efficiency is defined as the ratio of achievable network spectral efficiency (SE) and the total investment in the installation of the network infrastructure, i.e., capital expenditure (CAPEX), which includes deployment of BSs, backhauling links plus the operational expenditure (OPEX), including maintenance, energy costs, and land lease.

We assume that TBSs have the same estimated deployment cost as mmWave BSs whose values are available in [107, 108]. We estimate the CAPEX and OPEX as follows. The deployment of each TBS needs \$38K, each RBS costs \$33K, and hybrid BS requires \$48K. Since hybrid BSs require sophisticated transceiver which operates at different frequencies, it is assumed that the transceiver costs more than the cost of BSs operating in a single frequency. For OPEX, we consider \$2.8K per RBS and

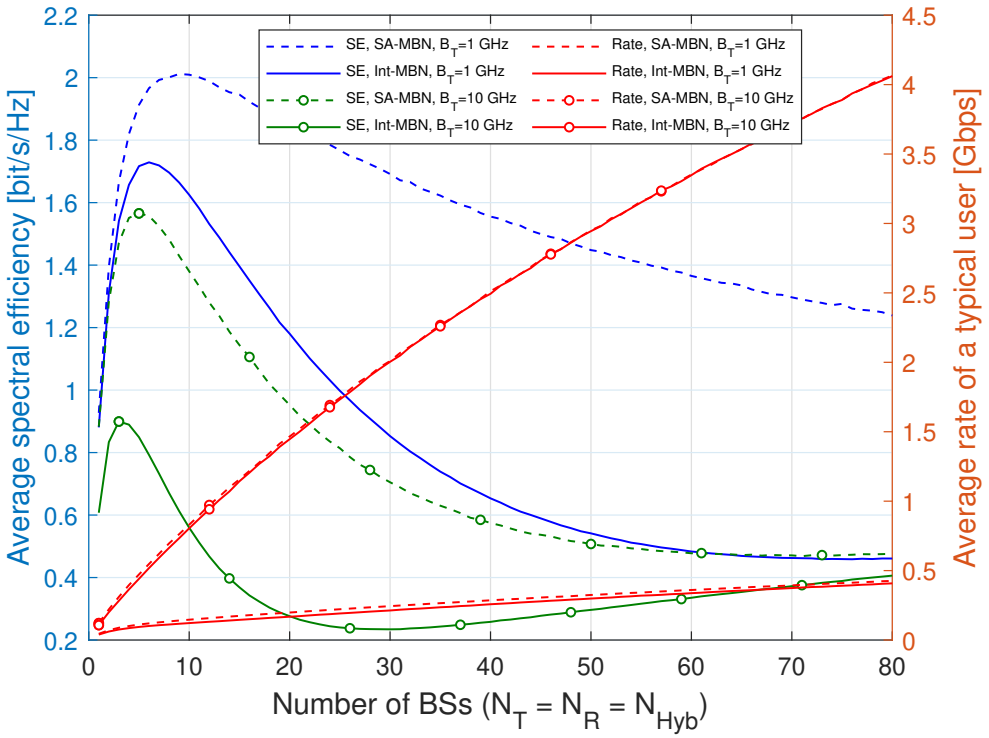


Figure 4.5: Impact of increasing the number of BSs at various THz bandwidth on Spectral efficiency and data rate *Adapted from [98]. © 2024 IEEE.*

$2.7K per TBS and $3.2K per hybrid BS. Note that these values are used as estimates for comparison and do not include all practical aspects. We define the deployment cost efficiency for SA-MBN and Int-MBN as

$$\mathrm{DCE}_{\mathrm{SA}} = \frac{(N_R + N_T)\mathrm{SE}^{\mathrm{SA}}}{N_R C_R + N_T C_T}, \quad \mathrm{DCE}_{\mathrm{Int}} = \frac{2N_{\mathrm{Hyb}}\mathrm{SE}^{\mathrm{Int}}}{N_{\mathrm{Hyb}} C_H}, \tag{4.5}$$

where $\mathrm{SE}^{\mathrm{SA}}$ and $\mathrm{SE}^{\mathrm{Int}}$ are the average user spectral efficiency of SA-MBN and Int-MBN, respectively. Also, C_T, C_R, and C_H denote the cost of TBS, RBS, and hybrid BS, respectively.

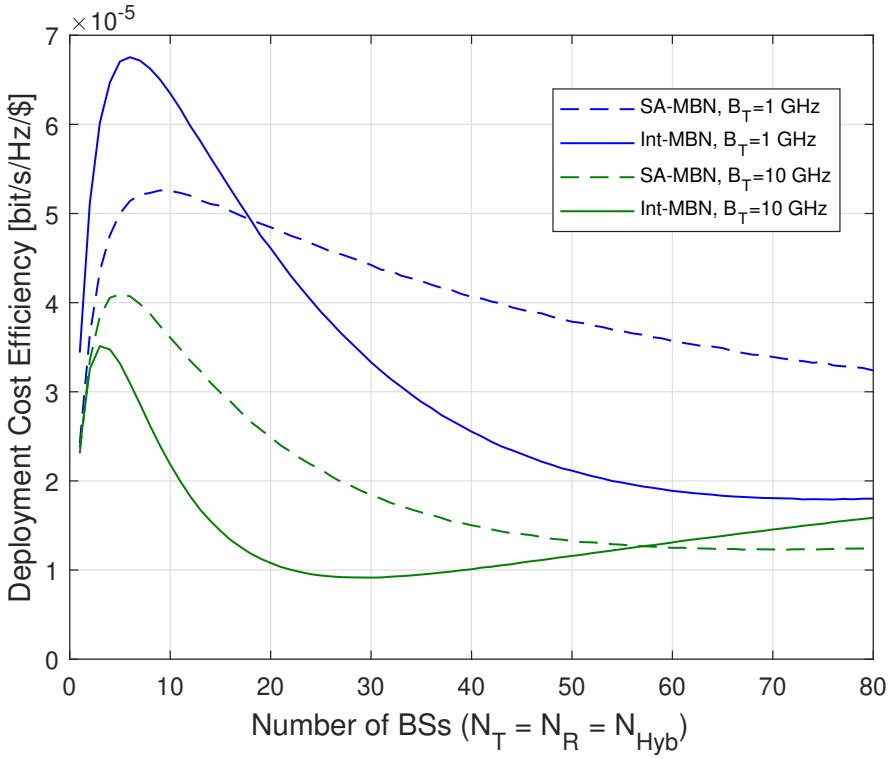


Figure 4.6: Impact of increasing the number of BSs at various THz bandwidth on deployment cost efficiency *Adapted from [98]. © 2024 IEEE.*

Fig. 4.6 depicts DCE as the number of BSs increases. Since the cost of SA-MBN is higher than Int-MBN, scaling the SE of the two configurations with cost demonstrates the improved DCE of Int-MBN. Also, for $B_T = 10$ GHz, the THz association shows that the system rapidly tends to operate in THz, which imposes much less interference on the desired signals. Therefore, after 30 BSs, where THz is more than 0.95 percent used, without experiencing too much interference, DCE starts to increase again. Compared with SA-MBN, these results are primarily observable in Int-MBN since both THz and RF bands are available to be selected at each deployed BS.

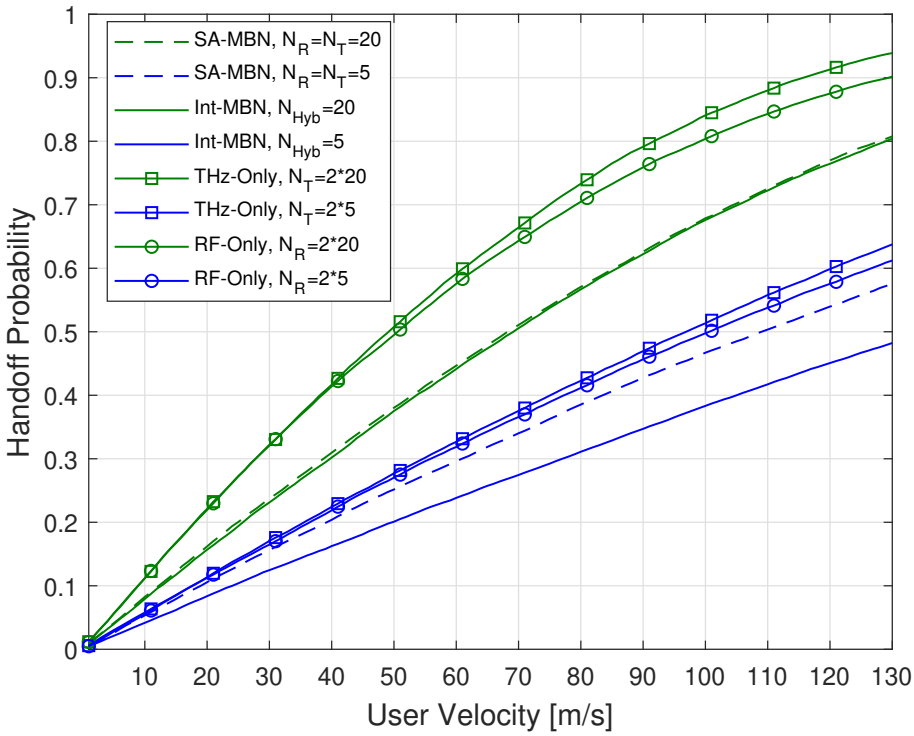


Figure 4.7: HO probability versus the velocity of user for different network architecture *Adapted from [72]. © 2024 IEEE.*

4.5.5 Handoff Probability in MBNs

Figure 4.7 demonstrates the HO probability of the user of interest moving at a predefined velocity considering both SA-MBN and Int-MBN deployments and THz-Only and RF-Only. It can be observed that both MBN deployments outperform the use of only one frequency band. It is intuitive that for a higher number of BSs and velocity, HO probability increases. For higher number of BSs, Int-MBN has a marginal gain over SA-MBN. In comparison, for a lower number of BSs, Int-MBN consistently outperforms SA-MBN when velocity increases. Seamless group HOs would be critical due to multi-connectivity considerations in MBN.

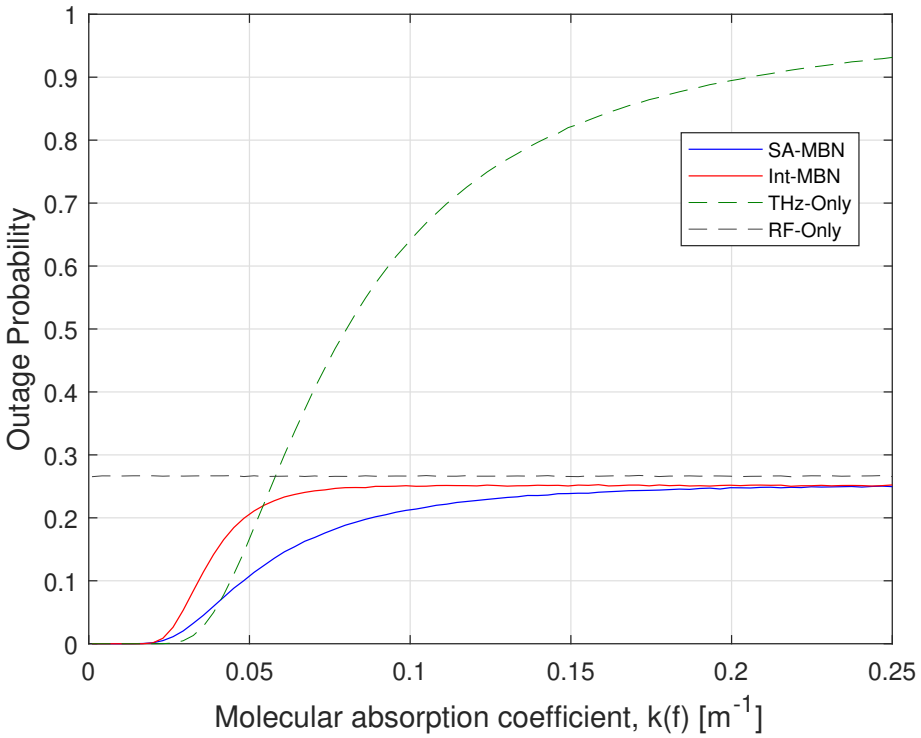


Figure 4.8: Outage probability versus the molecular absorption coefficient for average network rate threshold of 1 GHz, $N_T = N_R = N_{\text{Hyb}} = 15$ *Adapted from [72]. © 2024 IEEE.*

4.5.6 Significance of MBNs over Standalone Counterparts

Fig. 4.8 illustrates the influence of varying the frequency-dependent molecular absorption coefficient $k(f)$ values on outage probability for various network architectures. In this figure, the molecular absorption-aware association metric is employed. It can be observed that the SA-MBN outperforms all other scenarios as $k(f)$ increases, eventually approaching RF-Only as no THz transmission is possible at high $k(f)$. For the same reason, THz-only performance degrades significantly. It is also observed that for small values of $k(f)$, the use of THz-only is advantageous. Based on the results depicted in figures 4.8 and 4.9, it is crucial to select the transmission bands

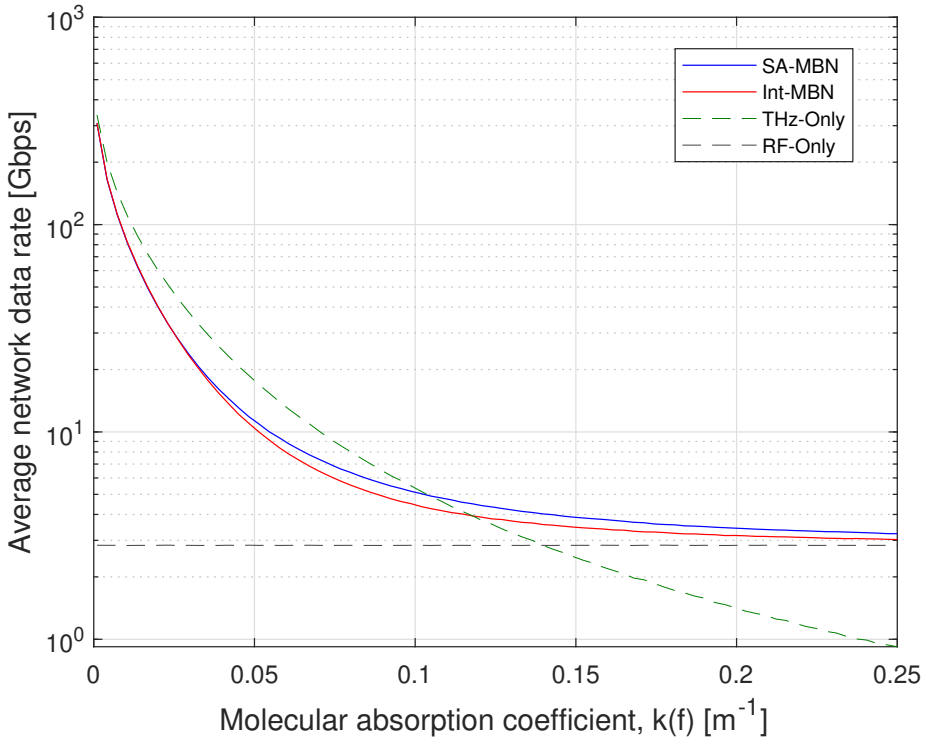


Figure 4.9: Average network data rate versus the molecular absorption coefficient
Adapted from [72]. © 2024 IEEE.

adaptively.

Similar insights can be obtained from Figure 4.9 that depicts the average network data rate for a SA-MBN, an Int-MBN, and THz-only and RF-only networks for various molecular absorption coefficients.

4.6 Summary

In this chapter, we comparatively analyzed two potential MBN architectures. We also highlighted the potential challenges of such configurations. In the simulation results, it shown that the proposed user offloading metric can achieve the performance of the maximum data rate association rule. It is shown that SA-MBN requires

a higher number of BSs to achieve the same performance as Int-MBN for a given target data rate and equal proportion of RBSs and TBSs. However, SA-MBN can achieve the performance of Int-MBN if we allow unequal proportion of RBS and TBS deployments.

Copyright and Co-authorship Statement

This chapter has resulted in the following publication: M. A. Saeidi, H. Tabassum, and M.-S. Alouini, "Multi-Band Wireless Networks: Architectures, Challenges, and Comparative Analysis," *IEEE Communications Magazine*, vol. 62, no. 1, pp. 80–86, Jan. 2024, doi: 10.1109/MCOM.003.2300078 [98]. © 2024 IEEE.

I was the primary author of this work and was responsible for the development of the main idea, technical analysis, numerical studies, interpretation of results, and manuscript preparation. Author 2 supervised the work and refined the manuscript. Author 3 reviewed the manuscript and provided feedback.

A portion of this chapter has resulted in the following publication: S. Aboagye, M. A. Saeidi, H. Tabassum, Y. Tayyar, E. Hossain, H.-C. Yang, and M.-S. Alouini, "Multi-Band Wireless Communication Networks: Fundamentals, Challenges, and Resource Allocation," *IEEE Transactions on Communications*, vol. 72, no. 7, pp. 4333–4361, Jul. 2024, doi: 10.1109/TCOMM.2024.3366816 [72]. © 2024 IEEE.

I contributed to the corresponding portion used in this chapter, including the numerical studies and interpretation of results. Authors 1 and 3–7 contributed to the development and refinement of that manuscript.

Chapter 5

Resource Allocation in Cooperative Mid-band/THz Networks in the Presence of Mobility

5.1 Overview

Chapter 4 characterized MBNs and showed that different frequency bands can be jointly used to balance data rate, coverage, mobility support, and deployment cost. Building on that discussion, this chapter studies a concrete cooperative MBN architecture that combines the upper mid-band (UMB) and THz bands. The objective is to investigate how UMB/THz cooperation can support high-rate transmission while reducing the mobility and blockage limitations of THz-only networks.

The UMB, ranging from 7 to 24 GHz, has recently gained attention because it offers more spectrum than the conventional sub-6 GHz band while providing better propagation and coverage characteristics than mmWave frequencies [3, 4]. In particular, the UMB operating in the 7.75–8.4 GHz range is intended for both outdoor and indoor applications in urban and suburban areas, aiming to meet high data rate and

moderate mobility requirements [4]. On the other hand, frequency bands above 100 GHz, such as the THz band, can provide substantial bandwidth for data-intensive services [109]. The technical feasibility of using frequency bands above 100 GHz for international mobile telecommunications has also been investigated in the ITU report [110].

The UMB and THz bands have complementary characteristics. THz links can provide very large bandwidths, but their coverage is limited by molecular absorption, blockage, and severe propagation loss. These limitations become more critical for mobile users, since a user may frequently switch serving BSs or lose the line-of-sight link. In contrast, UMB links provide more stable coverage and better mobility support, although with less bandwidth than THz links. Therefore, a UMB/THz MBN can exploit the large bandwidth of the THz band while using the UMB to improve coverage and mobility support.

In addition to multi-band operation, cooperative transmission can improve the reliability of high-frequency networks. By allowing each user to be served by a cluster of BSs, coordinated multi-point transmission can increase the received power, provide spatial diversity, and reduce the impact of blockage. However, cooperation also creates a new mobility-related trade-off. Increasing the cooperative cluster size may improve the instantaneous rate, but it can also increase the number of handovers when users move. Therefore, in mobile cooperative MBNs, user association, cluster size, beamforming, and handover cost must be designed jointly.

Another important aspect is the use of extremely large antenna arrays (ELAAs). ELAAs can provide high beamforming gains to compensate for the propagation loss at high frequencies. However, their large aperture extends the near-field region, making

conventional far-field channel models less accurate. As a result, the channel should account for the spherical wavefront and the distance between the user and each antenna element. For mobile users, the relative velocity with respect to each antenna element should also be included in the channel model. This motivates the use of near-field and velocity-aware channel modeling in the proposed cooperative UMB/THz MBN framework.

To this end, this chapter develops a resource allocation framework for cooperative UMB/THz MBNs in the presence of user mobility. The framework captures near-field channels, user velocity, QoS requirements, blockage, cooperative BS clustering, and hybrid beamforming. It addresses the following main question: how can UMB/THz multi-band operation and cooperative transmission support high-frequency THz networks under mobility, and what are the resulting trade-offs between sum-rate, handovers, cluster size, and beamforming accuracy?

The main contributions of this chapter can be summarized as follows:

• We propose a cooperative UMB/THz MBN, where BSs are equipped with ELAAs, and formulate the achievable rate based on accurate NFC channel modeling. This model considers spherical wavefronts to capture not only the impact of the distance between users and each antenna element, but also the relative velocity of users with respect to each antenna element. We then formulate resource allocation problems in two scenarios (a) to maximize the system sum-rate considering *stationary* users and (b) maximize the HO-aware sum-rate for *moving* users. We optimize analog and digital beamforming, and user association variables, subject to power constraints, maximum cluster size, users' QoS, blockages, etc.

• We develop a unified solution to the sum-rate maximization problem for stationary users in cooperative THz/UMB MBN. First, we derive a closed-form feasible solution for the analog beamformers and reformulate the power budget constraint accordingly. We then transform the coupled user association and digital beamforming variables using big-M approach. Using FP and quadratic transformation, we then handle the non-convexity of the ratio terms in the objective function and constraints. We prove that applying a quadratic transformation to the FP problem with a logarithmic fractional objective and QoS constraints yields a problem that is equivalent to the original FP. Additionally, we use the Majorization-Minimization (MM) method to address the binary user association variables. Finally, we transform the non-convex problem into a convex one and propose an iterative algorithm to jointly solve for digital beamformers and user association variables.

• Building on the proposed framework for stationary users, we derive two solutions for the HO-aware resource allocation problem. In the first solution, we maximize the HO-aware rate subject to HO-aware QoS and soft HO constraints. We prove that the HO-aware rate is a *log-concave* function after applying the quadratic transformation. Using this property, we demonstrate that the non-convex optimization problem can be tightly approximated by a convex problem leveraging Jensen's inequality and logarithmic properties. In the second approach, we formulate a multi-objective optimization problem (MOOP) that aims to maximize the system sum-rate, while minimizing the total number of HOs. This problem is transformed into a weighted-objective optimization and solved using the algorithm proposed for the stationary scenario.

• Numerical results validate the effectiveness of the proposed algorithm and quantify the performance gains of cooperative MBN compared to a stand-alone THz network, both with and without mobility. While requiring more complexity for channel estimation, the results highlight the importance of NFC-based beamforming with ELAAs, as opposed to its FFC-based counterpart. Additionally, it is observed that increasing the user-centric cluster size is not always beneficial in mobility scenarios within cooperative systems.

The rest of this chapter is organized as follows. Section 5.2 presents the considered cooperative UMB/THz MBN system model, including the near-field communication model, THz and UMB channel models, and downlink SINR and rate formulations. Section 5.3 formulates and solves the joint user association and hybrid beamforming problem for stationary users in cooperative UMB/THz MBNs. Section 5.4 extends the proposed framework to mobility-aware resource allocation by incorporating HO-aware rate and HO-aware optimization approaches. Section 5.5 presents numerical results and discussions under different deployment, blockage, molecular absorption, antenna, near-field modeling, mobility, cluster-size, and imperfect CSI settings. Finally, Section 5.7 summarizes the chapter.

Notations: Boldface lowercase letters represent vectors and boldface capital letters to represent matrices. $\mathbb{B}^{B\times K}$ and $\mathbb{R}^{B\times K}$ denote the $B\times K$ space of binary-valued and real-valued numbers, respectively. $\mathbb{C}^{B\times K}$ represent the $B\times K$ space of complex numbers. $|\boldsymbol{x}|$ is L2 norm of complex vector $\boldsymbol{x}$ and is the absolute value for scalar values. $\|\boldsymbol{X}\|_F$ denotes the frobenius norm of matrix $\boldsymbol{X}$. $\boldsymbol{X}^H$ is the conjugate transpose of complex matrix $\boldsymbol{X}$, and $\boldsymbol{X}^{\mathbb{T}}$ is only the transpose of the $\boldsymbol{X}$. $\Re\{\boldsymbol{X}\}$ denotes the real part of $\boldsymbol{X}$, and x^* is the conjugate of complex number x. $\boldsymbol{X}=[.,.,.]$ denotes

different column of matrix $\boldsymbol{X}$. $\nabla \boldsymbol{f}$ is the gradient of function $\boldsymbol{f}$. $\odot$ and $\otimes$ denote the Hadamard and Kronecker products, respectively.

5.2 System Model and Assumptions

We consider an MBN, where B^{T} THz BSs (TBSs) and B^{U} UMB BSs (UBSs) are deployed along a corridor of length D and all BSs are connected to a central processor unit (CPU) through high capacity backhaul links. $\mathcal{B}_{\mathcal{T}} = \{1, \ldots, B^{\mathrm{T}}\}$ and $\mathcal{B}_{\mathcal{U}} = \{1, \ldots, B^{\mathrm{U}}\}$ denote the set of all TBSs and UBSs, respectively. Also, K single-antenna mobile users, are deployed on the corridor such that the velocity of user k is denoted by v_k. We assume that each user is equipped with two separate terminals to receive signals at both the UMB and THz spectrum [111].[1]

We consider that a cluster of BSs in each frequency band can serve a user in a cooperative manner. Each TBS is equipped with a uniform linear array (ULA) of M^{T} antennas connected to a hybrid beamformer (HBF). Each HBF at the TBSs comprises of an analog beamformer and a digital beamformer (precoder). The analog beamformer, which is a network of phase shifters, is connected to the digital beamformer via N^{T} RF chains ($M^{\mathrm{T}} \geq N^{\mathrm{T}}$) either with fully or partially-connected architecture. Each UBS has M^{U} antennas connected to HBF with N^{U} RF chains, such that $M^{\mathrm{U}} \geq N^{\mathrm{U}}$ [30].

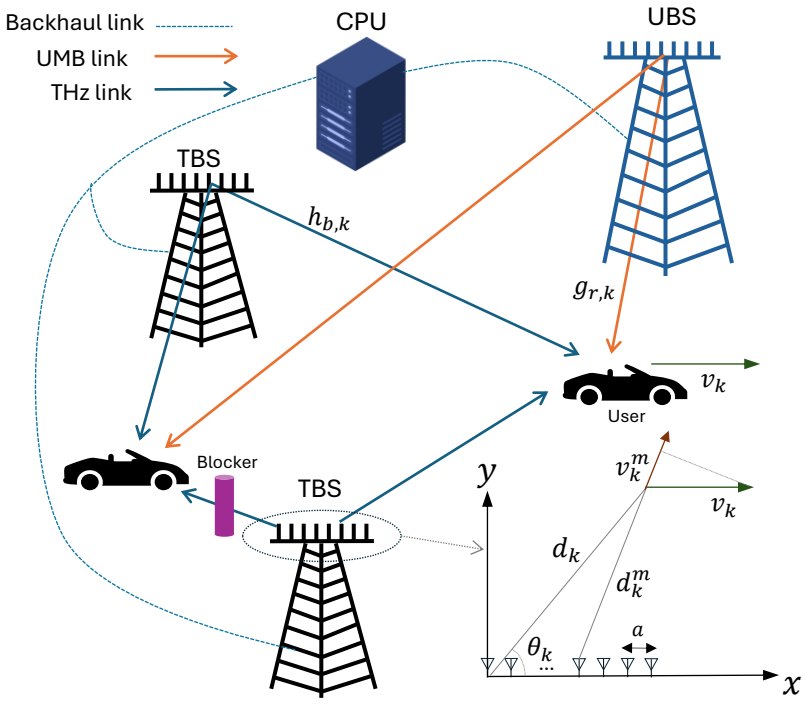


Figure 5.1: System model illustrating BSs and users and NFC model. *Adapted from [114]. © 2025 IEEE.*

5.2.1 Near Field Communication Model

In this subsection, we introduce the NFC modeling regardless of the frequency band, and the notations used here are general. Without loss of generality, according to Fig. 5.1, each ULA is deployed along the x-axis, which is in parallel with the users' movement. Denoting the distance between the first antenna element of BS $i \in \mathcal{B}_{\mathcal{T}} \cup \mathcal{B}_{\mathcal{U}}$, and user k by $d_{i,k} = d_{i,k}^{(0)}$ and the angle of the user with respect to the first antenna element of a BS by $\theta_{i,k}$, based on the near-field spherical wavefront

[1]In fixed wireless access and vehicular communication applications, as considered in this work, the practical challenges of implementing separate hardware, such as size constraints, are less stringent. Moreover, advanced unified MBN transceivers based on up/down conversion techniques can be utilized to enable the reception of signals across both frequency bands [72, 112, 113]. For further discussion of the practical challenges associated with MBNs and potential solution approaches, we refer the reader to [72, 98].

modeling, the exact distance between user k and m-th antenna of BS i is obtained as [115]:

$$d_{i,k}^{(m)}(d_{i,k},\theta_{i,k}) = \sqrt{d_{i,k}^2 + m^2x^2 - 2d_{i,k}mx\cos(\theta_{i,k})}, \tag{5.1}$$

where $0 \leq m \leq M-1$, and x is the antenna spacing, which can be set as half the wavelength of the operating carrier frequency (λ). Based on NFC model in (5.1), the array response of BS i with respect to user k at angle of departure (AoD) of $\theta_{i,k}$ is given by:

$$\boldsymbol{a}_{i,k}(d_{i,k},\theta_{i,k}) = \left[e^{-j\frac{2\pi}{\lambda}d_{i,k}^{(0)}},\ldots,e^{-j\frac{2\pi}{\lambda}d_{i,k}^{(M-1)}}\right]^{\mathbb{T}}. \tag{5.2}$$

Utilizing the distance between user k and the m-th antenna of BS i, we need to obtain the NFC-aware velocity of user k, which can vary with respect to each antenna element. To do so, the projection of v_k onto the line connecting the m-th antenna to the user, i.e. $d_{i,k}^{(m)}$ is derived using:

$$v_{i,k}^{(m)}(d_{i,k},\theta_{i,k}) = \frac{d_{i,k}\cos(\theta_{i,k}) - mx}{d_{i,k}^{(m)}}v_k. \tag{5.3}$$

Then, the Doppler-shift of user k at BS i is obtained by:

$$\boldsymbol{v}_{i,k}(d_{i,k},\theta_{i,k},v_k) = \left[e^{-j\frac{2\pi}{\lambda}v_{i,k}^{(0)}\tau},\ldots,e^{-j\frac{2\pi}{\lambda}v_{i,k}^{(M-1)}\tau}\right]^{\mathbb{T}}, \tag{5.4}$$

where τ is the duration between two trajectory points. Based on the array response in (5.2) and Doppler-shift in (5.4), we define the velocity-aware effective array response as [116]:

$$\boldsymbol{b}_{i,k}(d_{i,k},\theta_{i,k},v_k) = \boldsymbol{a}_{i,k}(d_{i,k},\theta_{i,k}) \odot \boldsymbol{v}_{i,k}(d_{i,k},\theta_{i,k},v_k). \tag{5.5}$$

From this point onward, the indices $b \in \mathcal{B}_{\mathcal{T}}$ and $r \in \mathcal{B}_{\mathcal{R}}$ are used for the b-th TBS and r-th UBS, respectively.

The analysis of parameter estimation for CSI acquisition, such as user location and velocity, and the associated overhead, is beyond the scope of this work. However, data-driven approaches and recent advancements in integrated sensing and communication, as well as cooperative strategies involving low Earth orbit satellites, provide promising directions for improving the estimation of these parameters and enhancing the overall quality of CSI acquisition [117–120]. In addition, the overhead associated with CSI transmission can be reduced through dynamic CSI compression techniques, such as the method proposed in [121].

5.2.2 Channel Model

5.2.2.1 THz Channel

We consider the LoS transmission links in the THz band. According to [72, 109], and based on Beer-Lambert law, the path-loss between TBS b and user k is given by:

$$PL_{b,k} = \left(\frac{c(G_{\mathrm{Tx}}^{\mathrm{T}} G_{\mathrm{Rx}}^{\mathrm{T}})^{\frac{1}{2}}}{4\pi f_{\mathrm{T}} d_{b,k}} \right) e^{-\frac{1}{2}\mathcal{K}(f_{\mathrm{T}}) d_{b,k}}, \tag{5.6}$$

where c is the speed of light, f_{T} is the carrier frequency of the THz band, and $\mathcal{K}(f_{\mathrm{T}})$ is the molecular absorption coefficient, that can be obtained using high-resolution transmission (HITRAN) molecular absorption database [122]. $G_{\mathrm{Tx}}^{\mathrm{T}}$ and $G_{\mathrm{Rx}}^{\mathrm{T}}$ are the gain of each directional antenna element in the arrays of the transmitter and receiver at the THz band, respectively [123]. The re-radiation of the absorbed signals by

molecules affects signal propagation, and the signals experience the following path-loss due to molecular absorption noise [109, 124]:

$$\tilde{PL}_{b,k} = \left(\frac{c(G_{\mathrm{Tx}}^{\mathrm{T}} G_{\mathrm{Rx}}^{\mathrm{T}})^{\frac{1}{2}}}{4\pi f_{\mathrm{T}} d_{b,k}} \right) \left(1 - e^{-\mathcal{K}(f_{\mathrm{T}}) d_{b,k}}\right)^{\frac{1}{2}}. \tag{5.7}$$

We can then formulate the channels at the THz band by combining the path losses and the effective array response as:

$$\boldsymbol{h}_{b,k} = PL_{b,k} \boldsymbol{b}_{b,k}(d_{b,k}, \theta_{b,k}, v_k) \tag{5.8}$$

$$\tilde{\boldsymbol{h}}_{b,k} = \tilde{PL}_{b,k} \boldsymbol{b}_{b,k}(d_{b,k}, \theta_{b,k}, v_k), \tag{5.9}$$

where $\boldsymbol{h}_{b,k} \in \mathbb{C}^{M^{\mathrm{T}} \times 1}$ and $\tilde{\boldsymbol{h}}_{b,k}$ are the direct and molecular absorption THz channels between TBS b and user k, respectively.

Moreover, THz transmission is susceptible to blockage due to high penetration losses [69, 98]; thus, we incorporate link blockage probability. We define $\psi_{b,k} \in \mathbb{B}$ as the indicator variable of the THz band for blockage such that $\psi_{b,k} = 0$ if there is a blockage between user k and TBS b, and otherwise $\psi_{b,k} = 1$. The blockage variable ψ follows the Bernoulli distribution [79, 82]:

$$\psi_{b,k} = \begin{cases} 0, & \text{With probability } \; 1 - e^{-\xi d_{b,k}} \\ 1, & \text{With probability } \; e^{-\xi d_{b,k}}, \end{cases} \tag{5.10}$$

where ξ is the density of blockers in a given area. Considering that if the link between a user and a TBS is blocked, the corresponding channels must have zero vectors.

Therefore, the effective THz channels with blockage are obtained as follows:

$$\boldsymbol{h}_{b,k} = \psi_{b,k}\mathbf{1}_{M^{\mathrm{T}}\times 1} \odot \boldsymbol{h}_{b,k}, \tag{5.11}$$

where $\mathbf{1}_{M^{\mathrm{T}}\times 1}$ is a vector of ones. The same multiplication is also applied to $\tilde{\boldsymbol{h}}_{b,k}$ as well. Moreover, in order to impose the blockage on the user association variables, $\alpha^n_{b,k}$, the following constraints need to be considered.

$$\alpha^n_{b,k} \leq \psi_{b,k}, \forall b, k. \tag{5.12}$$

5.2.2.2 UMB Channel

The path-loss at the UMB frequency range between user k and UBS r is modeled by:

$$PL_{r,k} = \frac{c\ (G^{\mathrm{U}}_{\mathrm{Rx}} G^{\mathrm{U}}_{\mathrm{Tx}})^{\frac{1}{2}}}{4\pi f_{\mathrm{U}}} \left(\frac{1}{d_{r,k}}\right)^{\frac{1}{2}\alpha}, \tag{5.13}$$

where $G^{\mathrm{M}}_{\mathrm{Rx}}$, $G^{\mathrm{M}}_{\mathrm{Tx}}$ are the antenna gain of the transmitter and receiver at UMB, respectively, $d_{r,k}$ is the distance between UBS r and user k, α is the path-loss exponent, and f_{M} is the UMB carrier frequency. Therefore, the channel between UBS r and user k is given by:

$$\boldsymbol{g}_{r,k} = PL_{r,k}\left(\sqrt{\frac{\kappa}{1+\kappa}}\boldsymbol{b}_{r,k}(d_{r,k}, \theta_{r,k}, v_k) + \sqrt{\frac{1}{1+\kappa}}\boldsymbol{g}^{\mathrm{NLoS}}_{r,k}\right), \tag{5.14}$$

where $\boldsymbol{g}_{r,k} \in \mathbb{C}^{M^{\mathrm{U}}\times 1}$ is the channel between UBS r and k-th user, κ is the Rician factor, and $\boldsymbol{g}^{\mathrm{NLoS}}_{r,k}$ is the NLoS component following zero mean and unit variance complex Gaussian distribution.

5.2.3 Downlink SINR and Rate Model

This subsection explains the signal model in the cooperative MBN system. The received signal at the THz terminal of user k is given by:

$$y_k^{\mathrm{T}} = \underbrace{\sum_{b=1}^{B^{\mathrm{T}}} \alpha_{b,k}^n \boldsymbol{h}_{b,k}^H \boldsymbol{F}_b \boldsymbol{w}_{b,k} s_k}_{\text{Desired signals}} + \underbrace{\sum_{k' \neq k}^{K} \sum_{b=1}^{B^{\mathrm{T}}} \alpha_{b,k'}^n \boldsymbol{h}_{b,k}^H \boldsymbol{F}_b \boldsymbol{w}_{b,k'} s_{k'}}_{\text{Multi-user Interference}} + z^{\mathrm{T}} + \tilde{z}_k, \tag{5.15}$$

where $s_k \sim \mathcal{CN}(0,1)$ is the independent and identically distributed (i.i.d.) information symbol of user k. $\boldsymbol{w}_{b,k} \in \mathbb{C}^{N^{\mathrm{T}} \times 1}$ is the digital beamformer, which is responsible for encoding the k-th user's message signal. We consider the following HBF architectures, i.e.,

- **Fully-connected HBF:** where all antenna elements are connected to all of the RF chains through a network of $N^{\mathrm{T}} M^{\mathrm{T}}$ phase shifters. The analog beamformer at TBS b is denoted by $\boldsymbol{F}_b = [\boldsymbol{f}_{b,1}, \ldots, \boldsymbol{f}_{b,N^{\mathrm{T}}}] \in \mathbb{C}^{M^{\mathrm{T}} \times N^{\mathrm{T}}}$.

- **Partially-connected HBF:** where each RF chain is connected to $M^{\mathrm{T}}/N^{\mathrm{T}}$ sub-array of antennas. The analog beamformer can then be given by $\boldsymbol{F}_b = \mathrm{blkdiag}\left([\boldsymbol{f}_{b,1}, \ldots, \boldsymbol{f}_{b,N^{\mathrm{T}}}]\right) \in \mathbb{C}^{M^{\mathrm{T}} \times N^{\mathrm{T}}}$ with M^{T} phase shifters, where $\boldsymbol{f}_{b,i} \in \mathbb{C}^{\frac{M^{\mathrm{T}}}{N^{\mathrm{T}}} \times 1}$ is the analog beamformer connected to the i-th RF chain via $\frac{M^{\mathrm{T}}}{N^{\mathrm{T}}}$ phase shifters.

Moreover, in (5.15), $\alpha_{b,k}^n$ is the association variable, which equals one if user k associates with TBS b at trajectory point n and zero otherwise. z^{T} is the additive Gaussian noise with zero mean and variance of $\omega^{\mathrm{T}} N_0$, such that ω^{T} is the bandwidth at the THz band and N_0 is the power spectral density of the additive noise. Furthermore, $\tilde{z} \sim \mathcal{CN}(0, \tilde{\sigma}^2)$ denotes the molecular absorption noise, resulting from the re-radiation

of transmitted signals by airborne molecules and is modeled as additive Gaussian noise [125]. The molecular absorption noise can be comparable to both thermal noise and interference, and must be accounted for to avoid overestimating the SINR [109]. Building upon [125, 126], in the considered cooperative multi-antenna network with coherent transmissions, the variance of the molecular absorption noise at user k is given by:

$$\tilde{\sigma}_k^2 = \sum_{i=1}^{K} \sum_{j=1}^{B^{\mathrm{T}}} \sum_{m=1}^{B^{\mathrm{T}}} \tilde{\boldsymbol{h}}_{j,k}^H \bar{\boldsymbol{w}}_{j,i} \bar{\boldsymbol{w}}_{m,i}^H \tilde{\boldsymbol{h}}_{m,k}, \tag{5.16}$$

where $\bar{\boldsymbol{w}}_{j,i} = \boldsymbol{F}_j \boldsymbol{w}_{j,i}$.

Considering the interference and molecular absorption noise, the signal-to-interference-plus-noise ratio (SINR) at the THz band for user k is given by:

$$\gamma_k^{\mathrm{T}} = \frac{\left| \sum_{b=1}^{B^{\mathrm{T}}} \alpha_{b,k}^n \boldsymbol{h}_{b,k}^H \boldsymbol{F}_b \boldsymbol{w}_{b,k} \right|^2}{\sum_{k' \neq k}^{K} \left| \sum_{b=1}^{B^{\mathrm{T}}} \alpha_{b,k'}^n \boldsymbol{h}_{b,k}^H \boldsymbol{F}_b \boldsymbol{w}_{b,k'} \right|^2 + \tilde{\sigma}_k^2 + N_0 \omega^{\mathrm{T}}} \tag{5.17}$$

The received signal at UMB terminal of user k is given by:

$$y_k^{\mathrm{U}} = \underbrace{\sum_{r=1}^{B^{\mathrm{U}}} \beta_{r,k}^n \boldsymbol{g}_{r,k}^H \boldsymbol{Q}_r \boldsymbol{u}_{r,k} \hat{s}_k}_{\text{Desired signals}} + \underbrace{\sum_{k' \neq k}^{K} \sum_{r=1}^{B^{\mathrm{U}}} \beta_{r,k'}^n \boldsymbol{g}_{r,k}^H \boldsymbol{Q}_r \boldsymbol{u}_{r,k'} \hat{s}_{k'}}_{\text{Multi-user Interference}} + z^{\mathrm{U}}, \tag{5.18}$$

where $\boldsymbol{Q}_r = [\boldsymbol{q}_{r,1}, \ldots, \boldsymbol{q}_{r,N^{\mathrm{U}}}] \in \mathbb{C}^{M^{\mathrm{U}} \times N^{\mathrm{U}}}$ denotes the analog beamformer (can be either fully or partially-connected), $\boldsymbol{u}_{r,k} \in \mathbb{C}^{N^{\mathrm{U}} \times 1}$ is the digital beamformer for encoding the message of user k, i.e., $\hat{s}_k \sim \mathcal{CN}(0, 1)$, at the r-th UBS. z^{U} is the additive thermal noise following a Gaussian distribution with zero mean and variance of $N_0 \omega^{\mathrm{U}}$, wherein ω^{U}

is the channel bandwidth of the UMB. The association variable at point n is denoted by $\beta_{r,k}^n$, such that if user k is associated with UBS r, we have $\beta_{r,k}^n = 1$, and $\beta_{r,k}^n = 0$, otherwise.

Based on the received signal in (5.18), the SINR of the UMB at user k is given by:

$$\gamma_k^{\mathrm{U}} = \frac{\left| \sum_{r=1}^{B^{\mathrm{U}}} \beta_{r,k}^n \boldsymbol{g}_{r,k}^H \boldsymbol{Q}_r \boldsymbol{u}_{r,k} \right|^2}{\sum_{k' \neq k}^{K} \left| \sum_{r=1}^{B^{\mathrm{U}}} \beta_{r,k'}^n \boldsymbol{g}_{r,k}^H \boldsymbol{Q}_r \boldsymbol{u}_{r,k'} \right|^2 + N_0 \omega^{\mathrm{U}}}. \tag{5.19}$$

Based on the SINR terms in (5.17) and (5.19), and the dual-connectivity from both THz and UMB [111], the overall achievable rate at user k is obtained by:

$$R_k = \omega^{\mathrm{T}} \log_2(1 + \gamma_k^{\mathrm{T}}) + \omega^{\mathrm{U}} \log_2(1 + \gamma_k^{\mathrm{U}}) \tag{5.20}$$

5.3 Joint User Association and Hybrid Beamforming in Cooperative THz-UMB MBN

In this section, we formulate and solve the resource allocation problem with a focus on maximizing the total network sum-rate when users are stationary.

5.3.1 Problem Formulation

We formulate the downlink[2] sum-rate maximization problem as:

$$\mathcal{P}_1 : \underset{\substack{\alpha^n_{b,k},\boldsymbol{w}_{b,k},\boldsymbol{F}_b \\ \beta^n_{r,k},\boldsymbol{u}_{r,k},\boldsymbol{Q}_r}}{\text{maximize}} \; U_1 = \sum_{k=1}^{K} R_k(\alpha^n_{b,k}, \beta^n_{r,k}, \boldsymbol{w}_{b,k}, \boldsymbol{u}_{b,k}, \boldsymbol{F}_b, \boldsymbol{Q}_r)$$

$$\text{s.t.} \quad \mathbf{C_1} : \sum_{b=1}^{B^{\mathrm{T}}} \alpha^n_{b,k} \leq \mathcal{C}^{\mathrm{T}}_k, \; \forall k, \quad \mathbf{C_2} : \sum_{r=1}^{B^{\mathrm{U}}} \beta^n_{r,k} \leq \mathcal{C}^{\mathrm{U}}_k, \forall k$$

$$\mathbf{C_3} : \sum_{k=1}^{K} \alpha^n_{b,k} \leq N^{\mathrm{T}}, \forall b, \qquad \mathbf{C_4} : \sum_{k=1}^{K} \beta^n_{r,k} \leq N^{\mathrm{U}}, \forall r$$

$$\mathbf{C_5} : \sum_{k=1}^{K} \left| \alpha^n_{b,k} \boldsymbol{F}_b \boldsymbol{w}_{b,k} \right|^2 \leq P^{\mathrm{T}}_{\max}, \forall b,$$

$$\mathbf{C_6} : \sum_{k=1}^{K} \left| \beta^n_{r,k} \boldsymbol{Q}_r \boldsymbol{u}_{r,k} \right|^2 \leq P^{\mathrm{U}}_{\max}, \forall r, \;\; \mathbf{C_7} : \alpha^n_{b,k} \leq \psi_{b,k}, \forall b, k,$$

$$\mathbf{C_8} : |[\boldsymbol{F}_b]_{i,j}| = 1, \forall i, j, b, \;\; \mathbf{C_9} : |[\boldsymbol{Q}_r]_{i,j}| = 1, \forall i, j, r,$$

$$\mathbf{C_{10}} : R_k \geq R^{\mathrm{Th}}_k, \forall k, \;\; \mathbf{C_{11}} : \alpha^n_{b,k}, \beta^n_{r,k} \in \mathbb{B}, \forall b, r, k.$$

In the optimization problem $\mathcal{P}_1$, to reduce the communication overhead, the number of BSs a user can connect to is limited in each frequency band. Hence, $\mathbf{C_1}$ and $\mathbf{C_2}$ denote the total number of BSs (user-centric cluster size) that a user can associate with in the THz ($\mathcal{C}^{\mathrm{T}}_k$) and UMB ($\mathcal{C}^{\mathrm{U}}_k$), respectively. $\mathbf{C_3}$ and $\mathbf{C_4}$ constrain the number of users served by each BS in the respective frequency band to the number of RF chains at the BS. Constraints $\mathbf{C_5}$ and $\mathbf{C_6}$ ensure that the transmit power allocated to all users associated with a given BS does not exceed the power budget, $P^{\mathrm{T}}_{\max}$ for THz

[2]In uplink, channel reciprocity in time division duplexing can be exploited to perform resource allocation by optimizing the users' power allocation and the receive combiners at the CPU, to maximize the uplink sum-rate.

and $P_{\max}^{\mathrm{U}}$ for UMB, respectively. $\mathbf{C_7}$ ensures that a THz link is not assigned to a TBS-user pair if it is blocked. $\mathbf{C_8}$ and $\mathbf{C_9}$ impose the constant modulus constraints on the analog beamformers at TBSs and UBSs, respectively, ensuring that each phase shifter maintains a constant amplitude. Moreover, $\mathbf{C_{10}}$ is the QoS constraint considering a minimum rate of R_k^{Th} for user k, and $\mathbf{C_{11}}$ denotes that the association variables are binary.

The optimization problem in $\mathcal{P}_1$ is mixed-integer nonlinear programming (MINLP), which is non-convex, NP-hard, and cannot be directly solved in polynomial time using traditional methods. The non-convexity of the problem arises from several factors: (i) The objective function is a non-convex function due to the fractional form and multiplication of the optimization variables, (ii) the power budget constraints in $\mathbf{C_5}$ and $\mathbf{C_6}$ are non-convex, (iii) the unit modulus constraints in $\mathbf{C_8}$ and $\mathbf{C_9}$ are non-convex, and (iv) the QoS constraint in $\mathbf{C_{10}}$ is also non-convex. It is worth noting that the QoS constraint can usually be reformulated in a convex format. However, in this problem, due to the different bandwidths in the considered MBN, i.e., ω^{T} and ω^{U}, the constraint cannot be reformulated into a convex function. Also, the binary association variables further render the problem intractable. Therefore, our proposed solution methodology is given as: **(1)** determining feasible analog beamformers to maximize array gain, **(2)** transforming the coupled user association and digital beamforming variables using big-M approach, **(3)** convexifying the SINR terms in the objective and QoS constraints using FP, **(4)** utilizing MM and penalty methods to address the user association variables, and **(5)** proposing an iterative algorithm to solve the joint user association and digital beamforming problem.

5.3.2 Analog Beamforming

By formulating the analog beamformer problem as an array gain maximization problem, which is independent of the digital beamformers and user associations, a sub-optimal solution for the analog beamformers $\boldsymbol{F}_b$ and $\boldsymbol{Q}_r$ can be determined by extending [30] for MBNs. That is, the upper bound of the array gain between TBS b and and user k as well as UBS r and user k can be obtained using Triangle and Cauchy-Schwartz inequality as follows [30]:

$$\begin{aligned}
&\|\boldsymbol{B}_b\boldsymbol{F}_b\boldsymbol{W}_b + \boldsymbol{G}_r\boldsymbol{Q}_r\boldsymbol{U}_r\|_F^2 \leq \|\boldsymbol{B}_b\boldsymbol{F}_b\|_F^2 \|\boldsymbol{W}_b\|_F^2 \\
&+ \|\boldsymbol{G}_r\boldsymbol{Q}_r\|_F^2 \|\boldsymbol{U}_r\|_F^2 + 2\Re(\langle \boldsymbol{B}_b\boldsymbol{F}_b\boldsymbol{W}_b, \boldsymbol{G}_r\boldsymbol{Q}_r\boldsymbol{U}_r\rangle) \\
&= \left(\left|\boldsymbol{b}_{b,k}^H \boldsymbol{f}_{b,k}\right|^2 + \|(\boldsymbol{B}_b)_{-k,:}(\boldsymbol{F}_b)_{:,-k}\|_F^2\right) \|\boldsymbol{W}_b\|_F^2 \\
&+ \left(\left|\boldsymbol{g}_{r,k}^H \boldsymbol{q}_{r,k}\right|^2 + \|(\boldsymbol{G}_r)_{-k,:}(\boldsymbol{Q}_r)_{:,-k}\|_F^2\right) \|\boldsymbol{U}_r\|_F^2 ,
\end{aligned} \tag{5.21}$$

where $\boldsymbol{B}_b = [\boldsymbol{b}_{b,1}, \ldots, \boldsymbol{b}_{b,K}]^H$ (i.e., $\boldsymbol{b}_{b,k} = \boldsymbol{a}_{b,k}$ for $v_{b,k} = 0$), $\boldsymbol{W}_b = [\alpha_{b,1}^n \boldsymbol{w}_{b,1}, \ldots, \alpha_{b,K}^n \boldsymbol{w}_{b,K}]$, $\boldsymbol{G}_r = [\boldsymbol{g}_{r,1}, \ldots, \boldsymbol{g}_{r,K}]^H$, $\boldsymbol{U}_r = [\beta_{r,1}^n \boldsymbol{u}_{r,1}, \ldots, \beta_{r,K}^n \boldsymbol{u}_{r,K}]$. $(\boldsymbol{X})_{-k,:}$ and $(\boldsymbol{X})_{:,-k}$ denote the sub-matrices of $\boldsymbol{X}$ after removing k-th row and k-th column, respectively. Also, it is noteworthy that the inner product $\langle \boldsymbol{B}_b\boldsymbol{F}_b\boldsymbol{W}_b, \boldsymbol{G}_r\boldsymbol{Q}_r\boldsymbol{U}_r\rangle = 0$ due to the orthogonality of THz and UMB frequencies.

Given $\boldsymbol{W}_b$ and $\boldsymbol{U}_r$, the analog beamforming for each pair of user and BS, which is

feasible and maximizes the array gain[3], is obtained by solving the following problem:

$$\begin{aligned} &\underset{\boldsymbol{f}_{b,k},\boldsymbol{q}_{r,k}}{\text{maximize}} \quad \left|\boldsymbol{b}_{b,k}^H \boldsymbol{f}_{b,k}\right|^2 + \left|\boldsymbol{g}_{r,k}^H \boldsymbol{q}_{r,k}\right|^2 \\ &\text{s.t.} \quad \mathbf{C_5}, \mathbf{C_6}, \mathbf{C_8}, \mathbf{C_9} \end{aligned} \tag{5.22}$$

Solving (5.22) requires the number of RF chains to be equal to the number of users, that is $N^{\mathrm{T}} = N^{\mathrm{U}} = K$. Hence, constraints $\mathbf{C_3}$ and $\mathbf{C_4}$ are inherently satisfied. For fully-connected HBF, the solution to (5.22) is given by:

$$\bar{\boldsymbol{F}}_b = \boldsymbol{B}_b^{\mathbb{T}} \tag{5.23}$$

$$\bar{\boldsymbol{Q}}_r = e^{j\angle\, \boldsymbol{G}_r^{\mathbb{T}}} \tag{5.24}$$

For the case of partially-connected, we have:

$$\bar{\boldsymbol{F}}_b = (\boldsymbol{I}_K \otimes \mathbf{1}_{\frac{M^{\mathrm{T}}}{K}}) \odot \boldsymbol{B}_b^{\mathbb{T}}, \tag{5.25}$$

$$\bar{\boldsymbol{Q}}_r = (\boldsymbol{I}_K \otimes \mathbf{1}_{\frac{M^{\mathrm{U}}}{K}}) \odot e^{j\angle\, \boldsymbol{G}_r^{\mathbb{T}}}. \tag{5.26}$$

The analog beamformers obtained above also conform to the unit modulus constraints, thus $\mathbf{C_8}$ and $\mathbf{C_9}$ are satisfied. In order to prove the feasibility of $\mathbf{C_5}$ and $\mathbf{C_6}$, we first reformulate the constraints as follows:

$$\begin{aligned} \mathbf{C_5} : \|\boldsymbol{F}_b \boldsymbol{W}_b\|_F^2 &= \mathrm{tr}(\boldsymbol{W}_b^H \boldsymbol{F}_b^H \boldsymbol{F}_b \boldsymbol{W}_b) \\ &= \mathrm{tr}(\boldsymbol{F}_b^H \boldsymbol{F}_b \boldsymbol{W}_b \boldsymbol{W}_b^H) \leq P_{\max}^{\mathrm{T}}, \end{aligned} \tag{5.27}$$

[3] $\|(\boldsymbol{B}_b)_{-k,:}(\boldsymbol{F}_b)_{:,-k}\|_F^2$ and $\|(\boldsymbol{G}_r)_{-k,:}(\boldsymbol{Q}_r)_{:,-k}\|_F^2$ in (5.21) represent the interference which need not be considered while maximizing the array gain. However, once the feasible analog beamformers are determined, the interference will be considered in the optimization.

$$\begin{aligned}\mathbf{C_6}: \|\boldsymbol{Q}_r\boldsymbol{U}_r\|_F^2 &= \mathrm{tr}(\boldsymbol{U}_r^H\boldsymbol{Q}_r^H\boldsymbol{Q}_r\boldsymbol{U}_r) \\ &= \mathrm{tr}(\boldsymbol{Q}_r^H\boldsymbol{Q}_r\boldsymbol{U}_r\boldsymbol{U}_r^H) \leq P_{\max}^{\mathrm{U}},\end{aligned} \tag{5.28}$$

where both (5.27) and (5.28) utilize cyclic property of trace of matrices. Then, we depict the feasibility of the analog beamformers for $\mathbf{C_5}$ and $\mathbf{C_6}$ as shown below.

Remark 5.1. *In a hybrid beamformer with M antenna elements and N RF chains, for large values of M, the analog beamformer $\boldsymbol{F}$ satisfies $\boldsymbol{F}^H\boldsymbol{F} \approx M\boldsymbol{I}_N$ in a fully-connected architecture, and $\boldsymbol{F}^H\boldsymbol{F} = \frac{M}{N}\boldsymbol{I}_N$ in a partially-connected architecture [127].*

Based on the result of **Remark 1**, the power budget constraints are reformulated as:

$$\mathbf{C_5}: \|\boldsymbol{W}_b\|_F^2 \leq \bar{P}_{\max}^{\mathrm{T}}, \;\; \mathbf{C_6}: \|\boldsymbol{U}_r\|_F^2 \leq \bar{P}_{\max}^{\mathrm{U}}, \tag{5.29}$$

where $\bar{P}_{\max}^{\mathrm{T}} = \frac{P_{\max}^{\mathrm{T}}}{M^{\mathrm{T}}}$ and $\bar{P}_{\max}^{\mathrm{U}} = \frac{P_{\max}^{\mathrm{U}}}{M^{\mathrm{U}}}$ for the case of fully-connected HBF, and $\bar{P}_{\max}^{\mathrm{T}} = \frac{P_{\max}^{\mathrm{T}}K}{M^{\mathrm{T}}}$ and $\bar{P}_{\max}^{\mathrm{U}} = \frac{P_{\max}^{\mathrm{U}}K}{M^{\mathrm{U}}}$ for partially-connected HBF. Therefore, the obtained analog beamformers are feasible for $\mathcal{P}_1$.

5.3.3 Problem Transformation

After obtaining the analog beamformers $\bar{\boldsymbol{F}}_b$ and $\bar{\boldsymbol{Q}}_r$, we first address the nonlinearity of the multiplication of the binary variables and beamformers in the objective function and the power budget constraints in $\mathbf{C_5}$ and $\mathbf{C_6}$, and QoS in $\mathbf{C_{10}}$. It is important to note that multiplying the binary variables by the digital beamformers results in a zero vector if a user is not associated with a given BS, making the constraints non-convex. To tackle this, we utilize the *big*-M method [128], which allows us to decouple the association variables from the SINR terms and the power budget constraints. First,

we introduce variables $p_{b,k} = |\boldsymbol{w}_{b,k}|^2$ and $\delta_{r,k} = |\boldsymbol{u}_{r,k}|^2$, denoting the power of the digital beamformers for each pair of user-BS at TBSs and UBSs, respectively. The following constraints guarantee that the beamformers remain non-zero whenever the association variables are non-zero, and zero, otherwise.

$$|\boldsymbol{w}_{b,k}|^2 \leq \alpha_{b,k}^n p_{b,k}, \tag{5.30a}$$

$$|\boldsymbol{u}_{r,k}|^2 \leq \beta_{r,k}^n \delta_{r,k}. \tag{5.30b}$$

However, both constraints are still non-convex. To tackle this, the right-hand side of (5.30a) and (5.30b) are expanded as follows:

$$|2\boldsymbol{w}_{b,k}|^2 \leq \left(\alpha_{b,k}^n + p_{b,k}\right)^2 - \left(\alpha_{b,k}^n - p_{b,k}\right)^2 \tag{5.31a}$$

$$|2\boldsymbol{u}_{r,k}|^2 \leq \left(\beta_{r,k}^n + \delta_{r,k}\right)^2 - \left(\beta_{r,k}^n - \delta_{r,k}\right)^2. \tag{5.31b}$$

After some manipulations, the above constraints can be written in the following convex formulation:

$$\left|[2\boldsymbol{w}_{b,k}^H, \alpha_{b,k}^n - p_{b,k}]\right| \leq \alpha_{b,k}^n + p_{b,k}, \tag{5.32a}$$

$$\left|[2\boldsymbol{u}_{r,k}^H, \beta_{r,k}^n - \delta_{r,k}]\right| \leq \beta_{r,k}^n + \delta_{r,k}, \tag{5.32b}$$

Moreover, the power budget constraints in $\mathbf{C_5}$ and $\mathbf{C_6}$ are reformulated as follows:

$$\mathbf{C_{5.2}} : \sum_{k=1}^{K} p_{b,k} \leq \bar{P}_{\max}^{\mathrm{T}}, \forall b, \quad \mathbf{C_{5.3}} : p_{b,k} \leq \alpha_{b,k}^n \bar{P}_{\max}^{\mathrm{T}}, \forall b, k$$

$$\mathbf{C_{6.2}} : \sum_{k=1}^{K} \delta_{r,k} \leq \bar{P}_{\max}^{\mathrm{U}}, \forall r, \quad \mathbf{C_{6.3}} : \delta_{r,k} \leq \beta_{r,k}^n \bar{P}_{\max}^{\mathrm{U}}, \forall r, k.$$

After the above-mentioned transformations and decoupling of the binary variables from the digital beamformers, the functions in U_1 and QoS constraints in $\mathbf{C_{10}}$ also reduce to $R_k(\boldsymbol{w}_{b,k}, \boldsymbol{u}_{b,k}, \bar{\boldsymbol{F}}_b, \bar{\boldsymbol{Q}}_r)$. For given analog beamformers and after decoupling the association variables and digital beamformers, $\mathcal{P}_1$ can be reformulated as follows:

$$\begin{aligned} \mathcal{P}_2 : \underset{\substack{\alpha^n_{b,k}, \boldsymbol{w}_{b,k}, p_{b,k} \\ \beta^n_{r,k}, \boldsymbol{u}_{r,k}, \delta_{r,k}}}{\text{maximize}} \quad & U_1 \\ \text{s.t.} \quad & \mathbf{C_{5.1}} : (5.32a), \mathbf{C_{5.2}}, \mathbf{C_{5.3}}, \mathbf{C_{6.1}} : (5.32b), \mathbf{C_{6.2}}, \mathbf{C_{6.3}}, \\ & \mathbf{C_1}, \mathbf{C_2}, \mathbf{C_7}, \mathbf{C_{10}}, \mathbf{C_{11}}. \end{aligned} \tag{5.33}$$

Problem $\mathcal{P}_2$ is still a MINLP due to the non-convexity of the objective function, QoS constraint, and the binary variables.

5.3.4 Joint User Association and Beamforming

5.3.4.1 FP-based Approach

First, we develop an equivalent transformation of the general FP problems with fractional objective function and constraints. This transformation leverages on quadratic transformation method introduced in [65, 66]. The transformation is presented in the following Lemma.

Lemma 5.1. *Given a sequence of logarithm functions of ratios* $\frac{A_{i,j}(\boldsymbol{X})}{B_{i,j}(\boldsymbol{X})}$*, for* $i \in \{1, \ldots, I\}$ *and* $j \in \{1, \ldots, J\}$*, and* $\mathcal{X}$ *being the convex feasible set of variable* $\boldsymbol{X}$*,*

the maximization problem subject to the QoS constraint formulated as:

$$\begin{aligned} &\underset{\boldsymbol{X}}{maximize} \quad \sum_i \sum_j \log\left(\frac{A_{i,j}(\boldsymbol{X})}{B_{i,j}(\boldsymbol{X})}\right) \\ &s.t. \quad \sum_j \log\left(\frac{A_{i,j}(\boldsymbol{X})}{B_{i,j}(\boldsymbol{X})}\right) \geq R_i^{Th}, \ \boldsymbol{X} \in \mathcal{X} \end{aligned} \tag{5.34}$$

is equivalent to

$$\begin{aligned} &\underset{\boldsymbol{X},\boldsymbol{y}}{maximize} \sum_i \sum_j \log\left(2\Re\left\{y_{i,j}^* \sqrt{A_{i,j}(\boldsymbol{X})}\right\} - |y_{i,j}|^2 B_{i,j}(\boldsymbol{X})\right) \\ &s.t. \sum_j \log\left(2\Re\left\{y_{i,j}^* \sqrt{A_{i,j}(\boldsymbol{X})}\right\} - |y_{i,j}|^2 B_{i,j}(\boldsymbol{X})\right) \geq R_i^{Th}, \forall i, \end{aligned} \tag{5.35}$$

where the optimal $\boldsymbol{y}$ *is* $y_{i,j}^o = \frac{\sqrt{A_{i,j}(\boldsymbol{X})}}{B_{i,j}(\boldsymbol{X})}$.

Proof. See Section 5.6.1. □

Based on the transformation in **Lemma 5.1**, for fixed auxiliary variable $\mu_k \in \mathbb{C}$, the THz SINR expression at iteration i can be formulated in the following quadratic format:

$$\gamma_k^{\mathrm{T},(i)} = 2\Re\left\{\mu_k^* \sum_{b=1}^{B^{\mathrm{T}}} \boldsymbol{w}_{b,k}^H \bar{\boldsymbol{F}}_b^H \boldsymbol{h}_{b,k}\right\} - |\mu_k|^2 \left(\sum_{k' \neq k}^{K} \left|\sum_{b=1}^{B^{\mathrm{T}}} \boldsymbol{h}_{b,k}^H \bar{\boldsymbol{F}}_b \boldsymbol{w}_{b,k'}\right|^2 + \tilde{\sigma}_k^2 + N_0 \omega^{\mathrm{T}}\right) \tag{5.36}$$

where the respective auxiliary variable can be obtained as:

$$\mu_k = \frac{\sum_{b=1}^{B^{\mathrm{T}}} \boldsymbol{w}_{b,k}^H \bar{\boldsymbol{F}}_b^H \boldsymbol{h}_{b,k}}{\sum_{k' \neq k}^{K} \left|\sum_{b=1}^{B^{\mathrm{T}}} \boldsymbol{h}_{b,k}^H \bar{\boldsymbol{F}}_b \boldsymbol{w}_{b,k'}\right|^2 + \tilde{\sigma}_k^2 + N_0 \omega^{\mathrm{T}}}. \tag{5.37}$$

By introducing the quadratic auxiliary variable $\zeta_k \in \mathbb{C}$, the quadratic format of the SINR terms in the UMB at iteration i is derived as:

$$\gamma_k^{\mathrm{U},(i)} = 2\Re\left\{\zeta_k^* \sum_{r=1}^{B^{\mathrm{U}}} \boldsymbol{u}_{r,k}^H \bar{\boldsymbol{Q}}_r^H \boldsymbol{g}_{r,k}\right\} - |\zeta_k|^2 \left(\sum_{k'\neq k}^{K} \left|\sum_{r=1}^{B^{\mathrm{U}}} \boldsymbol{g}_{r,k}^H \bar{\boldsymbol{Q}}_r \boldsymbol{u}_{r,k'}\right|^2 + N_0\omega^{\mathrm{U}}\right), \quad (5.38)$$

and the variable ζ_k is given by:

$$\zeta_k = \frac{\sum_{r=1}^{B^{\mathrm{U}}} \boldsymbol{u}_{r,k}^H \bar{\boldsymbol{Q}}_r^H \boldsymbol{g}_{r,k}}{\sum_{k'\neq k}^{K} \left|\sum_{r=1}^{B^{\mathrm{U}}} \boldsymbol{g}_{r,k}^H \bar{\boldsymbol{Q}}_r \boldsymbol{u}_{r,k'}\right|^2 + N_0\omega^{\mathrm{U}}}. \quad (5.39)$$

Replacing the non-convex SINR terms in the objective function and the QoS constraints by the derived quadratic format, problem $\mathcal{P}_2$ reduces to the following convex-binary optimization problem $\mathcal{P}_3$ at i-th iteration.

$$\mathcal{P}_3 : \underset{\substack{\alpha_{b,k}^n, \boldsymbol{w}_{b,k}, p_{b,k} \\ \beta_{r,k}^n, \boldsymbol{u}_{r,k}, \delta_{r,k}}}{\text{maximize}} \sum_{k=1}^{K} \omega^{\mathrm{T}} \log(1+\gamma_k^{\mathrm{T},(i)}) + \omega^{\mathrm{U}} \log(1+\gamma_k^{\mathrm{U},(i)})$$

$$\text{s.t.} \mathbf{C_{10}} : \omega^{\mathrm{T}} \log(1+\gamma_k^{\mathrm{T},(i)}) + \omega^{\mathrm{U}} \log(1+\gamma_k^{\mathrm{U},(i)}) \geq R_k^{\mathrm{Th}}, \forall k,$$

$$\mathbf{C_1}, \mathbf{C_2}, \mathbf{C_{5.1}}, \mathbf{C_{5.2}}, \mathbf{C_{5.3}}, \mathbf{C_{6.1}}, \mathbf{C_{6.2}}, \mathbf{C_{6.3}}, \mathbf{C_7}, \mathbf{C_{11}}.$$

A globally optimal solution of $\mathcal{P}_3$ can be obtained by utilizing standard algorithms, such as CVX [129] combined with MOSEK solver [86]. To deal with the binary variables in $\mathcal{P}_3$, Mosek employs branch-and-bound method in an iterative manner. Although a globally optimal solution to $\mathcal{P}_3$ can be achieved, in the worst case, the computational complexity of this method can grow exponentially with a large number of BSs and users in the network.

5.3.4.2 Majorization-minimization-based Approach

In order to reduce the computational complexity that arises from the binary variables, $\alpha_{b,k}^n$ and $\beta_{r,k}^n$, we seek an MM-based method [130, 131] and penalty approach. First, by introducing the following constraints replacing $\mathbf{C_{11}}$, the association variables can only take binary values:

$$\mathbf{C_{11.1}}: 0 \leq \alpha_{b,k}^n \leq 1, \forall b,k, \mathbf{C_{11.2}}: \sum_{b=1}^{B^{\mathrm{T}}} \sum_{k=1}^{K} \left(\alpha_{b,k}^n - {\alpha_{b,k}^n}^2\right) \leq 0,$$

$$\mathbf{C_{11.3}}: 0 \leq \beta_{r,k}^n \leq 1, \forall r,k, \mathbf{C_{11.4}}: \sum_{r=1}^{B^{\mathrm{U}}} \sum_{k=1}^{K} \left(\beta_{r,k}^n - {\beta_{r,k}^n}^2\right) \leq 0,$$

where $\mathbf{C_{11.2}}$ and $\mathbf{C_{11.4}}$ are non-convex. To overcome this, we use the penalty method by penalizing the objective function by adding these two sets of constraints with penalty factors of $\Gamma_1 \gg 0$ and $\Gamma_2 \gg 0$ for $\mathbf{C_{11.2}}$ and $\mathbf{C_{11.4}}$, respectively. Therefore, problem $\mathcal{P}_3$ can be reformulated into $\mathcal{P}_4$ as follows:

$$\begin{aligned}
\mathcal{P}_4: \quad & \underset{\substack{\alpha_{b,k}^n, \boldsymbol{w}_{b,k}, p_{b,k} \\ \beta_{r,k}^n, \boldsymbol{u}_{r,k}, \delta_{r,k}}}{\text{maximize}} \quad U_1 - \Gamma_1 \mathcal{Z}_1 - \Gamma_2 \mathcal{Z}_2 \\
\text{s.t.} \quad & \mathbf{C_1}, \mathbf{C_2}, \mathbf{C_{5.1}}, \mathbf{C_{5.2}}, \mathbf{C_{5.3}}, \mathbf{C_{6.1}}, \mathbf{C_{6.2}}, \mathbf{C_{6.3}}, \\
& \mathbf{C_7}, \mathbf{C_{10}}, \mathbf{C_{11.1}}, \mathbf{C_{11.3}},
\end{aligned}$$

where $\mathcal{Z}_1 = \sum_{b=1}^{B_T} \sum_{k=1}^{K} \alpha_{b,k}^n - {\alpha_{b,k}^n}^2$ and $\mathcal{Z}_2 = \sum_{r=1}^{B_R} \sum_{k=1}^{K} \beta_{r,k}^n - {\beta_{r,k}^n}^2$. The objective function in $\mathcal{P}_4$ is still non-convex due to adding the penalty functions. At this point, the MM approach is utilized, where the second quadratic term of the penalty function is

replaced by its first-order Taylor approximation.

$$\bar{\boldsymbol{f}}^{(i)}(x) \triangleq \boldsymbol{f}(x^{(i-1)}) + \nabla^{\mathbb{T}}\boldsymbol{f}\left(x^{(i-1)}\right)\left(x - x^{(i-1)}\right), \tag{5.40}$$

where $x = \alpha_{b,k}^n$ for the THz association variables and $x = \beta_{r,k}^n$ for the UMB association variables. Thus, $\mathcal{P}_5$ can be approximated by the following problem at iteration i.

$$\begin{aligned}
\mathcal{P}_5 : \underset{\substack{\alpha_{b,k}^n, \boldsymbol{w}_{b,k}, p_{b,k} \\ \beta_{r,k}^n, \boldsymbol{u}_{r,k}, \delta_{r,k}}}{\text{maximize}} \quad & U_1^{(i)} - \Gamma_1 \mathcal{L}_1^{(i)} - \Gamma_2 \mathcal{L}_2^{(i)} \\
\text{s.t.} \quad & \mathbf{C_1}, \mathbf{C_2}, \mathbf{C_{5.1}}, \mathbf{C_{5.2}}, \mathbf{C_{5.3}}, \mathbf{C_{6.1}}, \mathbf{C_{6.2}}, \mathbf{C_{6.3}}, \\
& \mathbf{C_7}, \mathbf{C_{10}}, \mathbf{C_{11.1}}, \mathbf{C_{11.3}}.
\end{aligned} \tag{5.41}$$

where

$$\mathcal{L}_1^{(i)} = \sum_{b=1}^{B^{\mathrm{T}}} \sum_{k=1}^{K} \left(\alpha_{b,k}^n \left(1 - 2\alpha_{b,k}^{n,(i-1)}\right) + \left(\alpha_{b,k}^{n,(i-1)}\right)^2 \right),$$

$$\mathcal{L}_2^{(i)} = \sum_{r=1}^{B^{\mathrm{U}}} \sum_{k=1}^{K} \left(\beta_{r,k}^n \left(1 - 2\beta_{r,k}^{n,(i-1)}\right) + \left(\beta_{r,k}^{n,(i-1)}\right)^2 \right),$$

and superscript $(i-1)$ denotes the value of the respective variable at the previous iteration. Problem $\mathcal{P}_5$ is a convex optimization that can be solved using CVX [129]. The proposed solution for solving $\mathcal{P}_1$ is summarized in **Algorithm 1**.

5.3.5 Convergence and Complexity Analysis of Algorithm 1

5.3.5.1 Convergence

To show the convergence of the iterative part of **Algorithm 1**, from Step 7 to Step 19, we first denote the variables of $\mathcal{P}_4$ by $\boldsymbol{x} = \boldsymbol{w}_{b,k}, \boldsymbol{u}_{r,k}, \forall b, r, k$, and $\boldsymbol{\alpha} = \alpha_{b,k}^n, \beta_{r,k}^n, \forall b, r, k$. Let $\mathcal{X}$ denotes the feasible set, and define the objective function as $f(\boldsymbol{x}, \boldsymbol{\alpha}) = U_1(\boldsymbol{x}) +$

Algorithm 4 Joint User Association and Hybrid Beamforming Solution to $\mathcal{P}_1$ and $\hat{\mathcal{P}}_1$

Input: Index of trajectory point, n, maximum number of iterations $L_{\max}$, stopping accuracy ϵ_1 and ϵ_2, penalty factors, Γ_1 and Γ_2. Initialize $\boldsymbol{w}_{b,k}, \boldsymbol{u}_{r,k}, \alpha_{b,k}^n, \beta_{r,k}^n, \gamma_k^{\mathrm{T},(0)}, \gamma_k^{\mathrm{U},(0)}$ $\forall b, r, k$ with feasible values.

Obtain the analog beamformers $\bar{\boldsymbol{F}}_b$, $\forall b$ and $\bar{\boldsymbol{Q}}_r$, $\forall r$
if Fully-connected HBF **then**
    Use (5.23) and (5.24)
else
    Use (5.25) and (5.26)
end if
for $i = 0, 1, ..., L_{\max}$ **do**
    Update $\mu_k^{(i)} \forall k$ using (5.37)
    Update $\zeta_k^{(i)} \forall k$ using (5.39)
    if Fully-connected HBF **then**
        Use $\bar{P}_{\max}^{\mathrm{T}} = \frac{P_{\max}^{\mathrm{T}}}{M^{\mathrm{T}}}$ and $\bar{P}_{\max}^{\mathrm{U}} = \frac{P_{\max}^{\mathrm{U}}}{M^{\mathrm{U}}}$
    else
        Use $\bar{P}_{\max}^{\mathrm{T}} = \frac{P_{\max}^{\mathrm{T}} K}{M^{\mathrm{T}}}$ and $\bar{P}_{\max}^{\mathrm{U}} = \frac{P_{\max}^{\mathrm{U}} K}{M^{\mathrm{U}}}$
    end if
    Solve problem $\mathcal{P}_5$ and obtain $\boldsymbol{w}_{b,k}^{(i)}$, $\boldsymbol{u}_{r,k}^{(i)}$, $\alpha_{b,k}^{n,(i)}$, $\beta_{r,k}^{n,(i)}$, $\forall b, r, k$
    Update $\gamma_k^{\mathrm{T},(i)}$ using (5.36)
    Update $\gamma_k^{\mathrm{U},(i)}$ using (5.38)
    Until $|U_1^{(i)} - U_1^{(i-1)}| < \epsilon_1$ & $|(\mathcal{L}_1^{(i)} + \mathcal{L}_2^{(i)}) - (\mathcal{L}_1^{(i-1)} + \mathcal{L}_2^{(i-1)})| < \epsilon_2$
end for

Output: The optimal resource allocation at point n: $\boldsymbol{w}_{b,k}^o, \boldsymbol{F}_b^o, \boldsymbol{u}_{r,k}^o, \boldsymbol{Q}_r^o, \alpha_{b,k}^{n,o}, \beta_{r,k}^{n,o}$, $\forall b, r, k$.

$P(\boldsymbol{\alpha})$, where U_1 is the sum-rate before applying the quadratic transformation, and $P(\boldsymbol{\alpha}) = -\Gamma_1 \mathcal{Z}_1 - \Gamma_2 \mathcal{Z}_2 \leq 0$, for all $\boldsymbol{\alpha} \in [0, 1]$. At the i-th iteration, the sum-rate is lower-bounded by $U_1^{(i)}(\boldsymbol{x}) \leq U_1(\boldsymbol{x})$ via the quadratic transformation (see (5.57)), such that $U_1^{(i)}(\boldsymbol{x}^{(i-1)}) = U_1(\boldsymbol{x}^{(i-1)})$. The penalty term, being a convex function, is lower-bounded by its first-order Taylor expansion as $P^{(i)}(\boldsymbol{\alpha}) \leq P(\boldsymbol{\alpha}) \leq 0$, where $P^{(i)}(\boldsymbol{\alpha}^{(i-1)}) = P(\boldsymbol{\alpha}^{(i-1)})$. Accordingly, the surrogate objective function at iteration i is defined as: $f^{(i)}(\boldsymbol{x}, \boldsymbol{\alpha}) = U_1^{(i)}(\boldsymbol{x}) + P^{(i)}(\boldsymbol{\alpha})$, which forms the objective of $\mathcal{P}_5$.

The solution at the i-th iteration is given by: $\boldsymbol{x}^{(i)}, \boldsymbol{\alpha}^{(i)} = \arg\max_{\boldsymbol{x},\boldsymbol{\alpha}\in\mathcal{X}} f^{(i)}(\boldsymbol{x}, \boldsymbol{\alpha})$. Since $\mathcal{P}_5$ is a convex optimization problem, it satisfies the inequality:

$$f(\boldsymbol{x}^{(i)}, \boldsymbol{\alpha}^{(i)}) \geq f^{(i)}(\boldsymbol{x}^{(i)}, \boldsymbol{\alpha}^{(i)}) \geq f^{(i)}(\boldsymbol{x}^{(i-1)}, \boldsymbol{\alpha}^{(i-1)}) = f(\boldsymbol{x}^{(i-1)}, \boldsymbol{\alpha}^{(i-1)}), \tag{5.43}$$

which shows that the objective of $\mathcal{P}_4$ increases monotonically with iterations. Furthermore, since all variables are bounded due to power and QoS constraints, the sequence $f(\boldsymbol{x}^{(i)}, \boldsymbol{\alpha}^{(i)})$ is finite. Also, as $P(\boldsymbol{\alpha}^{(i)}) \leq 0$, the objective function of $\mathcal{P}_4$ is bounded above by the sum-rate: $f(\boldsymbol{x}^{(i)}, \boldsymbol{\alpha}^{(i)}) \leq U_1(\boldsymbol{x}^{(i)})$. If sufficiently large penalty factors Γ_1 and Γ_2 are chosen, then $P(\boldsymbol{\alpha}^{(i)}) \to 0$, and equality holds. Consequently, $U_1(\boldsymbol{x}^{(i)})$ converges to a sub-optimal point of $\mathcal{P}_2$.

5.3.5.2 Complexity Analysis

The worst-case complexity of **Algorithm 1** is obtained as follows. In the first stage of the algorithm, the complexity is dominated by the analog beamforming derivations. If fully-connected HBF is selected, the complexity is $\mathcal{O}((B^{\mathrm{T}}M^{\mathrm{T}} + B^{\mathrm{U}}M^{\mathrm{U}})K)$, and if partially connected HBF is chosen, the complexity reduces to $\mathcal{O}(B^{\mathrm{T}}M^{\mathrm{T}} + B^{\mathrm{U}}M^{\mathrm{U}})$. For the iterative part of the algorithm, since only one convex optimization problem is solved in each iteration, the algorithm exhibits polynomial computational complexity [132]. Steps 8, 9, 16, and 17 have a combined complexity of $\mathcal{O}(4K)$, which is dominated by the complexity of Step 15. Complexity of solving the convex optimization in step 15 is $\mathcal{O}(((K^2 + 2K)(B^{\mathrm{T}} + B^{\mathrm{U}}))^3)$, which is of cubic order in terms of

the number of variables [133].[4] Therefore, the overall complexity of **Algorithm 1** is $\mathcal{O}((B^{\mathrm{T}}M^{\mathrm{T}} + B^{\mathrm{U}}M^{\mathrm{U}})K + I_A(((K^2+2K)(B^{\mathrm{T}}+B^{\mathrm{U}}))^3)$ for fully-connected HBF, and $\mathcal{O}(B^{\mathrm{T}}M^{\mathrm{T}} + B^{\mathrm{U}}M^{\mathrm{U}} + I_A(((K^2+2K)(B^{\mathrm{T}}+B^{\mathrm{U}}))^3)$ for partially-connected HBF, where I_A is the number of iterations required for Steps 7–19 to converge.

5.4 Mobility-aware User Association and Hybrid Beamforming in Cooperative THz/UMB MBN

In this section, we address mobility-aware resource allocation in a cooperative MBN. In traditional single-band multi-BS networks, HO occurs whenever a user switches from one BS to another. Each HO reduces effective data rate by decreasing the time available for data transmission. In cooperative networks, the number of HOs can be higher since users have more BSs to associate with simultaneously. However, the data rate reduction may not necessarily happen due to cooperation gain of the BSs' cluster. In this section, we develop two approaches to handle HO-aware resource allocation in order to mitigate the impact of HOs on cooperative MBN.

First, we define the HO-aware effective sum-rate, which serves as the key performance metric that accounts for mobility and HO cost. We assume that a HO occurs at trajectory point n whenever a user associates with a BS at point n that it was not associated with at the previous trajectory point $n-1$. For the k-th user, the total number of HOs at point n is given by $\varsigma_k^{\mathrm{T}} = \sum_{b=1}^{B^{\mathrm{T}}} \left(1-\alpha_{b,k}^{n-1}\right)\alpha_{b,k}^{n}$ for the THz band, and

[4]The MOSEK implementation is based on the homogeneous model, the Nesterov–Todd search direction, and a Mehrotra-like predictor-corrector algorithm [134]. These features enhance numerical robustness and practical efficiency. Although interior-point methods have cubic theoretical complexity, MOSEK leverages sparsity and structure to reduce this complexity. Based on MOSEK's FLOP logs for our problem, the total computational complexity scales approximately as $\mathcal{O}(((K^2+2K)(B^{\mathrm{T}}+B^{\mathrm{U}}))^2)$, which is the square of the number of variables.

by $\varsigma_k^{\mathrm{U}} = \sum_{r=1}^{B^{\mathrm{U}}} \left(1 - \beta_{r,k}^{n-1}\right) \beta_{r,k}^{n}$ for the UMB band. Each HO results in a transmission interruption by reducing the available time for data transmission, which in turn decreases the effective rate at point n [38, 40, 42, 135]. This interruption is modeled by cost factors $\eta^{\mathrm{T}} \in (0,1)$ and $\eta^{\mathrm{U}} \in (0,1)$ for the THz and UMB bands, respectively. These cost factors represent the fraction of transmission time lost due to HO delays, and their values may vary depending on the specific HO procedures implemented by the operator [42]. Therefore, the remaining time available for transmission over THz and UMB frequency bands is $1-\eta^{\mathrm{T}}\varsigma_k^{\mathrm{T}}$ and $1-\eta^{\mathrm{U}}\varsigma_k^{\mathrm{U}}$, respectively. Based on this, the effective rate of user k at point n is given by $\bar{R}_k = \bar{R}_k^{\mathrm{T}} + \bar{R}_k^{\mathrm{U}}$, where

$$\bar{R}_k^{\mathrm{T}} = \left[\left(1 - \eta^{\mathrm{T}} \sum_{b=1}^{B^{\mathrm{T}}} \left(1 - \alpha_{b,k}^{n-1}\right) \alpha_{b,k}^{n}\right) \omega^{\mathrm{T}} \log_2(1+\gamma_k^{\mathrm{T}})\right]^{+} \tag{5.44}$$

and

$$\bar{R}_k^{\mathrm{U}} = \left[\left(1 - \eta^{\mathrm{U}} \sum_{r=1}^{B^{\mathrm{U}}} \left(1 - \beta_{r,k}^{n-1}\right) \beta_{r,k}^{n}\right) \omega^{\mathrm{U}} \log_2(1+\gamma_k^{\mathrm{U}})\right]^{+}, \tag{5.45}$$

where $[x]^{+} = \max\{0, x\}$. The HO-aware system sum-rate is then formulated as follows:

$$U_2 = \sum_{k=1}^{K} \bar{R}_k = \sum_{k=1}^{K} \bar{R}_k^{\mathrm{T}} + \bar{R}_k^{\mathrm{U}}. \tag{5.46}$$

5.4.1 HO Cost-based Approach

In this approach, we incorporate the direct impact of HO on the system sum-rate by setting the HO-aware sum-rate as the objective function. Hence, the HO cost-based

resource allocation problem is formulated as follows:

$$\hat{\mathcal{P}}_1 : \underset{\substack{\alpha^n_{b,k},\boldsymbol{w}_{b,k},\boldsymbol{F}_b \\ \beta^n_{r,k},\boldsymbol{u}_{r,k},\boldsymbol{Q}_r}}{\text{maximize}} \; U_2 \tag{5.47}$$

$$\text{s.t.} \quad \mathbf{C_1}, \mathbf{C_2}, \mathbf{C_3}, \mathbf{C_4}, \mathbf{C_5}, \mathbf{C_6}, \mathbf{C_7}, \mathbf{C_8}, \mathbf{C_9}, \bar{\mathbf{C}}_{\mathbf{10}} : \bar{R}_k \geq R_k^{\text{Th}},$$

$$\mathbf{C_{11}}, \mathbf{C_{12}} : \sum_{b=1}^{B^{\text{T}}} \alpha_{b,k}^{n-1} \alpha_{b,k}^{n} + \sum_{r=1}^{B^{\text{U}}} \beta_{r,k}^{n-1} \beta_{r,k}^{n} \geq S^{\text{HO}},$$

where $\bar{\mathbf{C}}_{\mathbf{10}}$ denotes the HO-aware QoS constraint. In addition, we added constraint $\mathbf{C_{12}}$ to ensure that each user maintains a connection to at least S^{HO} BSs from the previous point. On top of the challenges identified in Section III.A for solving $\mathcal{P}_1$, such as non-convexity and the presence of integer variables, we encounter an additional obstacle in solving $\hat{\mathcal{P}}_1$. This arises from the multiplication of binary variables by achievable rates, specifically in the terms associated with HO costs. It is worth noting that although the binary variables can be decoupled from the beamformers as demonstrated in Section III.C, this approach does not allow for the decoupling of binary variables in the HO cost terms in U_2. Therefore, **Algorithm 1** cannot be directly employed to solve $\hat{\mathcal{P}}_1$.

In order to solve $\hat{\mathcal{P}}_1$, after obtaining the analog beamformer as in Section III.B and transforming the problem as in Section III.C, we apply quadratic programming

and MM approach at i-th iteration of the algorithm resulting in the following:

$$\hat{\mathcal{P}}_2 : \underset{\substack{\alpha^n_{b,k},\boldsymbol{w}_{b,k},p_{b,k}\\ \beta^n_{r,k},\boldsymbol{u}_{r,k},\delta_{r,k}}}{\text{maximize}} \quad U_2^{(i)} - \Gamma_1 \mathcal{L}_1^{(i)} - \Gamma_2 \mathcal{L}_2^{(i)} \tag{5.48}$$

$$\text{s.t.} \quad \mathbf{C_1}, \mathbf{C_2}, \mathbf{C_{5.1}}, \mathbf{C_{5.2}}, \mathbf{C_{5.3}}, \mathbf{C_{6.1}}, \mathbf{C_{6.2}}, \mathbf{C_{6.3}},$$

$$\mathbf{C_7}, \bar{\mathbf{C}}_{\mathbf{10}}, \mathbf{C_{11.1}}, \mathbf{C_{11.3}}, \mathbf{C_{12}},$$

where $U_2^{(i)}$ is the HO-aware sum-rate utility at i-th iteration of the Algorithm. $\hat{\mathcal{P}}2$ remains a non-convex problem due to the multiplication of variables in both the objective function and constraint $\bar{\mathbf{C}}\mathbf{10}$.

Lemma 5.2. *After applying the quadratic transformations in* (5.36) *and* (5.38)*, each HO-aware rate expression, i.e.,* $\bar{R}_k^{\mathrm{T}}$ *for* $\mathbf{C_{13}} : 0 \leq \sum_{b=1}^{B^{\mathrm{T}}} \left(1 - \alpha_{b,k}^{n-1}\right) \alpha_{b,k}^n < \frac{1}{\eta^{\mathrm{T}}}$ *and* $\bar{R}_k^{\mathrm{U}}$ *for* $\mathbf{C_{14}} : 0 \leq \sum_{r=1}^{B^{\mathrm{U}}} \left(1 - \beta_{r,k}^{n-1}\right) \beta_{r,k}^n < \frac{1}{\eta^{\mathrm{U}}}$*, is a log-concave function.*

Proof. See Section 5.6.2. □

In **Lemma** 5.2, we showed that each HO-aware rate is a log-concave function. However, since the sum of log-concave functions is not necessarily log-concave [88], the objective function $U_2^{(i)}$ and QoS constraints (sum over THz and UMB rates per user) are still non-convex. First, in order to handle the objective function, we propose the following lemma:

Lemma 5.3. *The objective* $U_2^{(i)}$ *can be replaced with the following concave lower bound:* $\tilde{U}_2^{(i)} = \frac{1}{2K} \sum_{k=1}^{K} \log(\bar{R}_k^{\mathrm{T}}) + \log(\bar{R}_k^{\mathrm{U}}) + 1$.

Proof. The lower bound of $U_2^{(i)}$ can be obtained as follows:

$$\tilde{U}_2^{(i)} = \frac{1}{2K}\sum_{k=1}^{K}\log(\bar{R}_k^{\mathrm{T}})+\log(\bar{R}_k^{\mathrm{U}})+1 \overset{(a)}{\leq} \log(\sum_{k=1}^{K}\frac{1}{2K}\bar{R}_k)+1$$
$$\overset{(b)}{\leq} \frac{1}{2K}\sum_{k=1}^{K}\bar{R}_k \overset{(c)}{\leq} \sum_{k=1}^{K}\bar{R}_k = U_2^{(i)}. \tag{5.49}$$

Following Jensen's inequality for concave function $f(x)$, we have:

$$\lambda_1 f(x_1) + \cdots + \lambda_K f(x_K) \leq f(\lambda_1 x_1 + \cdots + \lambda_K x_K). \tag{5.50}$$

By setting $\lambda_k = \frac{1}{2K}\ \forall k$ and having $x_k = \bar{R}_k$, inequality (a) can be obtained. Moreover, the inequality in (b) is justified using the property of the logarithm function: $x-1 \geq \log(x),\ \forall x > 0$, and (c) is trivial. $\tilde{U}_2^{(i)}$ is also a concave function since $\bar{R}_k^{\mathrm{T}}$ and $\bar{R}_k^{\mathrm{U}}, \forall\ k$ are log-concave functions as proved in **Lemma 2**. □

Next, we need to tackle the non-convexity of the HO-aware QoS constraints in $\bar{\mathbf{C}}_{\mathbf{10}}$. This constraint can be written as:

$$\log\left(\frac{1}{2}R_k^{\mathrm{T}} + \frac{1}{2}R_k^{\mathrm{U}}\right) \geq \log\left(\frac{1}{2}R_k^{\mathrm{Th}}\right), \forall\ k. \tag{5.51}$$

By seeking Jensen's inequality and setting $\lambda_1 = \lambda_2 = \frac{1}{2}$ in (5.50), we have:

$$\log(\frac{1}{2}\bar{R}_k) = \log\left(\frac{1}{2}R_k^{\mathrm{T}} + \frac{1}{2}R_k^{\mathrm{U}}\right) \geq \tilde{R}_k$$
$$= \frac{1}{2}\log(R_k^{\mathrm{T}}) + \frac{1}{2}\log(R_k^{\mathrm{U}}) \geq \log\left(\frac{1}{2}R_k^{\mathrm{Th}}\right) = \tilde{R}_k^{\mathrm{Th}}, \forall\ k.$$

After the above-mentioned discussions, problem $\hat{\mathcal{P}}_2$ is approximated with the following convex optimization at the i-th iteration of the Algorithm:

$$\hat{\mathcal{P}}_3 : \underset{\substack{\alpha^n_{b,k},\boldsymbol{w}_{b,k},p_{b,k} \\ \beta^n_{r,k},\boldsymbol{u}_{r,k},\delta_{r,k}}}{\text{maximize}} \quad \tilde{U}_2^{(i)} - \Gamma_1 \mathcal{L}_1^{(i)} - \Gamma_2 \mathcal{L}_2^{(i)} \tag{5.52}$$

$$\text{s.t.} \quad \mathbf{C_1}, \mathbf{C_2}, \mathbf{C_{5.1}}, \mathbf{C_{5.2}}, \mathbf{C_{5.3}}, \mathbf{C_{6.1}}, \mathbf{C_{6.2}}, \mathbf{C_{6.3}},$$

$$\mathbf{C_7}, \tilde{\mathbf{C}}_{\mathbf{10}} : \tilde{R}_k \geq \tilde{R}_k^{\text{Th}}, \mathbf{C_{11.1}}, \mathbf{C_{11.3}}, \mathbf{C_{12}}, \mathbf{C_{13}}, \mathbf{C_{14}},$$

where $\mathbf{C_{13}}$ and $\mathbf{C_{14}}$ are included to ensure the conditions of **Lemma 2**. $\hat{\mathcal{P}}_3$ is a convex optimization problem that CVX can solve efficiently. Therefore, by properly adjusting **Algorithm 1**, we can solve $\hat{\mathcal{P}}_1$, while the convex optimization solved inside the algorithm is $\hat{\mathcal{P}}_3$.

5.4.2 MOOP-based Approach

In the second HO-aware resource allocation approach, we utilize the same system sum-rate formulation used in U_1 and solve the following MOOP problem, which maximizes the system sum-rate as the first objective and minimizes the total number of HOs as the second objective.

$$\bar{\mathcal{P}}_1 : \underset{\substack{\alpha^n_{b,k},\boldsymbol{w}_{b,k},\boldsymbol{F}_b \\ \beta^n_{r,k},\boldsymbol{u}_{r,k},\boldsymbol{Q}_r}}{\text{maximize}} \quad U_1 \quad , \quad \underset{\alpha^n_{b,k},\beta^n_{r,k}}{\text{minimize}} \quad U_{\text{HO}} \tag{5.53}$$

$$\text{s.t.} \quad \mathbf{C_1}, \mathbf{C_2}, \mathbf{C_3}, \mathbf{C_4}, \mathbf{C_5}, \mathbf{C_6}, \mathbf{C_7}, \mathbf{C_8}, \mathbf{C_9}, \mathbf{C_{10}}, \mathbf{C_{11}}, \mathbf{C_{12}},$$

where U_{HO} is the total number of HOs and is given by:

$$U_{\mathrm{HO}} = \sum_{k=1}^{K} \left(\sum_{b=1}^{B^{\mathrm{T}}} (1 - \alpha_{b,k}^{n-1}) \alpha_{b,k}^{n} + \sum_{r=1}^{B^{\mathrm{U}}} (1 - \beta_{r,k}^{n-1}) \beta_{r,k}^{n} \right). \tag{5.54}$$

To solve $\bar{\mathcal{P}}_1$, the method of *weighted sum method* is utilized to transform both objective functions into a single one [136]. In this approach, the second objective function U_{HO} is added to the first objective function with a weight denoted by $\Xi \in \mathbb{R}^{+}$. Therefore, $\bar{\mathcal{P}}_1$ is transformed into the following problem:

$$\begin{aligned} \bar{\mathcal{P}}_2 : \underset{\substack{\alpha_{b,k}^{n}, \boldsymbol{w}_{b,k}, \boldsymbol{F}_b \\ \beta_{r,k}^{n}, \boldsymbol{u}_{r,k}, \boldsymbol{Q}_r}}{\text{maximize}} \quad & U_1 - \Xi U_{\mathrm{HO}} \\ \text{s.t.} \quad & \mathbf{C_1}, \mathbf{C_2}, \mathbf{C_3}, \mathbf{C_4}, \mathbf{C_5}, \mathbf{C_6}, \mathbf{C_7}, \mathbf{C_8}, \mathbf{C_9}, \mathbf{C_{10}}, \mathbf{C_{11}}, \mathbf{C_{12}}, \end{aligned} \tag{5.55}$$

Problem $\bar{\mathcal{P}}_2$ can be solved directly using **Algorithm 1**, which includes obtaining the analog beamformer using (5.22) and the subsequent transformations.

5.5 Numerical Results and Discussions

In this section, we evaluate the proposed MBN performance and resource allocation methods.

5.5.1 Simulation Set-up and Parameters

We consider a corridor with a length of $D = 350$ [m] and width of 250 [m], where BSs are deployed on both sides uniformly. The minimum vertical distance from the BSs is 30 [m]. Unless otherwise stated, the parameters used in the simulations are as follows: The carrier frequency of THz, $f_{\mathrm{T}} = 0.4$ [THz], and its bandwidth is $\omega^{\mathrm{T}} = 0.8$

[GHz]. The carrier frequency at UMB is $f_{\mathrm{U}} = 8$ [GHz] and the bandwidth is set to be $\omega^{\mathrm{U}} = 100$ [MHz]. The antenna gains are $G^{\mathrm{T}}_{\mathrm{Tx}} = 15$ [dB], $G^{\mathrm{T}}_{\mathrm{Rx}} = 8$ [dB], $G^{\mathrm{U}}_{\mathrm{Tx}} = 10$ [dB], and $G^{\mathrm{U}}_{\mathrm{Rx}} = 8$ [dB]. Power budgets are $P^{\mathrm{T}}_{\max} = 25$ [dBm] and $P^{\mathrm{U}}_{\max} = 40$ [dBm]. Cluster sizes are set as $\mathcal{C}^{\mathrm{T}}_k = 2$ and $\mathcal{C}^{\mathrm{U}}_k = 2$. For scenarios where BSs operate on only one frequency band, the cluster size is set to $\mathcal{C}_k = \mathcal{C}^{\mathrm{T}}_k + \mathcal{C}^{\mathrm{U}}_k = 4$ for fairness in comparison.

5.5.2 Benchmarks

The proposed MBN architecture is compared against the *THzOn* scenario, where no UBS is deployed in the corridor, and for a fair comparison, the total number of BSs is the same in both cases. We have *MBN-Algo1*, which denotes the proposed MBN architecture as well as resource allocation in **Algorithm 1**. The notation FC is used for the fully-connected HBF case, while PC is used for the partially-connected HBF. Meanwhile, *THzOn-Algo1* represents the deployment of only TBSs and using **Algorithm 1**. *MBN-B1* refers to the case where, in the proposed MBN, the $\mathcal{C}^{\mathrm{T}}_k$ nearest TBSs and $\mathcal{C}^{\mathrm{U}}_k$ nearest UBSs with the strongest channel gains are selected for the k-th user's association in the THz and UMB bands, respectively. The proposed beamforming method in **Algorithm 1** is then applied. The same approach is used in *THzOn-B1*, where only TBSs are deployed, and the $\mathcal{C}_k = \mathcal{C}^{\mathrm{T}}_k + \mathcal{C}^{\mathrm{U}}_k$ nearest TBSs are selected for user association. Moreover, *MBN-ZF* employs the same nearest user–BS association strategy, but uses regularized zero-forcing (RZF) for beamforming [137]. In this method, the digital beamformers at the b-th TBS and r-th UBS are given by: $\bar{\boldsymbol{W}}_b = \boldsymbol{H}_b^H \left(\boldsymbol{H}_b \boldsymbol{H}_b^H + e_b^{\mathrm{T}} \boldsymbol{I}\right)^{-1}$, $\bar{\boldsymbol{U}}_r = \boldsymbol{G}_r^H \left(\boldsymbol{G}_r \boldsymbol{G}_r^H + e_r^{\mathrm{U}} \boldsymbol{I}\right)^{-1}$, where $\boldsymbol{H}_b = \boldsymbol{F}_b^H [\alpha^n_{b,1} \boldsymbol{h}_{b,1}, \ldots, \alpha^n_{b,K} \boldsymbol{h}_{b,K}]$, with $e_b^{\mathrm{T}} > 0$ used to regularize $\boldsymbol{H}_b \boldsymbol{H}_b^H$,

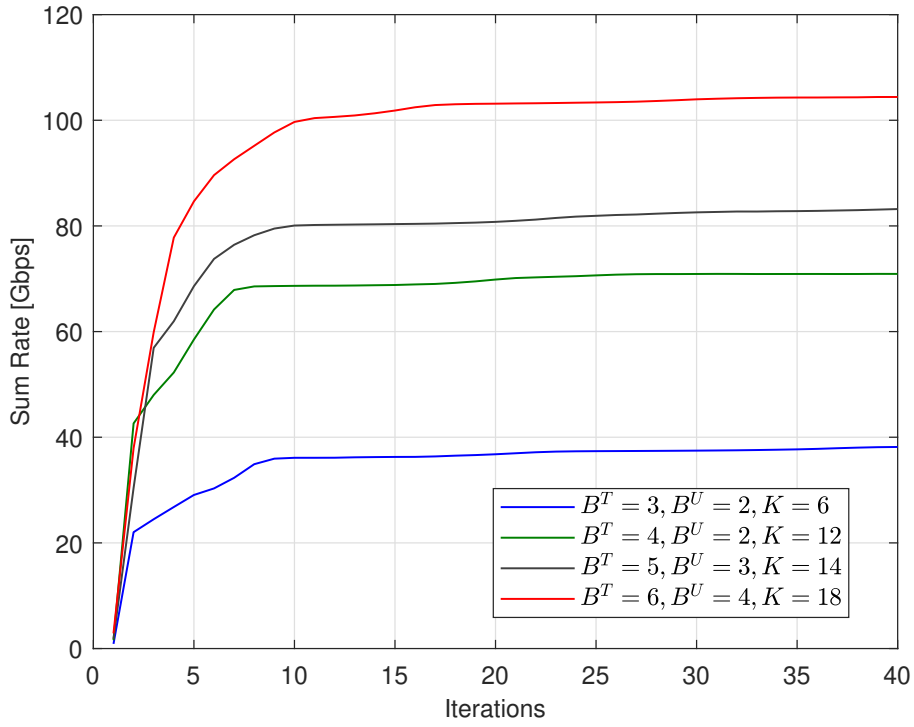


Figure 5.2: Convergence of Algorithm 1 for different system parameters *Adapted from [114]. © 2025 IEEE.*

and $\boldsymbol{G}_r = \boldsymbol{Q}_r^H[\beta_{r,1}^n \boldsymbol{g}_{r,1}, \ldots, \beta_{r,K}^n \boldsymbol{g}_{r,K}]$, with $e_r^{\mathrm{U}} > 0$ used to regularize $\boldsymbol{G}_r \boldsymbol{G}_r^H$. To satisfy the power budget constraints, the final beamformers are normalized as follows: $\boldsymbol{W}_b = \bar{\boldsymbol{W}}_b \sqrt{\frac{\bar{P}_{\max}^{\mathrm{T}}}{\sum_{k=1}^K \alpha_{b,k}^n \mathrm{tr}(\bar{\boldsymbol{W}}_b \bar{\boldsymbol{W}}_b^H)}}$, and $\boldsymbol{U}_r = \bar{\boldsymbol{U}}_r \sqrt{\frac{\bar{P}_{\max}^{\mathrm{U}}}{\sum_{k=1}^K \beta_{r,k}^n \mathrm{tr}(\bar{\boldsymbol{U}}_r \bar{\boldsymbol{U}}_r^H)}}$. A similar approach is also taken for *THzOn-ZF*.

5.5.3 Convergence of Algorithm 1

Figure 5.2 illustrates the convergence behavior of the proposed algorithm under different system parameters, including the number of users and BSs. It can be observed that **Algorithm 1** converges, as mathematically shown in Section III.E.1. As the number of variables increases, more iterations are required for convergence due to the

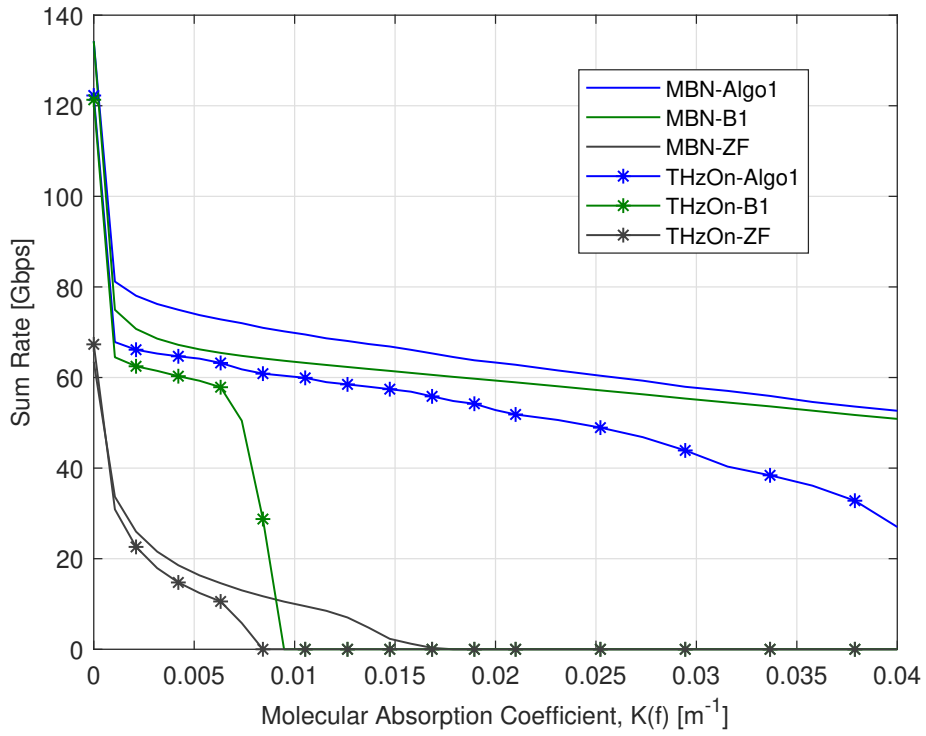


Figure 5.3: Network sum-rate at different molecular absorption coefficients, $R^{\mathrm{Th}} = 0.5$ [Gbps], $K = 12$, $B^{\mathrm{T}} = 4$, $B^{\mathrm{U}} = 2$, $B^{\mathrm{T}}_{\mathrm{THzOn}} = 6$, $M^{\mathrm{T}} = 504$, $M^{\mathrm{U}} = 84$ *Adapted from [114]. © 2025 IEEE.*

expanded search space and increased complexity of the optimization problem in each iteration.

5.5.4 Numerical Results for Stationary Users

5.5.4.1 Impact of Molecular Absorption Coefficient

Different environmental conditions, including humidity, temperature, and gas molecule ratios, can impact the molecular absorption coefficient, thereby affecting transmission

performance over the THz frequency band. Varying the molecular absorption coefficients for a given carrier frequency can provide insights related to different environmental conditions [72, 98]. Figure 5.3 illustrates the impact of increasing molecular absorption coefficients on the proposed MBN network, THzOn, and their respective benchmarks. As shown, the performance of all approaches declines with increasing $\mathcal{K}(f_{\mathrm{T}})$. This is because the molecular absorption loss in the direct channels $|h_{b,s,n}|^2_{b,k}$ increases significantly, and the molecular absorption noise further degrades the SINR in the THz band. The sum-rate for the proposed MBN network decreases at a slower rate, outperforming the scenario where only THz BSs are deployed. Also, the reason MBN-B1 approaches the performance of MBN-Algo1 is that a higher absorption coefficient leads to greater absorption loss, which increases with distance. When the molecular absorption loss becomes dominant, the optimal user association determined by Algorithm 1 converges to the $\mathcal{C}_k$ nearest BSs association to user k. Additionally, the proposed algorithm exhibits better performance compared to other benchmarks. Specifically, the sum-rate of THzOn-B1 rapidly approaches zero, while the proposed method maintains superior performance. Moreover, RZF in the MBN also outperforms THzOn-ZF, though both approaches see their sum-rates decline more quickly compared to the proposed Algorithm 1.

5.5.4.2 Impact of Blockage Density

Figure 5.4 illustrates the impact of increasing the density of blockers. As shown, the proposed MBN outperforms THzOn as it can benefit from the availability of UMB frequency as blockage density increases. Additionally, the proposed **Algorithm 1**

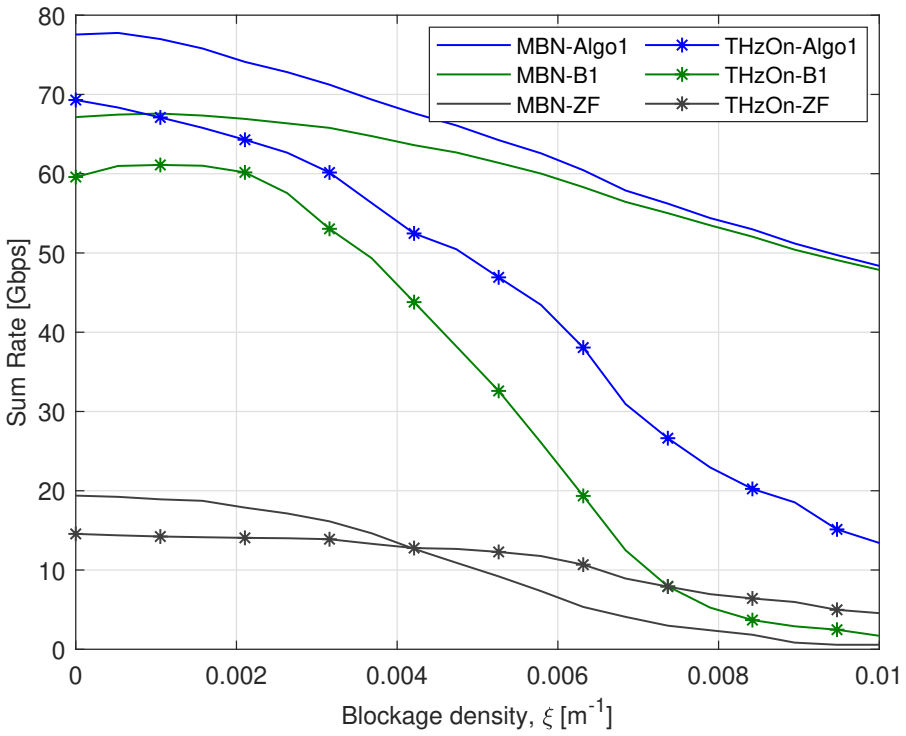


Figure 5.4: System sum-rate versus density of blockers, ξ. $R^{\mathrm{Th}} = 0.5$ [Gbps], $K = 12$, $B^{\mathrm{T}} = 4$, $B^{\mathrm{U}} = 2$, $B^{\mathrm{T}}_{\mathrm{THzOn}} = 6$, $M^{\mathrm{T}} = 504$, $M^{\mathrm{U}} = 84$ *Adapted from [114].* *© 2025 IEEE.*

outperforms other benchmarks. At high blockage densities, the performance of MBN-B1 converges to that of the proposed algorithm because the optimal association in MBN-Algo1 tends to favor non-blocked links, which are typically the nearest BSs to the user, as shown in (5.10). Also, after a certain point, THzOn-ZF outperforms MBN-ZF, which is attributed to the non-optimal resource allocation in this scheme.

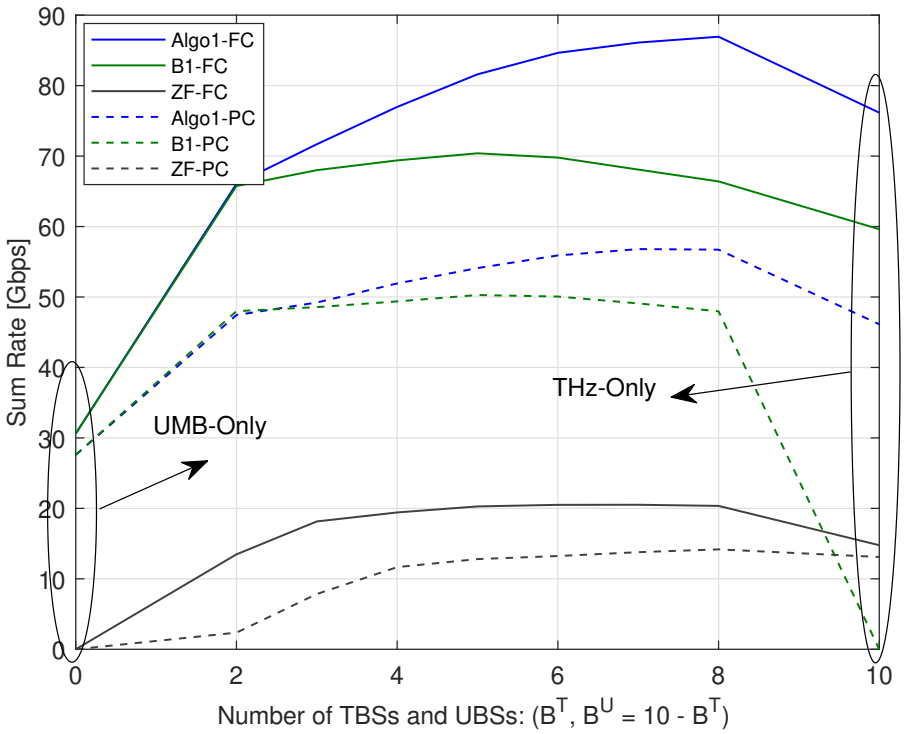


Figure 5.5: System sum-rate versus number of TBSs and UBSs. The total number of BSs is 10 and increasing TBSs decreases UBSs as B^{U} varies within the range of $[10, 0]$. $R^{\mathrm{Th}} = 0.5$ [Gbps], $K = 12$, $M^{\mathrm{T}} = 504$, $M^{\mathrm{U}} = 84$. *Adapted from [114].* *© 2025 IEEE.*

5.5.4.3 Significance of MBN Deployment with Fully-Connected and Partially-Connected HBF

Figure 5.5 illustrates the impact of deploying different types of BSs. We assume a total of 10 BSs, and as the number of TBSs on the x-axis increases, the number of UBSs decreases accordingly. For example, when the x-axis value is $B^{\mathrm{T}} = 6$, the number of UBSs is $B^{\mathrm{U}} = 4$. In this figure, all of the channel characteristics, including blockage with $\xi = 0.002$, are considered in the transitioning from UMB-only to MBN and then to THz-only network. It is evident that the MBN can outperform both UMB-only

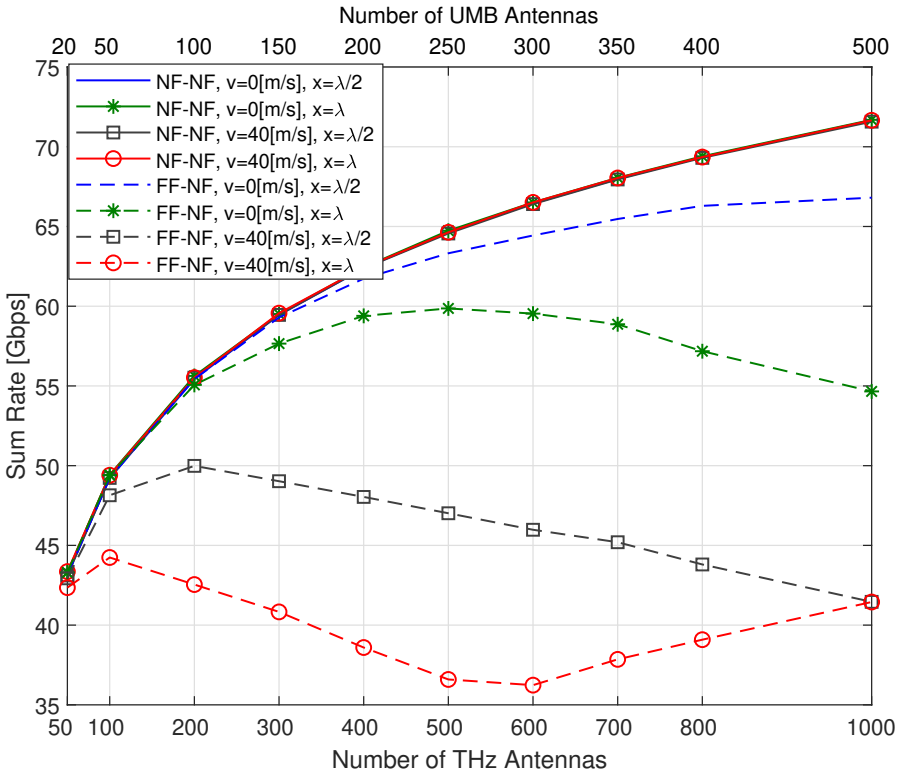


Figure 5.6: System sum-rate versus the different number of antennas of UMB and THz bands for various velocities of users and antenna spacing. $R^{\mathrm{Th}} = 0.1$ [Gbps], $K = 8$, $B^{\mathrm{T}} = 4$, $B^{\mathrm{U}} = 2$, $\mathcal{C}_k^{\mathrm{T}} = 4$, $\mathcal{C}_k^{\mathrm{U}} = 2$ *Adapted from [114]. © 2025 IEEE.*

and THz-Only setups for a predefined deployment ratio. The figure also demonstrates the superiority of the proposed algorithm over other benchmarks. Moreover, in both the proposed Algorithm 1 and Benchmark B1, where beamforming is optimized, FC HBF consistently outperforms PC HBF due to utilizing more phase shifters at the cost of higher energy consumption. However, it can also be observed that even with PC HBF, the MBN still outperforms the single-band counterparts.

5.5.4.4 Number of Antennas and NFC vs FFC

Figure 5.6 illustrates the impact of increasing the number of antennas in both frequency bands on the system sum-rate across different user velocities and antenna spacing. In the cases labeled 'NF-NF,' resource allocation is performed using **Algorithm-1** based on NFC channel modeling, whereas in 'FF-NF,' resource allocation is carried out using FFC channel modeling, with NFC considered as the ground truth. Increasing both the number of antennas and the antenna spacing expands the array aperture, which in turn increases the Fraunhofer distance. Below this distance, NFC provides a more accurate representation compared to FFC. It can be observed that increasing antenna spacing amplifies the sensitivity to the choice of channel modeling. Furthermore, in the presence of mobility, when the antenna spacing is half a wavelength, i.e., $x = \frac{\lambda}{2}$, increasing user velocity further degrades system performance due to mismatched beamforming under the FFC assumption. However, it can also be observed that when $x = \lambda$, the performance degradation levels off at a certain point, which may be attributed to the energy spread effect [25]. Moreover, Figures 5.7 and 5.8 further illustrate the differences between the array gain of NFC-based and FFC-based beamforming for both THz and UMB. As shown, increasing the antenna spacing, which increases the array aperture, also increases sensitivity to the choice between near- and far-field modeling. A larger spacing or the use of subarrays with spacing greater than half a wavelength enables NFC, which is advantageous as it introduces an additional degree of freedom by enabling beamfocusing in both the angular and distance domains [4]. It is also observed that higher user velocities amplify the mismatch resulting from far-field beamforming. When the antenna spacing is set to half a wavelength, i.e., $x = \frac{\lambda}{2}$, both THz and UMB experience beam squint due to

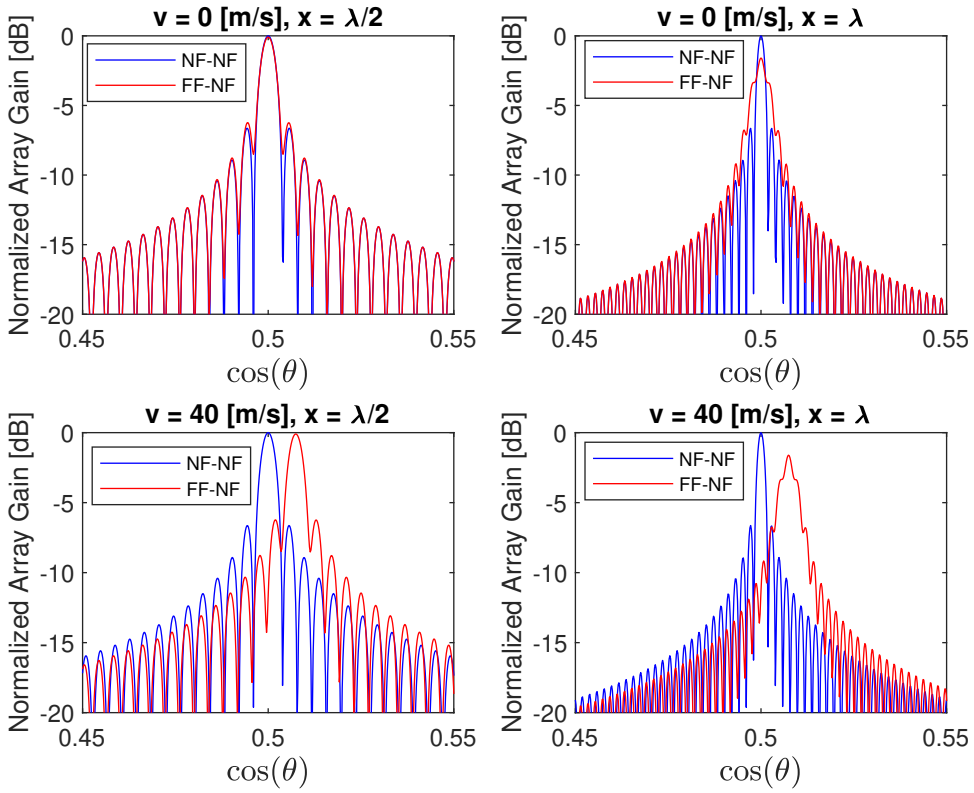


Figure 5.7: Normalized Array Gain at the THz band for different user velocity and antenna spacing, considering the distance from the transmitter is 40 [m], $\cos(\theta) = 0.5$ and $M^{\mathrm{T}} = 504$

the far-field assumption. It is also worth noting that employing near-field modeling increases the complexity of channel acquisition in practice, as discussed in [138]. This introduces a trade-off between the performance improvement offered by NFC-based beamforming and the additional overhead required for CSI acquisition.

5.5.5 Numerical Results for Mobile Users

In the simulation of mobile users moving at a speed of 40 [m/s], we consider three consecutive optimization windows (trajectory points), each of length 100 [ms] [42],

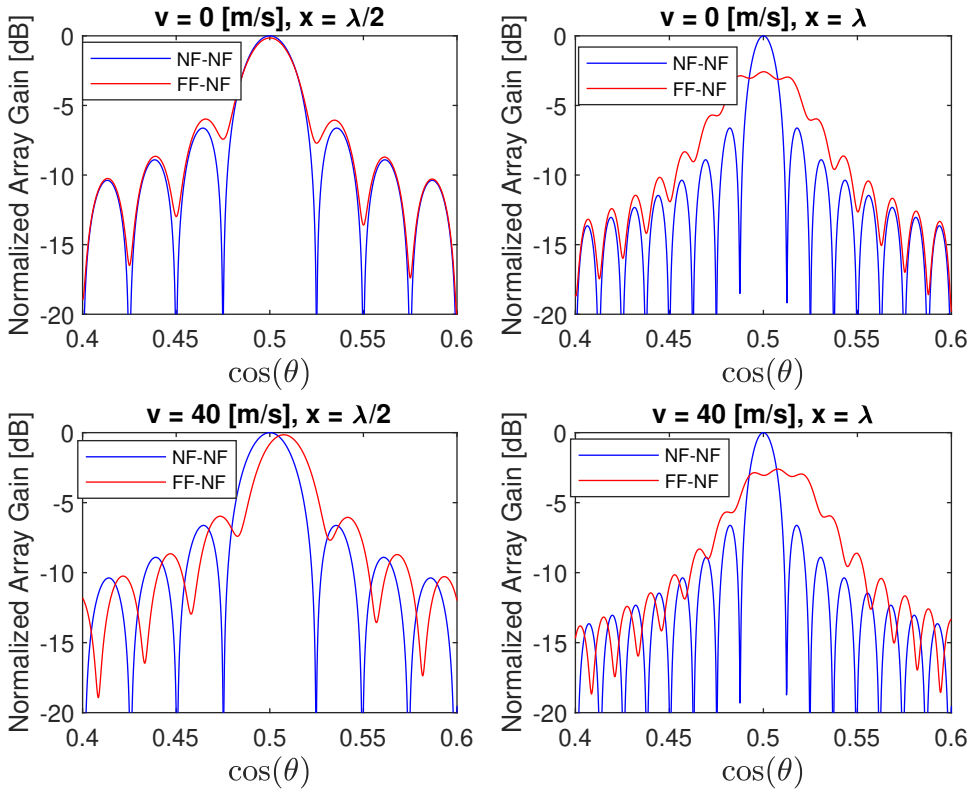


Figure 5.8: Normalized Array Gain at the UMB for different user velocity and antenna spacing, considering the distance from the transmitter is 40 [m], $\cos(\theta) = 0.5$, and and $M^{\mathrm{U}} = 84$

and report the average HO-aware sum-rate in (5.46) and the number of HOs in (5.54). The remaining parameters are specified in the description of each simulation result. In these results, *MBN-Algo1-Cost* represents the first proposed HO-aware resource allocation, while *MBN-Algo1-MO* refers to the second HO-aware method. Similar approaches are applied to the THz-Only case. For benchmarks labeled "(No Cost)," resource allocation is performed via **Algorithm 1** without HO consideration (i.e., HO-agnostic), while their achievable rates are computed using the HO-aware rate expressions in (5.44) and (5.45).

5.5.5.1 Impact of Cluster Size

Figure 5.9 illustrates the trade-off between cluster size and the number of HOs when no HO control is applied. For the THzOn case, we consider two configurations: one with a cluster size of $\mathcal{C}_k = \mathcal{C}_k^{\mathrm{T}} + \mathcal{C}_k^{\mathrm{U}}$, as used in the other results of this work, and one with $\mathcal{C}_k = \mathcal{C}_k^{\mathrm{T}}$. It is evident that increasing the cluster size enhances the system sum-rate due to greater cooperative beamforming capability. On the other hand, the total number of HOs initially increases with cluster size and then decreases. While more HOs can degrade system performance, the improvement from larger cluster sizes can offset this effect. As shown, when the cluster size increases from $\mathcal{C}_k = 1 + 1$ to $\mathcal{C}_k = 2 + 2$, the green curve for THzOn experiences a performance drop, followed by improvement beyond that point. A similar trend is observed for the black curve of THzOn, where the performance decreases from $\mathcal{C}_k = 3$ to $\mathcal{C}_k = 5$, coinciding with the highest number of HOs. It is also observed that MBN results in fewer HOs and consequently achieves a higher HO-aware sum-rate.

5.5.5.2 Impact of HO Cost

In Fig. 5.10, the impact of increasing the HO cost for both frequency bands on the proposed methods is illustrated. First, note that the proposed MBN without HO consideration in the optimization, referred to as MBN-Algo1(No Cost), outperforms its counterpart with only THz BSs deployed, i.e., THzOn-Algo1(No Cost). The THzOn scenario is observed to be more sensitive to increasing HO costs. Thus, even without accounting for HOs during resource allocation, the proposed MBN network outperforms its single-band counterpart. While the performance of MBN-Algo1-MO does not drop as the HO cost increases, the main challenge with this method is determining

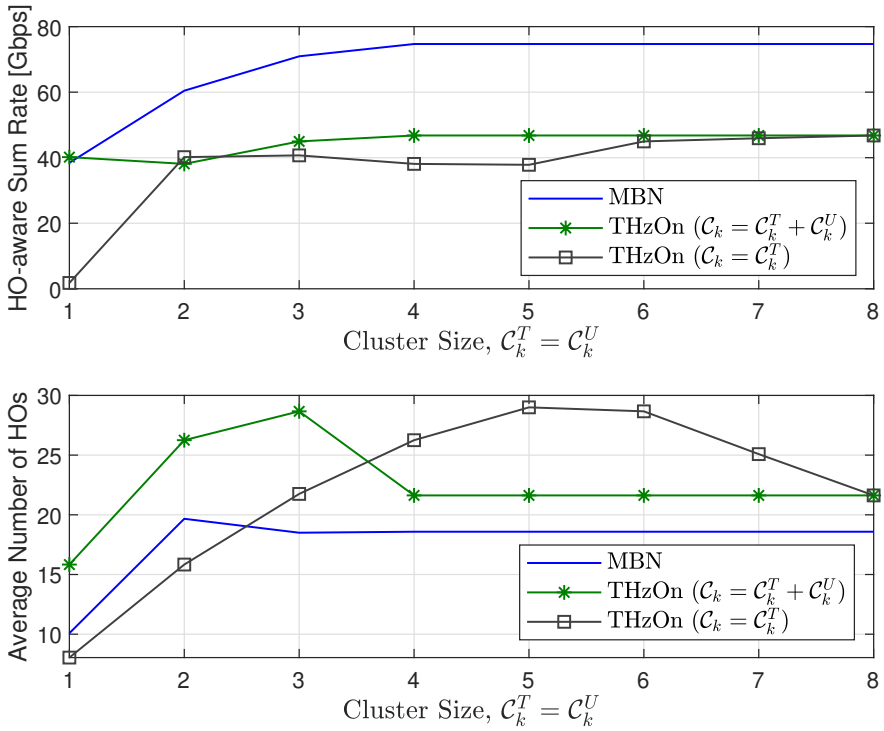


Figure 5.9: HO-aware sum-rate and the total number of HOs versus different cluster size ($\mathcal{C}_k^{\mathrm{T}} = \mathcal{C}_k^{\mathrm{U}}$), where $R^{\mathrm{Th}} = 0.5$ [Gbps], $K = 12$, $B^{\mathrm{T}} = 5$, $B^{\mathrm{U}} = 3$, $B_{\mathrm{THzOn}}^{\mathrm{T}} = 8$, and $\eta^{\mathrm{T}} = \eta^{\mathrm{U}} = 0.4$. *Adapted from [114]. © 2025 IEEE.*

an optimal value for Ξ, which serves as a hyperparameter and varies across scenarios Moreover, while MBN-B1(No Cost) is more robust to HOs, it lacks the flexibility to optimize user association and mitigate the severe effects of HOs. The proposed HO-aware methods also improve the performance of the THzOn case, although it underperforms its MBN counterpart as the HO cost increases.

Fig. 5.11 shows the impact of increasing the HO cost on the number of HOs. As seen, increasing the HO cost significantly reduces the number of HOs in MBN-Algo1-Cost for both MBN and THzOn scenarios. Also, the MOOP-based method maintains a nearly constant number of HOs, which is due to the absence of a direct relationship

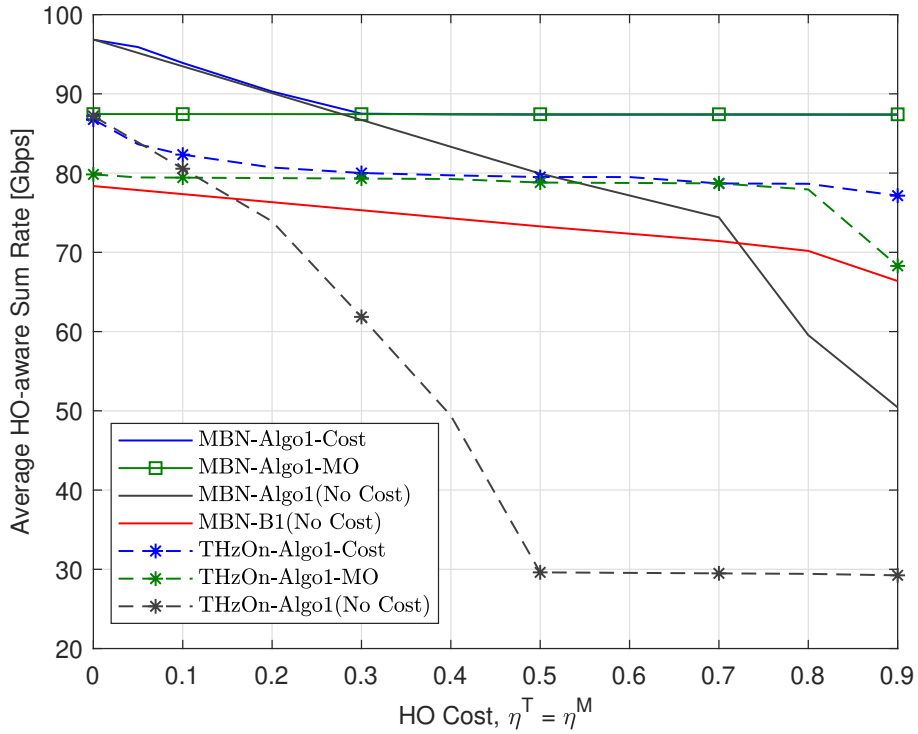


Figure 5.10: HO-aware sum-rate versus different HO cost ($\eta^{\mathrm{T}} = \eta^{\mathrm{U}}$), where $R^{\mathrm{Th}} = 0.5$ [Gbps], $K = 15$, $B^{\mathrm{T}} = 5$, $B^{\mathrm{U}} = 3$, $B^{\mathrm{T}}_{\mathrm{THzOn}} = 8$, and $\Xi = 1$. *Adapted from [114]. © 2025 IEEE.*

between HO and rate in its optimization formulation. MBN-B1(No Cost) results in fewer HOs compared to MBN-Algo1(No Cost), although it lacks the capability to achieve optimal resource allocation. Moreover, we note that even without HO consideration in the resource allocation, MBN imposes a smaller number of HOs in mobile scenarios.

5.5.5.3 Impact of Minimum Rate Requirement

Fig. 5.12 illustrates the impact of increasing the minimum rate requirement on the HO-aware system sum-rate. The proposed HO-aware resource allocation consistently

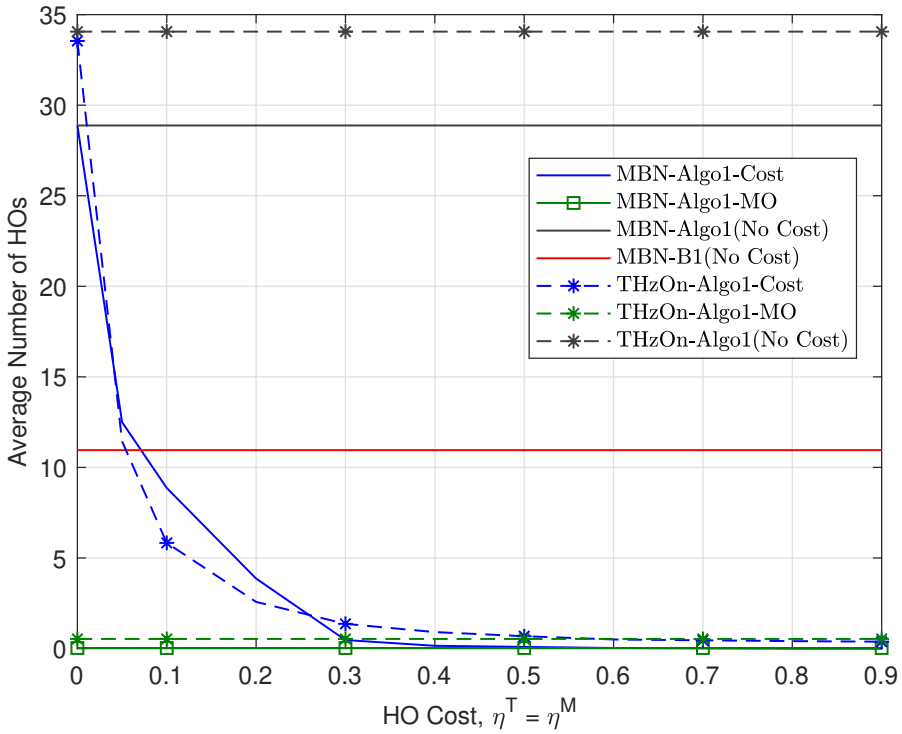


Figure 5.11: Total number of HOs versus different HO cost ($\eta^{\mathrm{T}} = \eta^{\mathrm{U}}$), where $R^{\mathrm{Th}} = 0.5$ [Gbps], $K = 15$, $B^{\mathrm{T}} = 5$, $B^{\mathrm{U}} = 3$, $B^{\mathrm{T}}_{\mathrm{THzOn}} = 8$, and $\Xi = 1$. *Adapted from [114]. © 2025 IEEE.*

outperforms the other methods. Additionally, due to a higher number of HOs in the THzOn system compared to MBN, THzOn demonstrates inferior performance. It is also observed that increasing the minimum rate requirement has a more adverse effect on THzOn-Algo1-MO than on THzOn-Algo1-Cost, highlighting the importance of incorporating HO-aware QoS constraints.

5.5.5.4 Impact of Imperfect CSI

Since obtaining perfect CSI is challenging in practice, we examine the robustness of the proposed methods to CSI uncertainties by considering a statistical CSI error

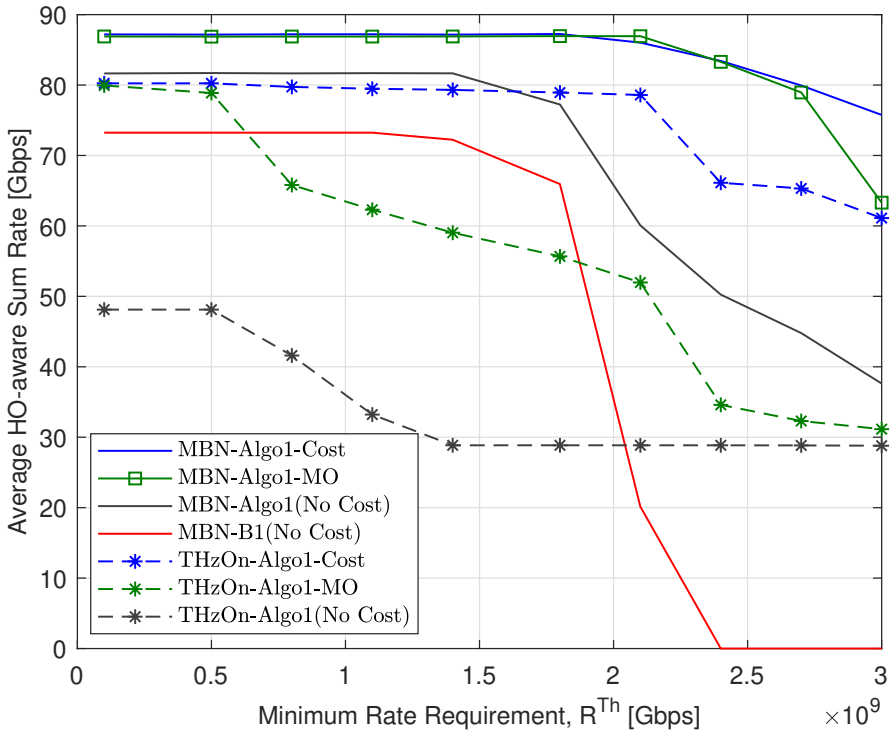


Figure 5.12: HO-aware sum-rate versus minimum rate requirement R^{Th}, where $K = 15$, $B^{\mathrm{T}} = 5$, $B^{\mathrm{U}} = 3$, $B^{\mathrm{T}}_{\mathrm{THzOn}} = 8$, $\Xi = 1$, and $\eta^{\mathrm{T}} = \eta^{\mathrm{U}} = 0.4$ *Adapted from [114]. © 2025 IEEE.*

model, as in [139]. For instance, for THz links, the channel is expressed as:

$$\boldsymbol{h}^{\mathrm{Ideal}}_{b,k} = \boldsymbol{h}_{b,k} + \Delta\boldsymbol{h}_{b,k}, \quad \forall b, k, \tag{5.56}$$

where $\boldsymbol{h}^{\mathrm{Ideal}}_{b,k}$ denotes the perfect CSI, $\boldsymbol{h}_{b,k}$ is the estimated CSI, and $\Delta\boldsymbol{h}_{b,k}$ represents the CSI estimation error. The error term $\Delta\boldsymbol{h}_{b,k}$ is modeled as a circularly symmetric complex Gaussian vector. Following [139, 140], the error variance is defined as $\hat{\chi} = \chi||\boldsymbol{h}_{b,k}||^2_2$, where $\chi \in [0, 1)$ quantifies the level of CSI uncertainty. Assuming the same value of χ for all channels, Figure 5.13 illustrates the impact of imperfect CSI on the proposed methods. As the CSI uncertainty increases, the system performance of all

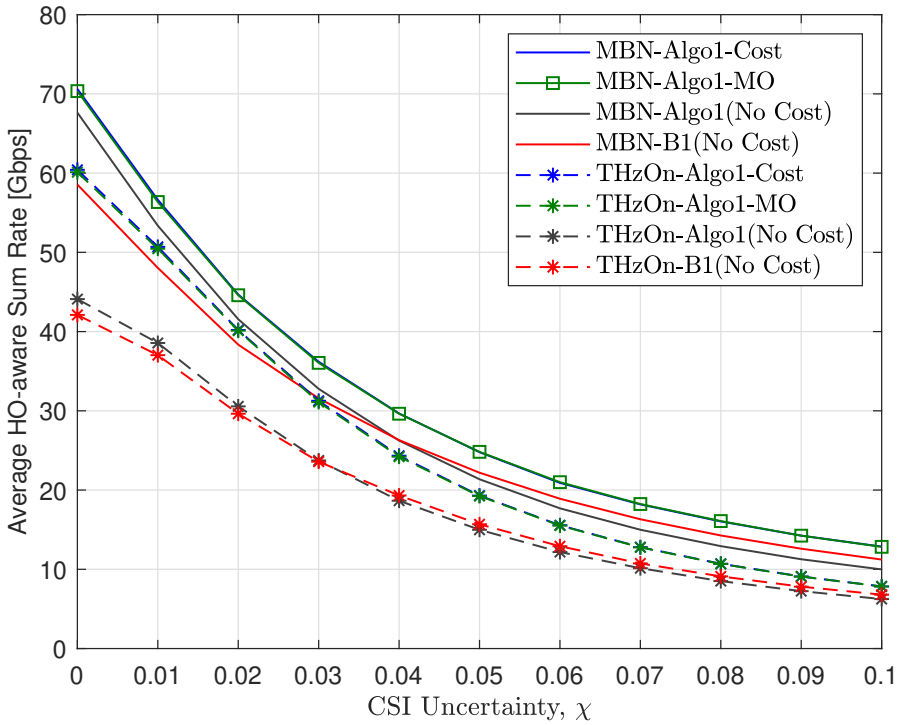


Figure 5.13: HO-aware system sum-rate versus CSI uncertainty χ, where $K = 12$, $B^{\mathrm{T}} = 4$, $B^{\mathrm{U}} = 2$, $B^{\mathrm{T}}_{\mathrm{THzOn}} = 6$, $\Xi = 1$, and $\eta^{\mathrm{T}} = \eta^{\mathrm{U}} = 0.4$ *Adapted from [114].* *© 2025 IEEE.*

methods degrades. This degradation is attributed to the large number of antennas in ELAAs, which amplifies the CSI error variance as described in (5.56). It can be observed that for both MBN and THz-On, the performance of MBN-B1 and THzOn-B1 approaches and outperforms that of the proposed HO-unaware methods as the CSI uncertainty increases. This is because, when the estimation error is proportional to the channel gain, resource allocation strategies based on associating with the nearest BSs tend to be less sensitive to CSI errors. Nevertheless, the proposed HO-aware methods still outperform these benchmarks due to their explicit consideration of HOs in the objective function.

5.6 Complete Proof

5.6.1 Proof of Lemma 5.1

Given the optimal and feasible $\boldsymbol{X}^o$ in (5.34), define $y^o_{i,j} = \frac{\sqrt{A_{i,j}(\boldsymbol{X}^o)}}{B_{i,j}(\boldsymbol{X}^o)}$. Then, it can be shown that the term i and j in the objective function in (5.35) at $y^o_{i,j}$ is: $T_{i,j}(\boldsymbol{X}^o, y^o_{i,j}) = \frac{A_{i,j}(\boldsymbol{X}^o)}{B_{i,j}(\boldsymbol{X}^o)}$. Hence, by taking the logarithm, we have:

$$\log\left(2\Re\left\{y^{o,*}_{i,j}\sqrt{A_{i,j}(\boldsymbol{X}^o)}\right\} - \left|y^o_{i,j}\right|^2 B_{i,j}(\boldsymbol{X}^o)\right) = \log\left(\frac{A_{i,j}(\boldsymbol{X}^o)}{B_{i,j}(\boldsymbol{X}^o)}\right)$$

and by summing over all i and j, the objective function in (5.35) at $(\boldsymbol{X}^o, \boldsymbol{y}^o)$ matches the objective of the problem in (5.34).

Also, since $\boldsymbol{X}^o$ is feasible in (5.34), we have: $\sum\limits_j \log\left(\frac{A_{i,j}(\boldsymbol{X}^o)}{B_{i,j}(\boldsymbol{X}^o)}\right) \geq R_i^{Th}$, and by using the result of the previous step and evaluating the terms in $(\boldsymbol{X}^o, \boldsymbol{y}^o)$, we have:

$$\sum_j \log\left(2\Re\left\{y^{o,*}_{i,j}\sqrt{A_{i,j}(\boldsymbol{X}^o)}\right\} - \left|y^o_{i,j}\right|^2 B_{i,j}(\boldsymbol{X}^o)\right) \geq R_i^{Th}, \forall i.$$

Therefore, the objective and feasible set are equivalent for $\boldsymbol{y}^o$. Now for optimality in the second problem, suppose there exists a feasible solution $(\boldsymbol{X}', \boldsymbol{y}')$ to (5.35) such that $\sum_i \sum_j \log\left(2\Re\left\{y'^{*}_{i,j}\sqrt{A_{i,j}(\boldsymbol{X}')}\right\} - \left|y'_{i,j}\right|^2 B_{i,j}(\boldsymbol{X}')\right) > \sum_i \sum_j \log\left(\frac{A_{i,j}(\boldsymbol{X}^o)}{B_{i,j}(\boldsymbol{X}^o)}\right)$. Since the maximum of the term inside the logarithm occurs at $y_{i,j} = \frac{\sqrt{A_{i,j}(\boldsymbol{X})}}{B_{i,j}(\boldsymbol{X})}$, we have:

$$2\Re\left\{y'^{*}_{i,j}\sqrt{A_{i,j}(\boldsymbol{X}')}\right\} - \left|y'_{i,j}\right|^2 B_{i,j}(\boldsymbol{X}') \leq \frac{A_{i,j}(\boldsymbol{X}')}{B_{i,j}(\boldsymbol{X}')}. \tag{5.57}$$

Moreover, as the logarithm is a monotonically increasing function, by taking the

logarithm and summing over j and i on (5.57), we have:

$$\sum_i \sum_j \log\left(2\Re\left\{y'^{*}_{i,j}\sqrt{A_{i,j}\left(\boldsymbol{X}'\right)}\right\} - \left|y'_{i,j}\right|^2 B_{i,j}\left(\boldsymbol{X}'\right)\right)$$
$$\leq \sum_i \sum_j \log\left(\frac{A_{i,j}\left(\boldsymbol{X}'\right)}{B_{i,j}\left(\boldsymbol{X}'\right)}\right),$$

which implies:

$$\sum_i \sum_j \log\left(\frac{A_{i,j}\left(\boldsymbol{X}'\right)}{B_{i,j}\left(\boldsymbol{X}'\right)}\right) > \sum_i \sum_j \log\left(\frac{A_{i,j}\left(\boldsymbol{X}^o\right)}{B_{i,j}\left(\boldsymbol{X}^o\right)}\right),$$

which contradicts the optimality of $\boldsymbol{X}^o$ in the first problem in (5.34). Therefore, $\boldsymbol{X}^o$ is the optimal solution to the first problem while $(\boldsymbol{X}^o, \boldsymbol{y}^o)$ is optimal for the second problem.

5.6.2 Proof of Lemma 5.2

Each HO-aware rate expression after applying the quadratic transformation can be given in the format of $f(x,y) = (1-y)\log(1+x-x^2)$. To show that $f(x,y)$ is a log-concave function, we need to show that the Hessian of $g(x,y) = \log(f(x,y))$ is negative-definite and $f(x,y) > 0$. The Hessian of $g(x,y)$ is given by:

$$\boldsymbol{H} = \nabla^2 g = \begin{bmatrix} \frac{(-2x^2+2x-3)\log(-x^2+x+1)-(1-2x)^2}{(-x^2+x+1)^2\log^2(-x^2+x+1)} & 0 \\ 0 & -\frac{1}{(1-y)^2} \end{bmatrix}$$

We need to show that the determinant of $\boldsymbol{H}$ is non-negative, which implies that the leading element of $\boldsymbol{H}$ must be non-positive. It can be concluded that the leading element $\boldsymbol{H}_{11}(x) \leq 0, \quad \forall x \in (0,1)$. This region also complies with the condition of $x - x^2 > 0$, which is necessary to avoid negative SINR. Therefore, both of the

eigenvalues of $\boldsymbol{H}$ are non-positive and the Hessian of $g(x, y)$ is negative semidefinite. On the other hand, with the condition of $f(x, y) > 0$ and the fact that HO cost cannot be negative, we have $0 \leq y < 1$ and $x - x^2 > 1$, where the latter is already met. Therefore, we can conclude that for $y \in [0, 1)$ and $x \in (0, 1)$, the function $f(x, y)$ is log-concave. Also, the condition of $y \in [0, 1)$ results in $0 \leq \sum_{b=1}^{B^{\mathrm{T}}} \left(1 - \alpha_{b,k}^{n-1}\right) \alpha_{b,k}^{n} < \frac{1}{\eta^{\mathrm{T}}}$ for THz and $0 \leq \sum_{r=1}^{B^{\mathrm{U}}} \left(1 - \beta_{r,k}^{n-1}\right) \beta_{r,k}^{n} < \frac{1}{\eta^{\mathrm{U}}}$ for UMB.

5.7 Summary

This chapter presents a comprehensive sum-rate maximization framework for cooperative UMB/THz MBNs through the joint optimization of hybrid beamforming and user association, subject to constraints such as power budget, QoS, and maximum cluster size, for both stationary and moving users. The proposed framework incorporates NFC channel modeling, fully and partially connected hybrid beamforming architectures, and users' mobility.

Specifically, the coexistence of UMB/THz in the considered MBN improves the system performance under higher molecular absorption as well as THz link blockage. While near-field channel modeling requires more parameters to be known in practice, performing NFC-based beamforming, as opposed to mismatched FFC-based beamforming, becomes essential when using ELAAs. Moreover, we observe that increasing the cluster size in cooperative systems is not always beneficial in mobility scenarios due to HO costs. Additionally, the proposed HO-aware resource allocation methods effectively mitigate the impact of HOs, particularly when the cost of each HO and the minimum rate requirement of each user are high. Moreover, investigating the integrated deployment of MBN, as proposed in [72], within the considered framework

is a potential research topic. Exploring the feasibility of ML-based methods for solving the optimization problem, such as the approach in [141], is another direction for future research.

Copyright and Co-authorship Statement

This chapter has resulted in the following publication: M. A. Saeidi and H. Tabassum, "Resource Allocation in Cooperative Mid-Band/THz Networks in the Presence of Mobility," *IEEE Transactions on Wireless Communications*, vol. 25, pp. 5046–5062, 2026, doi: 10.1109/TWC.2025.3615916 [114]. © 2025 IEEE.

I was the primary author of this work and was responsible for the problem formulation, system modeling, mathematical derivations, algorithm development, simulation design, numerical analysis, interpretation of results, and manuscript preparation. Author 2 supervised the work and refined the manuscript.

Chapter 6

Characterization of Beam-Squint and Beam-Split Effects in RIS-assisted Multi-Frequency Networks

6.1 Overview

Next-generation wireless networks are expected to operate across diverse transmission frequencies, either through wideband signaling around a designated carrier [109] or via simultaneous use of multiple carriers across distinct spectrum bands in multi-band networks (MBNs) [72, 98, 114]. While these strategies enhance data rates and spectral flexibility, they also exacerbate frequency-dependent beamforming effects in large-aperture arrays and reconfigurable intelligent surfaces (RISs). When a beamforming profile, whether implemented at an active array or programmed at an RIS, is designed at a specific frequency, any deviation between the configuration and operating frequencies distorts the intended spatial response. This mismatch manifests as *beam-squint*, where the main lobe shifts from the desired direction, and *beam-split*, where multiple dominant lobes appear. These impairments intensify at upper mid-band, mmWave, and THz frequencies, where large electrical apertures produce narrow beams that are inherently more sensitive to frequency variations. In highly

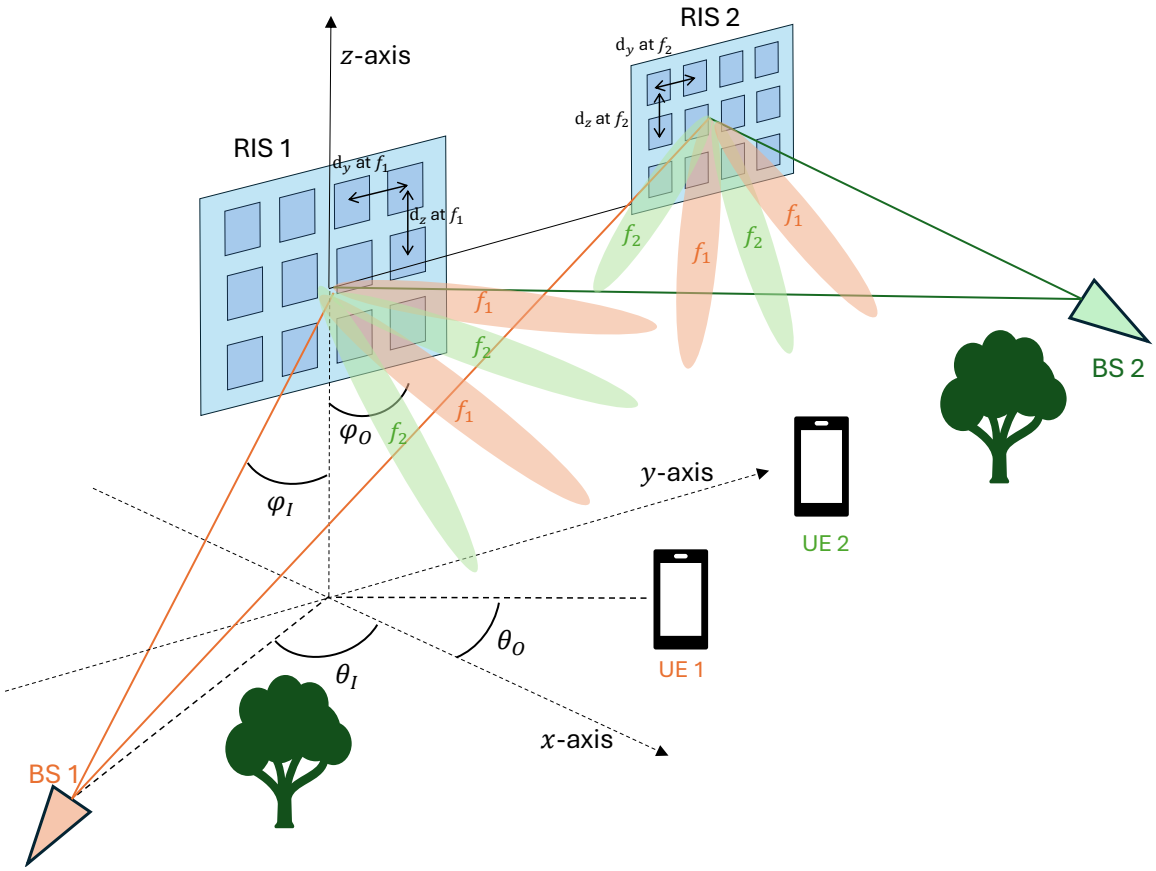


Figure 6.1: Two-RIS two-UE multi-band network: RIS 1 (RIS 2) is configured at f_1 (f_2); colored beams illustrate beam-split under cross-band evaluation, with angles (φ, θ) defined by the shown coordinate axes.

directional transmission and reflection scenarios, such as RIS-assisted links, even minor frequency-induced angular deviations can translate into substantial array gain loss, undermining the expected benefits of wideband and multi-band operation.

Consequently, multi-frequency systems require analytical tools capable of predicting how beam directions evolve with frequency. Such characterization is essential for frequency-aware beam management, wideband beamforming, and resource allocation, enabling efficient design without resorting to exhaustive angular scanning at

every carrier. By analytically capturing the evolution of beams under frequency mismatch, one can systematically mitigate beam-squint and beam-split effects rather than treating them as empirical artifacts.

In this chapter, we mathematically formalize the beam-squint and beam-split considering uniform planar array (UPA) RISs. We define beam-squint as the deviation of the main lobe from the desired elevation–azimuth angle pair, and beam-split as the emergence of additional dominant lobes when an RIS configuration designed at one frequency is evaluated at another. Specifically, we address the question of *which elevation and azimuth angle pairs maximize the RIS beamforming gain under incident frequency mismatch, while considering both continuous and b-bit quantized phase profiles?*

To this end, our contributions can be summarized as follows:

• We develop a unified mathematical framework to characterize the beam-squint and beam-split effects of UPA-RISs in the presence of both continuous phase and b-bit quantized phase profile. For the quantization case, we propose a correction procedure to reduce the quantization-induced error in the beam-squint/split analytical estimates.

• For the continuous phase profile of the RIS, we derive necessary and sufficient conditions to maximize beamforming gain as well as feasibility conditions under which beam-squint and beam-split exist. We then obtain elevation–azimuth squint/split angle pairs in closed-form as a function of the incidence and configuration angles and frequencies, and element spacing of the UPA-RIS.

• For b-bit phase quantization, we derive the quantization-error phasor via its Fourier expansion, and transform the quantized UPA beampattern into an explicit

superposition of harmonic array responses with affine phase slopes. This converts the non-affine quantized phase profile into tractable per-harmonic beamforming gain maximization problems. We then derive closed-form elevation–azimuth squint/split angle pairs, followed by deriving feasibility conditions, and a principled dominance ordering across harmonics.

• Since the per-harmonic beamforming gain maximizers (or peaks) can deviate from the true peaks of the *superposed* beampattern, we develop a local perturbation analysis around each nominal harmonic peak and derive a closed-form deviation of the peak. This provides an analytically grounded refinement that improves peak-location characterization without exhaustive 2D searches, especially for moderate RIS sizes.

• We validate, through Monte-Carlo simulations, the derived analytic expressions of squint/split, the split-existence conditions, and show how the proposed correction method reduced the quantization-induced deviation of the per-harmonic beamforming gain peaks. We then demonstrate the significance of the beam-split-aware scheduling in a multi-RIS, multi-user, and multi-band network in terms of the average sum-rate, with larger gains for larger RISs.

The rest of this chapter is organized as follows. Section 6.2 presents the RIS-assisted multi-frequency system model and formulates the beamforming gain maximization problem. Section 6.3 characterizes beam-squint and beam-split for continuous RIS phase profiles and derives the corresponding feasibility conditions. Section 6.4 extends the analysis to b-bit quantized RIS phase profiles using harmonic decomposition. Section 6.5 develops the per-harmonic peak correction mechanism to refine the predicted squint/split angle locations under harmonic superposition. Section 6.6 presents a beam-split-aware scheduling use case for a multi-RIS, multi-user,

and multi-band network. Section 6.7 provides numerical results and discussions to validate the analytical characterization and evaluate the proposed beam-split-aware scheduling approach. Finally, Section 6.9 summarizes the chapter.

Notations: Boldface lowercase/uppercase letters denote vectors/matrices. $\mathbb{R}$ and $\mathbb{C}$ denote the real and complex sets, respectively. $|\cdot|$ is the absolute value and $\|\cdot\|$ is the Euclidean norm; $\|\cdot\|_F$ is the Frobenius norm. $(\cdot)^T$ and $(\cdot)^H$ denote transpose and conjugate transpose, $\Re\{\cdot\}$ the real part, and $(\cdot)^*$ complex conjugation. $\nabla(\cdot)$ denotes the gradient, and $\otimes$ denotes Kronecker products.

6.2 System Model and Problem Statement

Assume that a plane wave of frequency f_I impinges on a UPA-based RIS from elevation angle $\varphi_I \in (-\frac{\pi}{2}, \frac{\pi}{2})$ and azimuth angle $\theta_I \in (-\frac{\pi}{2}, \frac{\pi}{2})$. The array response vector can then be modeled to obtain the respective phase shifts among the RIS elements as follows:

$$\boldsymbol{a}(\varphi_I, \theta_I, f_I, f_C) = [1, \ldots, e^{j2\pi(N_z-1)d_z \frac{\sin\varphi_I}{\lambda_I}}]^T \otimes [1, \ldots, e^{j2\pi(N_y-1)d_y \frac{\sin\theta_I \cos\varphi_I}{\lambda_I}}]^T \quad (6.1)$$

where N_z and N_y are the number of vertical and horizontal elements in the UPA, respectively, and $N = N_y N_z$ is the total number of elements. The wavelength of the incident signal is $\lambda_I = \frac{c}{f_I}$, with c being the speed of light. f_C is the design frequency of the RIS. The vertical element spacing is $d_z = \alpha_z \lambda_C = \frac{\alpha_z c}{f_C}$, and the horizontal element spacing is $d_y = \alpha_y \lambda_C = \frac{\alpha_y c}{f_C}$, where $\alpha_z > 0$ and $\alpha_y > 0$ are the factors controlling the element spacing. Hence, the array response $\boldsymbol{a}(\varphi_I, \theta_I, f_I, f_C)$ depends on the frequency of the incident signal and the design frequency of the RIS. Consider the diagonal reflection matrix of the RIS as $\boldsymbol{\Psi} = \text{diag}\left(e^{j\psi_1}, \ldots, e^{j\psi_N}\right)$, where $e^{j\psi_n}$ is

the response of the n-th unit modulus element with the phase response denoted by $\psi_n \in \mathcal{F}_1 = [0, 2\pi)$ for the continuous case and $\psi_n \in \mathcal{F}_2 = \{0, \Delta, \ldots, (B-1)\Delta\}$ for the b-bits quantized RIS. $\Delta = \frac{2\pi}{B}$ denotes the quantization level and $B = 2^b$.

The normalized beamforming gain of the RIS at an observation elevation angle of φ_O, and observation azimuth angle of θ_O is then given by [43]:

$$\begin{aligned} u(\varphi_O, \theta_O, \boldsymbol{\Psi}, \rho) &= \frac{1}{N}\left|\boldsymbol{a}^T(\varphi_O, \theta_O, f_I, f_C)\boldsymbol{\Psi}\boldsymbol{a}(\varphi_I, \theta_I, f_I, f_C)\right| \\ &= \frac{1}{N}\left|\sum_{n_y=0}^{N_y-1}\sum_{n_z=0}^{N_z-1} e^{j\frac{2\pi}{\rho}(n_z\alpha_z\zeta_{I,O}+n_y\alpha_y\xi_{I,O})}e^{j\psi_n}\right|, \end{aligned} \tag{6.2}$$

where $\rho = \frac{\lambda_I}{\lambda_C} = \frac{f_C}{f_I}$, $\zeta_{I,O} = \sin\varphi_I + \sin\varphi_O$, $\xi_{I,O} = \sin\theta_I\cos\varphi_I + \sin\theta_O\cos\varphi_O$, and $n = n_y + n_z N_y + 1$.

In the following optimization problem, our objective is to identify all elevation and azimuth angle pairs $\{(\varphi_O^*, \theta_O^*)\}$ at which the RIS beamforming gain attains its maximum.

$$\{(\varphi_O^*, \theta_O^*)\} = \arg\max_{\varphi_O, \theta_O} u(\varphi_O, \theta_O, \boldsymbol{\Psi}, \rho), \tag{6.3}$$

where $\boldsymbol{\Psi}$ is calculated by substituting $\rho = 1$ in the phases of the complex exponential given in (7.11) and equating them to a constant phase ϕ_0 as shown below, thus maximizing the beamforming gain:

$$2\pi(n_z\alpha_z\zeta_{I,D} + n_y\alpha_y\xi_{I,D}) + \psi_{n_y,n_z} \equiv \phi_{0,0}, \forall n_y, n_z, \tag{6.4}$$

where $\phi_{0,0} \equiv \phi_0 \pmod{2\pi}$, which ensures that the phases of all complex exponential terms in (7.11) are aligned to ϕ_0. Also, $\zeta_{I,D}$ and $\xi_{I,D}$ are obtained based on (φ_D, θ_D), which denotes the design reflection angles of the RIS at frequency f_C. In particular,

when $f_I = f_C$, i.e., $\rho = 1$, the maximum beamforming gain occurs at $(\varphi_D, \theta_D) \in \{(\varphi_O^*, \theta_O^*)\}$.

In the subsequent sections, we aim to solve the problem in (6.3), where $\mathbf{\Psi} \in \mathcal{F}_1$ corresponds to the continuous-phase setting and is obtained directly from (6.4), and $\mathbf{\Psi} \in \mathcal{F}_2$ corresponds to the quantized-phase setting, where $\mathbf{\Psi}$ is obtained by applying nearest-neighbor quantization to (6.4).

6.3 Beam-squint and Beam-split Existence and Characterization: Continuous RIS Phase Profiles

In this section, we first derive the necessary and sufficient conditions to maximize the beamforming gain. We then derive the squint/split angle pairs for the continuous RIS phase profile case, i.e., $\mathbf{\Psi} \in \mathcal{F}_1$, and analyze the beam-squint/split existence conditions for both elevation and azimuth directions.

6.3.1 Beam-squint and Beam-split Characterization

We aim to identify all sets of observation angles $\{(\varphi_O^*, \theta_O^*)\}$ for which the maximum beamforming gain is achieved when the RIS phase profiles are fixed continuous values and designed for f_C, but the frequency of the incident signal $f_I \neq f_C$, i.e., $\rho \neq 1$. The total phase at element (n_y, n_z) becomes:

$$\phi_{n_y,n_z} = \frac{2\pi}{\rho}(n_z \alpha_z \zeta_{I,O} + n_y \alpha_y \xi_{I,O}) + \psi_{n_y,n_z}. \tag{6.5}$$

By plugging (6.4) into (6.5), the total phase is obtained by:

$$\begin{aligned}\phi_{n_y,n_z} &= \frac{2\pi}{\rho}\left(n_z\alpha_z\zeta_{I,O} + n_y\alpha_y\xi_{I,O}\right) + \phi_{0,0} - 2\pi\left(n_z\alpha_z\zeta_{I,D} + n_y\alpha_y\xi_{I,D}\right) \\ &= 2\pi\left[n_z\alpha_z\left(\frac{1}{\rho}\zeta_{I,O} - \zeta_{I,D}\right) + n_y\alpha_y\left(\frac{1}{\rho}\xi_{I,O} - \xi_{I,D}\right)\right] + \phi_{0,0}, \end{aligned} \tag{6.6}$$

Next, to solve (6.3) in this scenario, we first present the following Lemma.

Lemma 6.1. *Using the triangle inequality, the beamforming gain can be bounded as follows:*

$$\tilde{u} = \left|\sum_{n_y=0}^{N_y-1}\sum_{n_z=0}^{N_z-1} e^{j\phi_{n_y,n_z}}\right| \leq \sum_{n_y=0}^{N_y-1}\sum_{n_z=0}^{N_z-1}\left|e^{j\phi_{n_y,n_z}}\right| = N_yN_z,$$

where $\tilde{u} = Nu$, and equality holds iff $\phi_{n_y+1,n_z} - \phi_{n_y,n_z} = 2\pi m_y$ and $\phi_{n_y,n_z+1} - \phi_{n_y,n_z} = 2\pi m_z$, for some integer values of $m_y, m_z \in \mathbb{Z}$; i.e., any difference between two neighboring phases must be an integer multiple of 2π.

Proof. See Section 6.8.1. □

The equality conditions in Lemma 6.1 can be rewritten using (6.6) as follows:

$$m_z = \frac{\phi_{n_y,n_z+1} - \phi_{n_y,n_z}}{2\pi} = \alpha_z\left(\frac{1}{\rho}\zeta_{I,O} - \zeta_{I,D}\right),\ m_z \in \mathbb{Z}, \tag{6.7}$$

$$m_y = \frac{\phi_{n_y+1,n_z} - \phi_{n_y,n_z}}{2\pi} = \alpha_y\left(\frac{1}{\rho}\xi_{I,O} - \xi_{I,D}\right),\ m_y \in \mathbb{Z}. \tag{6.8}$$

Substituting $\zeta_{I,O} = \sin\varphi_I + \sin\varphi_O$ and $\zeta_{I,D} = \sin\varphi_I + \sin\varphi_D$ into (6.7), and then solving for φ_O, we obtain:

$$\varphi_O^* = \arcsin\left(\rho\left(\zeta_{I,D} + \frac{m_z}{\alpha_z}\right) - \sin\varphi_I\right), \tag{6.9}$$

Moreover, substituting $\xi_{I,O} = \sin\theta_I \cos\varphi_I + \sin\theta_O \cos\varphi_O$ and $\xi_{I,D} = \sin\theta_I \cos\varphi_I + \sin\theta_D \cos\varphi_D$ into (6.8), and given φ_O^* from (6.9) and defining $\cos\varphi_O^* = \sqrt{1-\sin^2\varphi_O^*} > 0$, θ_O^* is obtained as follows:

$$\theta_O^* = \arcsin\left(\frac{\rho\left(\xi_{I,D} + \frac{m_y}{\alpha_y}\right) - \sin\theta_I\cos\varphi_I}{\sqrt{1-\left(\rho\left(\zeta_{I,D}+\frac{m_z}{\alpha_z}\right)-\sin\varphi_I\right)^2}}\right). \tag{6.10}$$

Note that we can identify the *beam-squint* by setting $m_z = 0$ and $m_y = 0$ in (6.9) and (6.10), respectively.

6.3.2 Beam-Squint/Split Existence Conditions (Elevation Angles)

It is important to identify the conditions under which beam-squint and beam-split occur, given the frequency ratio ρ and the element spacing factors α_y and α_z. The feasibility conditions can be obtained by limiting the arguments of $\arcsin(\cdot)$ within the interval of $[-1, 1]$. Let X_z denotes the argument of the $\arcsin(\cdot)$ function in (6.9). By enforcing the condition $|X_z| \leq 1$ and solving for m_z, we obtain the following integer set of feasible peaks, i.e.,

$$\mathcal{M}_z = \{m_z \in \mathbb{Z} : L_z \leq m_z \leq U_z\}, \tag{6.11}$$

where

$$L_z = \frac{\alpha_z}{\rho}\left(\sin\varphi_I - 1 - \rho\zeta_{I,D}\right), \tag{6.12}$$

$$U_z = \frac{\alpha_z}{\rho}\left(\sin\varphi_I + 1 - \rho\zeta_{I,D}\right). \tag{6.13}$$

Therefore, for a fixed pair (φ_I, φ_D), the set of feasible elevation indices (m_z) consists of the integers contained within the real interval $[L_z, U_z]$. The length of the interval is given by $U_z - L_z = \frac{2\alpha_z}{\rho}$, and its center is given as $\frac{L_z+U_z}{2} = \frac{\alpha_z}{\rho}\left(\sin\varphi_I - \rho\left(\sin\varphi_I + \sin\varphi_D\right)\right)$. Based on this analysis, we establish the following feasibility criteria:

- If $\frac{2\alpha_z}{\rho} < 1$, the interval is shorter than 1, so for any fixed (φ_I, φ_D), there can be *at most* one integer m_z in $[L_z, U_z]$ which represents beam-squint. However, *no beam-split* occurs in the elevation direction. Moreover, if the center of the interval $[L_z, U_z]$ lies strictly between two consecutive integers, then $\mathcal{M}_z = \emptyset$, and neither beam-squint nor beam-split can occur.

- If $1 \leq \frac{2\alpha_z}{\rho} < 2$, the interval length is *at least* 1 and less than 2, so for every (φ_I, φ_D), there is at least one integer m_z guaranteed in the interval which is referred to as beam-squint. On the other hand, depending on the center, there may be two integers within the interval. In this case, both *beam-split* and beam-squint occur.

- If $\frac{2\alpha_z}{\rho} \geq 2$: the interval always contains at least two integers, regardless of its center, so both beam-squint and beam-split always occur.

We can determine the existence of beam-squint regardless of α_z and (φ_I, φ_D) by substituting $m_z = 0$ in (6.9). The argument of the $\arcsin(\cdot)$ function is then given by $X_z^{(0)} = (\rho - 1)\sin\varphi_I + \rho\sin\varphi_D$. Applying the triangle inequality to $X_z^{(0)}$, we obtain:

$$\left|X_z^{(0)}\right| \leq |\rho - 1||\sin\varphi_I| + \rho|\sin\varphi_D| \leq |\rho - 1| + \rho. \tag{6.14}$$

In (6.14), if $\rho < 1$, then $|\rho - 1| + \rho = 1$. Consequently, for all (φ_I, φ_D), we have

$\left|X_z^{(0)}\right| \leq 1$, and therefore $m_z = 0 \in \mathcal{M}_z$. On the other hand, for $\rho > 1$, beam-squint in elevation occurs ($m_z = 0 \in \mathcal{M}_z$) if the angle pair (φ_I, φ_D) satisfies $\left|X_z^{(0)}\right| \leq 1$.

6.3.3 Beam-Squint/Split Existence Conditions (Azimuth Angles)

To analyze the feasibility of beam-squint and beam-split in the azimuth direction, we assume that a feasible peak in the elevation direction exists for some m_z. Let us define the numerator of the argument of arcsin(·) in (6.10) as:

$$X_y = \rho \left(\sin\theta_I \cos\varphi_I + \sin\theta_D \cos\varphi_D + \frac{m_y}{\alpha_y} \right) - \sin\theta_I \cos\varphi_I,$$

and let $A_O = \cos\varphi_O^* = \sqrt{1 - \sin^2\varphi_O^*}$, where $A_O \in (0, 1]$. Hence, the argument of the arcsin(·) function in (6.10) becomes $\frac{X_y}{A_O}$, and the azimuth angle θ_O^* is feasible if $\frac{|X_y|}{A_O} \leq 1$.

Solving $\frac{|X_y|}{A_O} \leq 1$ for m_y, we obtain the following integer set of feasible peaks, i.e.,

$$\mathcal{M}_y = \{m_y \in \mathbb{Z} : L_y \leq m_y \leq U_y\}, \tag{6.15}$$

with

$$L_y = \frac{\alpha_y}{\rho} \left(\sin\theta_I \cos\varphi_I - \rho\xi_{I,D} - A_O \right), \tag{6.16}$$

$$U_y = \frac{\alpha_y}{\rho} \left(\sin\theta_I \cos\varphi_I - \rho\xi_{I,D} + A_O \right). \tag{6.17}$$

Therefore, for fixed pairs of (φ_I, φ_D), (θ_I, θ_D), and for a feasible φ_O^*, the set of feasible azimuth indices (m_y) consists of the integers within the real interval $[L_y, U_y]$. The length of this interval is $U_y - L_y = \frac{2\alpha_y}{\rho} A_O$, and its center is given by $\frac{L_y + U_y}{2} = \frac{\alpha_y}{\rho} (\sin\theta_I \cos\varphi_I - \rho\xi_{I,D})$.

Using $A_O \in (0, 1]$ and the length of the interval, we establish the following feasibility criteria:

- If $\frac{2\alpha_y}{\rho} A_O < 1 \; \forall \varphi_O^*$, *at most* one feasible solution exists which is referred to as beam-squint, and beam-split never occurs. Depending on the center of the interval $[L_y, U_y]$, it is possible to have $\mathcal{M}_y = \emptyset$, and therefore, neither beam-squint nor beam-split occurs.

- If $1 \leq \frac{2\alpha_y}{\rho} A_O < 2$, the length of the interval is *at least* 1 and less than 2, so beam-split may occur depending on the center of the interval. Depending on the center, there may be two integers within the interval. In this case, both *beam-split* and beam-squint occur.

- If $\frac{2\alpha_y}{\rho} A_O \geq 2$, the interval contains at least two integers, so both beam-squint and beam-split always occur.

The feasibility of the main azimuth for $m_y = 0$ requires $\frac{\left|X_y^{(0)}\right|}{A_O} \leq 1$, where $X_y^{(0)} = (\rho - 1) \sin\theta_I \cos\varphi_I + \rho \sin\theta_D \cos\varphi_D$. This condition implies that $\left|X_y^{(0)}\right| \leq A_O = \cos\varphi_O^*$, which depends on all angles as well as ρ. Unlike in the elevation case, even when $\rho \leq 1$, azimuth squint is not guaranteed to occur, since A_O can be strictly less than 1.

Remark: For the continuous RIS phase profile, Fig. 6.2 illustrates an example of the beam-squint and beam-split angle pairs obtained by the proposed framework, together with the actual peaks obtained by exhaustive search. This framework utilizes feasibility analysis based on $\mathcal{M}_z$ and $\mathcal{M}_y$ to limit the search space over the integer pairs (m_z, m_y).

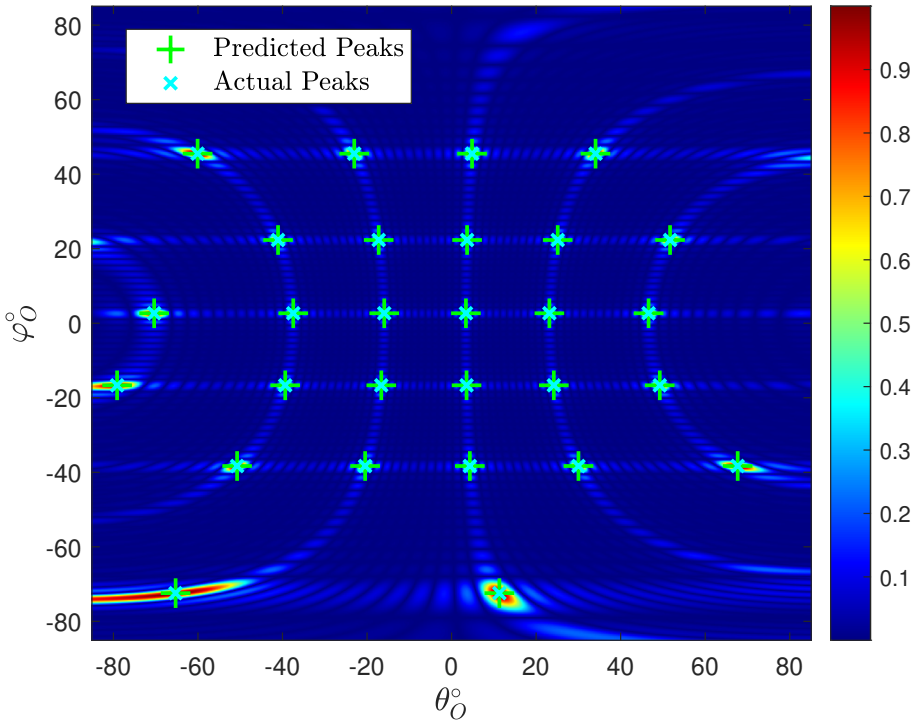


Figure 6.2: Continuous RIS beampattern under frequency mismatch for $\rho = 0.5$, $\alpha_y = \alpha_z = 1.5$, $N_y = 12$, $N_z = 12$, with incidence $(\varphi_I, \theta_I) = (-30°, -10°)$ and design direction $(\varphi_D, \theta_D) = (-24°, 44°)$.

6.4 Beam-squint and Beam-split Existence and Characterization: b-bit RIS Phase Profiles

In practice, the phase shifts of the RIS elements are typically not continuous, thus cannot take arbitrary values from the set $\mathcal{F}_1$. Therefore, in this section, we solve the problem in (6.3) considering b-bit quantized RIS, i.e., for $\boldsymbol{\Psi} \in \mathcal{F}_2$.

In the continuous-phase case, the per-element phase is exactly affine in (n_y, n_z), so the array sum achieves its maximum only when all unit-step phase increments are integer multiples of 2π (as given in **Lemma 6.1**). In contrast, under b-bit quantization, each element incurs an additional RIS phase shift error $\varepsilon(\psi_{n_y,n_z}) \in [-\Delta/2, \Delta/2)$,

which depends on the RIS element index (n_y, n_z) and is no longer affine in (n_y, n_z). As a result, **Lemma 6.1** can no longer be utilized directly on the quantized array sum.

Subsequently, in this section, considering the Fourier series of the RIS quantization error phasor, we reformulate the beam pattern as a superposition of harmonic beam patterns and then obtain the closed-form beam-squint/split angle pairs that maximize each harmonic. We then analyze the dominance ordering of the harmonics and study the existence of per-harmonic beam-squint/split.

6.4.1 Beam Pattern with b-bit Quantized RIS

Considering b-bit nearest-neighbor quantization, we define the quantization function, which takes the continuous phase (with $\rho = 1$) from (6.4) as input, as follows:

$$Q_b(\psi_{n_y,n_z}) = \Delta \left\lfloor \frac{\psi_{n_y,n_z}}{\Delta} \right\rceil, \tag{6.18}$$

where $\lfloor x \rceil$ denotes the *round* function.

We then define the RIS quantization phase error function as follows:

$$\varepsilon(\psi_{n_y,n_z}) = Q_b(\psi_{n_y,n_z}) - \psi_{n_y,n_z} = \Delta\left(\left\lfloor \frac{\psi_{n_y,n_z}}{\Delta} \right\rceil - \frac{\psi_{n_y,n_z}}{\Delta} \right). \tag{6.19}$$

Accordingly, the beam pattern S is obtained as:

$$\begin{aligned} S &= \sum_{n_y=0}^{N_y-1} \sum_{n_z=0}^{N_z-1} e^{j\left(\frac{2\pi}{\rho}(\alpha_z \zeta_{I,O} n_z + \alpha_y \xi_{I,O} n_y) + Q_b(\psi_{n_y,n_z})\right)} \\ &= \sum_{n_y=0}^{N_y-1} \sum_{n_z=0}^{N_z-1} e^{j\left(2\pi(\delta_z n_z + \delta_y n_y) + \varepsilon(\psi_{n_y,n_z})\right)}, \end{aligned} \tag{6.20}$$

where $\delta_z = \alpha_z \left(\frac{1}{\rho}\zeta_{I,O} - \zeta_{I,D}\right)$ and $\delta_y = \alpha_y \left(\frac{1}{\rho}\xi_{I,O} - \xi_{I,D}\right)$ are obtained by substituting $Q_b(\psi_{n_y,n_z})$ from (6.18) and performing some algebraic manipulations.

In particular, we define the RIS quantization error phasor by $g(x) = e^{j\varepsilon(x)}$, with $\varepsilon(x) = \Delta(\lfloor x \rceil - x)$, and for ease of exposition, we define x by:

$$x = \frac{\psi_{n_y,n_z}}{\Delta} = -B(n_z\alpha_z\zeta_{I,D} + n_y\alpha_y\xi_{I,D}) + x_0, \tag{6.21}$$

where $x_0 = \frac{\phi_0}{\Delta}$. However, as we show below, the RIS quantization error phasor function has Fourier series representation. Expanding this function into its harmonics decomposes the quantization effect into a weighted sum of array responses with *purely linear* phases. For each harmonic, the phase is again affine in (n_y, n_z), thus **Lemma 6.1** becomes applicable on a per-harmonic basis.

Since we aim to utilize the Fourier series of $g(x)$ in our analysis, it is necessary to show that the Fourier series of $g(x)$ exists and converges. This is given in **Lemma-2**.

Lemma 6.2. *The function $g : \mathbb{R} \to \mathbb{C}$, defined as $g(x) = e^{j\Delta(\lfloor x \rceil - x)}$, is 1-periodic and of bounded variation over one period. Moreover, its Fourier series converges pointwise to $\frac{1}{2}(g(x^+) + g(x^-))$, where $g(x^-)$ and $g(x^+)$ denote the left- and right-hand limits of g at x, respectively.*

Proof. See Section 6.8.2. □

Next, we obtain the Fourier series of $g(x)$ in the following.

Lemma 6.3 (Fourier series of g)**.** *Let $g(x) = e^{j\,\Delta(\lfloor x \rceil - x)}$ with period 1. For harmonic $\ell \in \mathbb{Z}$, we define the Fourier series coefficients as follows:*

$$\widehat{g}[\ell] = \int_0^1 g(x)\, e^{-j2\pi\ell x}\, dx. \quad \textit{Then} \quad \widehat{g}[\ell] = \frac{2(-1)^\ell \sin\left(\frac{\Delta}{2}\right)}{\Delta + 2\pi\ell}$$

Consequently, the Fourier series can be given as:

$$g(x) \sim \sum_{\ell \in \mathbb{Z}} \widehat{g}[\ell]\ e^{j2\pi\ell x}, \tag{6.22}$$

where $g(x)$ *converges point-wise to* $\frac{1}{2}\big(g(x^-) + g(x^+)\big)$.

Proof. See Section 6.8.3. □

6.4.2 Beam Pattern Transformation and Per-harmonic Peaks

At this point, we transform the beam pattern in (6.20) by substituting the Fourier series of the error phasor function $g(x)$, as given in (6.22), and obtain:

$$\begin{aligned}\widehat{S} &= \sum_{n_y=0}^{N_y-1} \sum_{n_z=0}^{N_z-1} e^{j\left(2\pi(\delta_z n_z + \delta_y n_y) + \varepsilon(\psi_{n_y,n_z})\right)} \\ &= \sum_{\ell \in \mathbb{Z}} \widehat{g}[\ell] \sum_{n_y=0}^{N_y-1} \sum_{n_z=0}^{N_z-1} e^{j2\pi(\delta_z n_z + \delta_y n_y)} e^{j2\pi\ell x}.\end{aligned} \tag{6.23}$$

where $x = \frac{\psi_{n_y,n_z}}{\Delta}$ is taken from (6.21). Therefore, $\widehat{S}$ becomes:

$$\widehat{S} = \sum_{\ell \in \mathbb{Z}} S_\ell = \sum_{\ell \in \mathbb{Z}} \widehat{g}[\ell] e^{j2\pi\ell x_0} \sum_{n_y=0}^{N_y-1} \sum_{n_z=0}^{N_z-1} e^{j2\pi(\beta_z(\ell) n_z + \beta_y(\ell) n_y)}, \tag{6.24}$$

where $\beta_z(\ell) = \delta_z - \ell B \alpha_z \zeta_{I,D}$ and $\beta_y(\ell) = \delta_y - \ell B \alpha_y \xi_{I,D}$.

Using (6.24), we first obtain the per-harmonic peak locations. To do so, we define the absolute value of the ℓ-th harmonic in (6.24) as given below:

$$T_\ell = |\widehat{g}[\ell]| \left| \sum_{n_y=0}^{N_y-1} \sum_{n_z=0}^{N_z-1} e^{j2\pi(\beta_z(\ell) n_z + \beta_y(\ell) n_y)} \right|. \tag{6.25}$$

$$\theta_O^*(\ell, m_z, m_y) = \arcsin\left(\frac{\rho\left(\xi_{I,D} + \frac{1}{\alpha_y}(m_y + \ell B\alpha_y\xi_{I,D})\right) - \sin\theta_I\cos\varphi_I}{\sqrt{1 - \left(\rho\left(\zeta_{I,D} + \frac{1}{\alpha_z}(m_z + \ell B\alpha_z\zeta_{I,D})\right) - \sin\varphi_I\right)^2}}\right). \tag{6.29}$$

Since the total phases are now affine in (n_y, n_z) for a given ℓ, by applying **Lemma** 6.1, the term T_ℓ is maximized if and only if the following set of equations holds:

$$\beta_z(\ell) = \delta_z - \ell B\alpha_z\zeta_{I,D} = m_z, \quad m_z \in \mathbb{Z}, \tag{6.26}$$

$$\beta_y(\ell) = \delta_y - \ell B\alpha_y\xi_{I,D} = m_y, \quad m_y \in \mathbb{Z}. \tag{6.27}$$

For a fixed ℓ, solving (6.26) for φ_O yields:

$$\varphi_O^*(\ell, m_z) = \arcsin\left(\rho\left(\zeta_{I,D} + \frac{1}{\alpha_z}(m_z + \ell B\alpha_z\zeta_{I,D})\right) - \sin\varphi_I\right), \tag{6.28}$$

and, given $\varphi_O^*(\ell, m_z)$, solving (6.27) for θ_O results in the optimal azimuth angle given at the top of the page in (6.29). From (6.28) and (6.29), we obtain the elevation and azimuth angles $\varphi_O^*(\ell, m_z)$ and $\theta_O^*(\ell, m_z, m_y)$ corresponding to the ℓ-th harmonic for each integer pair (m_z, m_y), respectively.

6.4.3 Existence of Per Harmonic Beam-Squint/Split

To analyze the existence of per-harmonic peaks, we first restrict attention to a finite set of "dominant" harmonics whose relative strength exceeds a threshold η, and then carry out the feasibility analysis only for ℓ in this set. Define the relative strength of the ℓ-th harmonic relative to $\ell = 0$ as:

$$\kappa_\ell = \frac{|\widehat{g}[\ell]|}{|\widehat{g}[0]|} = \frac{\Delta}{|\Delta + 2\pi\ell|} = \frac{1}{|1 + B\ell|}.^{1} \tag{6.30}$$

Let $\eta \in (0, 1]$ be a relative strength threshold. Then, the set of harmonics whose relative strength compared to $\ell = 0$ is at least η is given by:

$$\mathcal{K}_\eta = \{\ell \in [-K_-, K_+] \cap \mathbb{Z}\}, \tag{6.31}$$

where $K_+ = \left\lfloor (-1 + \frac{1}{\eta})B^{-1} \right\rfloor$ is obtained by solving $\kappa_\ell \geq \eta$ for $\ell \geq 0$, and $K_- = \left\lfloor (1 + \frac{1}{\eta})B^{-1} \right\rfloor$ is obtained by solving $\kappa_\ell \geq \eta$ for $\ell < 0$. Here, $\lfloor \cdot \rfloor$ denotes the floor function.

Now, for all selected harmonics $\ell \in \mathcal{K}_\eta$, we can apply the feasibility analysis described in Section III. In particular, the lower and upper bounds of the intervals containing feasible integers in the elevation direction, analogous to (6.12) and (6.13), for the ℓ-th harmonic are given by:

$$L_z(\ell) = \frac{\alpha_z}{\rho} \left(\sin \varphi_I - 1 - \rho \zeta_{I,D} \left(\ell B + 1 \right) \right), \ \forall \ell \in \mathcal{K}_\eta \tag{6.32}$$

$$U_z(\ell) = \frac{\alpha_z}{\rho} \left(\sin \varphi_I + 1 - \rho \zeta_{I,D} \left(\ell B + 1 \right) \right), \ \forall \ell \in \mathcal{K}_\eta \tag{6.33}$$

such that $L_z(0) = L_z$ and $U_z(0) = U_z$. In the ℓ-th interval, an integer m_z exists if $\lceil L_z(\ell) \rceil \leq \lfloor U_z(\ell) \rfloor$. If this condition is satisfied, then every integer $m_z \in \mathcal{M}_z(\ell) = [\lceil L_z(\ell) \rceil, \lfloor U_z(\ell) \rfloor]$ yields a feasible elevation angle $\varphi_O^*(\ell, m_z)$.

Assuming that $\varphi_O^*(\ell, m_z)$ exists such that $A_O(\ell) = \cos \varphi_O^*(\ell, m_z) > 0$, the lower and upper bounds of the interval containing feasible azimuth integers, similar to (6.16)

[1]For 1-bit quantization, we have $\kappa_0 = \kappa_{-1} > \kappa_\ell$ for all $\ell \in \mathbb{Z} \backslash \{-1, 0\}$, whereas for $b \geq 2$-bit quantization, $\kappa_0 > \kappa_\ell$ for all $\ell \in \mathbb{Z} \backslash \{0\}$.

and (6.17), for the ℓ-th harmonic are obtained by:

$$L_y(\ell) = \frac{\alpha_y}{\rho}(\sin\theta_I \cos\varphi_I - \rho\xi_{I,D}(\ell B + 1) - A_O(\ell)), \forall \ell \in \mathcal{K}_\eta \tag{6.34}$$

$$U_y(\ell) = \frac{\alpha_y}{\rho}(\sin\theta_I \cos\varphi_I - \rho\xi_{I,D}(\ell B + 1) + A_O(\ell)), \forall \ell \in \mathcal{K}_\eta \tag{6.35}$$

with $L_y(0) = L_y$ and $U_y(0) = U_y$. Similarly, in the ℓ-th interval, an integer m_y exists if $\lceil L_y(\ell) \rceil \leq \lfloor U_y(\ell) \rfloor$. If this condition holds, then every integer $m_y \in \mathcal{M}_y(\ell) = [\lceil L_y(\ell) \rceil, \lfloor U_y(\ell) \rfloor]$ yields a feasible azimuth angle $\theta_O^*(\ell, m_z, m_y)$. Further analysis on the feasibility of beam-squint and beam-split of the per-harmonic peak locations can then be carried out, analogous to the continuous-phase case.

6.5 Per-harmonic beam-squint/Split Angle Correction Mechanism

From (6.24), it is evident that the quantized beam pattern $\widehat{S}$ is expressed as the superposition of infinitely many harmonics. In particular, it can be observed in Fig. 6.3 that the superposition of multiple harmonics and their respective split beams can displace the actual per-harmonic peak positions resulting from the aggregate beam pattern S. To address this, we propose a correction method that refines the predicted peak locations by accounting for the interference introduced by other harmonics. Then, we analyze how the number of RIS elements and the frequency ratio can affect the accuracy of the quantized peak locations.

Considering the fact that multiplying a constant and mapping of $a \to a^2$ do not change the arg max of the function in (6.3), we define the aggregate beam power pattern as follows:

$$J(\varphi_O, \theta_O) = \left| \widehat{S}(\varphi_O, \theta_O) \right|^2, \tag{6.36}$$

and for a particular harmonic $\ell \in \mathbb{Z}$, define $J_\ell(\varphi_O, \theta_O) = |S_\ell(\varphi_O, \theta_O)|^2$. Moreover, for harmonic $\ell \in \mathcal{K}_\eta$ and integer pair $(m_z, m_y) \in \mathcal{M}_z(\ell) \times \mathcal{M}_y(\ell)$, we define

$$\boldsymbol{p} = \begin{bmatrix} \varphi_O^* (\ell, m_z) \\ \theta_O^* (\ell, m_z, m_y) \end{bmatrix},$$

which is the main lobe center for this harmonic and integer pair. At point $\boldsymbol{p}$, the *isolated* harmonic power has a local maximum, i.e., the gradient $\nabla J_\ell(\boldsymbol{p}) = 0$, and its Hessian is negative definite. The reason is that the necessary and sufficient conditions of maximum peaks in Lemma 6.1 are applied to obtain $\boldsymbol{p}$ for the ℓ-th harmonic.

Now, we aim to estimate the actual maximum of J in a neighborhood of $\boldsymbol{p}$, which is further closer to the actual location of this specific lobe once all harmonics are superposed.

6.5.1 Decomposition of Harmonic of Interest and Perturbation

First, we decompose the aggregate beam power pattern into the main harmonic of interest at point $\boldsymbol{p}$, denoted by $J_\ell(\boldsymbol{p})$, and the perturbation, which is the contribution of all other harmonics, denoted by $R_\ell(\boldsymbol{p})$, i.e.,

$$\begin{aligned} J(\boldsymbol{p}) = J_\ell(\boldsymbol{p}) + R_\ell(\boldsymbol{p}) &= \left| \sum_\ell S_\ell(\boldsymbol{p}) \right|^2 \\ &= J_\ell(\boldsymbol{p}) + \underbrace{\sum_{i \neq \ell} |S_i(\boldsymbol{p})|^2 + 2 \sum_{k<r} \Re \{ S_k(\boldsymbol{p}) S_r^*(\boldsymbol{p}) \}}_{R_\ell(\boldsymbol{p})}. \end{aligned} \tag{6.37}$$

We assume that the perturbation $R_\ell(\boldsymbol{p})$ is locally smaller than $J_\ell(\boldsymbol{p})$, so we can approximate the gradient of J using a first-order approximation as given in **Lemma 6.4**.

Lemma 6.4 (First-order approximation of a vector-valued function)**.** *Let* $\boldsymbol{F} : \mathbb{R}^n \to \mathbb{R}^m$ *be a vector-valued function that is differentiable at point* $\boldsymbol{a} \in \mathbb{R}^n$*, and let* $\boldsymbol{F}'(\boldsymbol{a}) \in \mathbb{R}^{m\times n}$ *denote its Jacobian matrix at* $\boldsymbol{a}$*. Then, for any vector* $\boldsymbol{h} \in \mathbb{R}^n$*, we have:* $\boldsymbol{F}(\boldsymbol{a}+\boldsymbol{h}) = \boldsymbol{F}(\boldsymbol{a})+\boldsymbol{F}'(\boldsymbol{a})\,\boldsymbol{h}+\boldsymbol{r}(\boldsymbol{h})$*, where the remainder* $\boldsymbol{r}(\boldsymbol{h}) \in \mathbb{R}^m$ *satisfies* $\lim_{\boldsymbol{h}\to\boldsymbol{0}} \frac{\|\boldsymbol{r}(\boldsymbol{h})\|}{\|\boldsymbol{h}\|} = 0$ *Hence,* $\boldsymbol{F}$ *admits the first-order approximation:* $\boldsymbol{F}(\boldsymbol{a}+\boldsymbol{h}) \approx \boldsymbol{F}(\boldsymbol{a}) + \boldsymbol{F}'(\boldsymbol{a})\boldsymbol{h}$ *with small error for small* $\boldsymbol{h}$*.*

Proof. This follows directly from the definition of differentiability and Theorems 2.4 and 4.1 of [142]. □

Now, in order to estimate the actual peak location, denoted by $\bar{\boldsymbol{p}}$, of the main lobe for harmonic ℓ near $\boldsymbol{p}$, we define the shift vector as $\Delta\boldsymbol{p} = \bar{\boldsymbol{p}} - \boldsymbol{p} = \begin{bmatrix} \Delta\varphi \\ \Delta\theta \end{bmatrix}$. Note that $\bar{\boldsymbol{p}}$ is a local maximum of the total beam power pattern J and therefore satisfies $\nabla J(\bar{\boldsymbol{p}}) = 0$. Assuming that the actual peak $\bar{\boldsymbol{p}}$ lies close to the nominal peak $\boldsymbol{p}$ of harmonic ℓ, we linearize the vector function $\nabla J(\bar{\boldsymbol{p}})$ around the point $\boldsymbol{p}$.

By applying **Lemma** 6.4 to $\nabla J(\bar{\boldsymbol{p}})$, we have:

$$\nabla J(\boldsymbol{p} + \Delta\boldsymbol{p}) \approx \nabla J(\boldsymbol{p}) + \boldsymbol{H}_J(\boldsymbol{p})\Delta\boldsymbol{p}, \tag{6.38}$$

where $\boldsymbol{H}_J$ is the Hessian of J, i.e., Jacobian matrix of ∇J. Moreover, $\nabla J(\boldsymbol{p}) = \nabla R_\ell(\boldsymbol{p})$ since $\nabla J_\ell(\boldsymbol{p}) = 0$. Solving (6.38) by using $\boldsymbol{H}_J(\boldsymbol{p})$ directly to find the angle shift $\Delta\boldsymbol{p}$ corresponds to a Newton step on the stationary condition $\nabla J(\bar{\boldsymbol{p}}) = 0$. However, since J is highly nonconvex and contains many extrema, this step can move the estimate toward side-lobes or even null directions.

6.5.2 Local Surrogate Function-based Optimization

Instead of using Newton step by using the Hessian of J, we maximize a local surrogate that preserves the dominant harmonic's negative-definite curvature while treating the residual R_ℓ as a linear perturbation. Since $\boldsymbol{p}$ is the isolated maximizer of J_ℓ, we have $\nabla J_\ell(\boldsymbol{p}) = \mathbf{0}$. We approximate

$$J(\boldsymbol{p} + \Delta\boldsymbol{p}) \approx \tilde{J}(\Delta\boldsymbol{p})$$

where $J(\boldsymbol{p} + \Delta\boldsymbol{p}) = J_\ell(\boldsymbol{p} + \Delta\boldsymbol{p}) + R_\ell(\boldsymbol{p} + \Delta\boldsymbol{p})$ and

$$\tilde{J}(\Delta\boldsymbol{p}) \triangleq J_\ell(\boldsymbol{p}) + \tfrac{1}{2}\Delta\boldsymbol{p}^T \boldsymbol{H}_{J_\ell}(\boldsymbol{p})\Delta\boldsymbol{p} + R_\ell(\boldsymbol{p}) + \nabla R_\ell^T(\boldsymbol{p})\Delta\boldsymbol{p}, \tag{6.39}$$

which is second order in J_ℓ and first order in R_ℓ around $\boldsymbol{p}$. Since $\boldsymbol{H}_{J_\ell}(\boldsymbol{p}) \prec 0$, $\tilde{J}(\Delta\boldsymbol{p})$ is strictly concave in $\Delta\boldsymbol{p}$, and setting $\nabla_{\Delta\boldsymbol{p}}\tilde{J} = \mathbf{0}$ yields:

$$\Delta\boldsymbol{p} = -\boldsymbol{H}_{J_\ell}^{-1}(\boldsymbol{p})\nabla R_\ell(\boldsymbol{p}). \tag{6.40}$$

Thus, (6.40) is the unique maximizer of this local surrogate, using the dominant harmonic's curvature rather than the full Hessian of J. The expressions of the gradient and the Hessian used in (6.40) are presented in Section 6.8.4.

The accuracy of the per-harmonic peak locations as well as peak shift-estimate in (6.40) depend on the assumption that the curvature of the harmonic ℓ, i.e., $\boldsymbol{H}_{J_\ell}(\boldsymbol{p})$, is dominant. It can be shown that $\boldsymbol{H}_{J_\ell}(\boldsymbol{p})$ scales on the order of

$$|\widehat{g}[\ell]|^2 \frac{1}{\rho^2} N_y^2 N_z^2 \left(\alpha_z^2(N_z^2 - 1) + \alpha_y^2(N_y^2 - 1)\right)$$

up to bounded trigonometric factors and a generally nonzero off-diagonal terms $J^{(\ell)}_{\varphi\theta}$. Therefore, increasing N_z and N_y strengthens the negative definiteness of $\boldsymbol{H}_{J_\ell}(\boldsymbol{p})$ and reduces the sensitivity of the actual peak locations to the interferences from other harmonics. By comparing Fig. 6.3 and Fig. 6.4, we note that increasing the number of RIS elements improves the accuracy of the per-harmonic peak locations, as well as that of the corrected peaks, by increasing the curvature of J at the per-harmonic peak locations. Moreover, since $\boldsymbol{H}_{J_\ell}(\boldsymbol{p})$ scales on the order of $1/\rho^2$, increasing the frequency ratio ρ reduces the magnitude of the per-harmonic curvature and makes the first-order shift approximation less reliable. In particular, since (6.40) involves $\boldsymbol{H}^{-1}_{J_\ell}(\boldsymbol{p})$, a smaller-curvature Hessian can yield a larger shift $\Delta\boldsymbol{p}$ which undermines the local assumption that $\Delta\boldsymbol{p}$ is small, which is required for the first-order approximation of the perturbation term $R_\ell(\boldsymbol{p}+\Delta\boldsymbol{p})$ around $\boldsymbol{p}$.

Finally, Fig. 6.5 illustrates the beampattern of a 1-bit RIS with many beam-splits, all of which are detected by the proposed framework. Similarly to the continuous case, we utilize feasibility analysis to limit the search over the integer pairs (m_z, m_y) for each harmonic $\ell \in \{-1, 0\}$.

6.5.3 Computational Complexity Analysis

A direct beamforming gain peak search evaluates the beampattern (or $J = |S|^2$) over a 2D grid $\mathcal{G}_\varphi \times \mathcal{G}_\theta$ with sizes G_φ and G_θ. Since each evaluation of $S(\varphi, \theta)$ requires summing over all $N = N_y N_z$ RIS elements, the exhaustive-scan complexity scales as $\mathsf{Cost}_{\text{sweep}} = O(G_\varphi G_\theta N)$, which can be prohibitive for fine angular grids (large G_φ and G_θ) and large RIS sizes. In contrast, the complexity of our proposed framework is based on the following two factors.

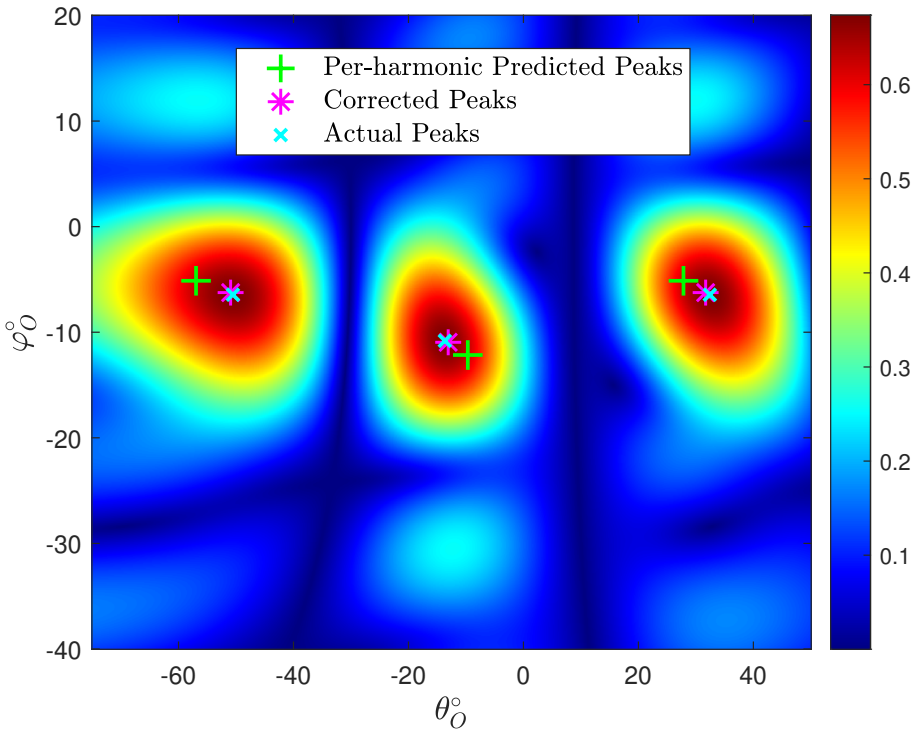


Figure 6.3: Quantized RIS beampattern under frequency mismatch for $\rho = 0.65$, $\alpha_y = \alpha_z = 0.5$, $N_y = 4$, $N_z = 6$, and $b = 1$, with incidence $(\varphi_I, \theta_I) = (-30°, -10°)$ and design direction $(\varphi_D, \theta_D) = (-24°, 44°)$.

6.5.3.1 Per-harmonic peak complexity

For each $\ell \in \mathcal{K}_\eta$, feasible peak indices satisfy $m_z \in \mathcal{M}_z(\ell) = \mathbb{Z} \cap [L_z(\ell), U_z(\ell)]$ and $m_y \in \mathcal{M}_y(\ell) = \mathbb{Z} \cap [L_y(\ell), U_y(\ell)]$. Using $|\mathbb{Z} \cap [L, U]| \leq (U - L) + 2$ and the interval lengths $U_z(\ell) - L_z(\ell) = \frac{2\alpha_z}{\rho}$ and $U_y(\ell) - L_y(\ell) = \frac{2\alpha_y}{\rho} A_O(\ell) \leq \frac{2\alpha_y}{\rho}$, we obtain $|\mathcal{M}_z(\ell)| = O(\alpha_z/\rho + 1)$ and $|\mathcal{M}_y(\ell)| = O(\alpha_y/\rho + 1)$ $\forall \ell \in \mathcal{K}_\eta$. Hence, the total number of per-harmonic candidate peaks satisfies:

$$\sum_{\ell \in \mathcal{K}_\eta} |\mathcal{M}_z(\ell)| \, |\mathcal{M}_y(\ell)| = O\Big(|\mathcal{K}_\eta| \left(\tfrac{\alpha_z}{\rho} + 1 \right) \left(\tfrac{\alpha_y}{\rho} + 1 \right) \Big).$$

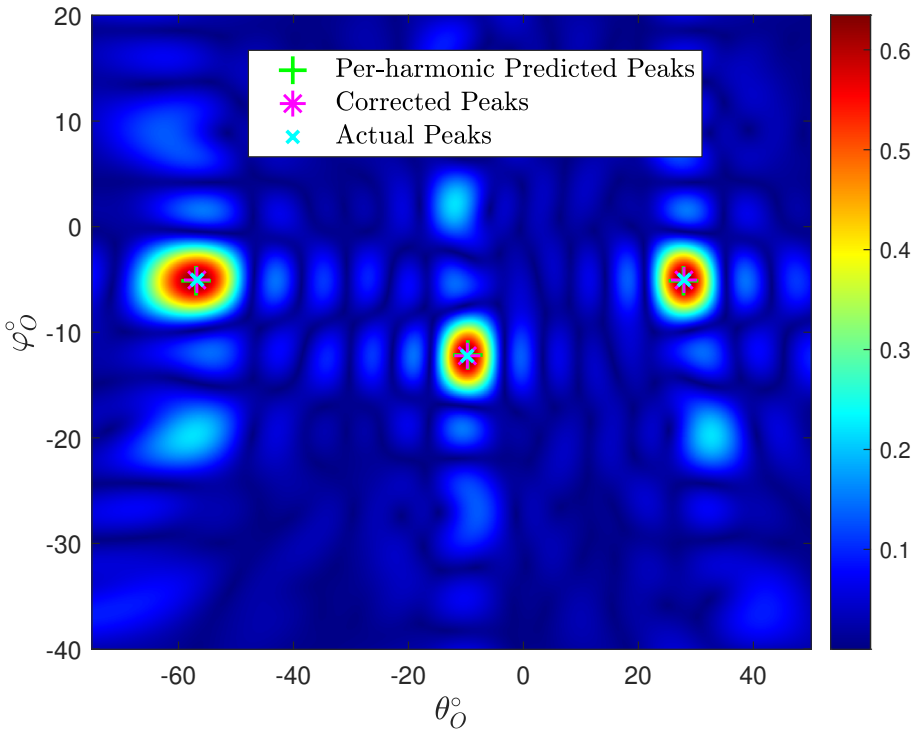


Figure 6.4: Quantized RIS beampattern under frequency mismatch for $\rho = 0.65$, $\alpha_y = \alpha_z = 0.5$, $N_y = 12$, $N_z = 16$, and $b = 1$, with incidence $(\varphi_I, \theta_I) = (-30°, -10°)$ and design direction $(\varphi_D, \theta_D) = (-24°, 44°)$.

Each candidate angle pair is obtained in constant time from (6.28)–(6.29). Therefore, the total complexity is obtained by:

$$\mathsf{Cost}_{\text{har}} = O\Big(|\mathcal{K}_\eta| \left(\tfrac{\alpha_z}{\rho} + 1\right) \left(\tfrac{\alpha_y}{\rho} + 1\right)\Big). \tag{6.41}$$

6.5.3.2 Perturbation-based correction complexity

The correction in (6.40) is applied to each nominal per-harmonic candidate. In implementation, the total field and its derivatives are evaluated using a finite truncation set of harmonics $\mathcal{K}_c$ (e.g., $\ell \in [-L_c, L_c]$), with cardinality $|\mathcal{K}_c|$. Thus, for each candidate

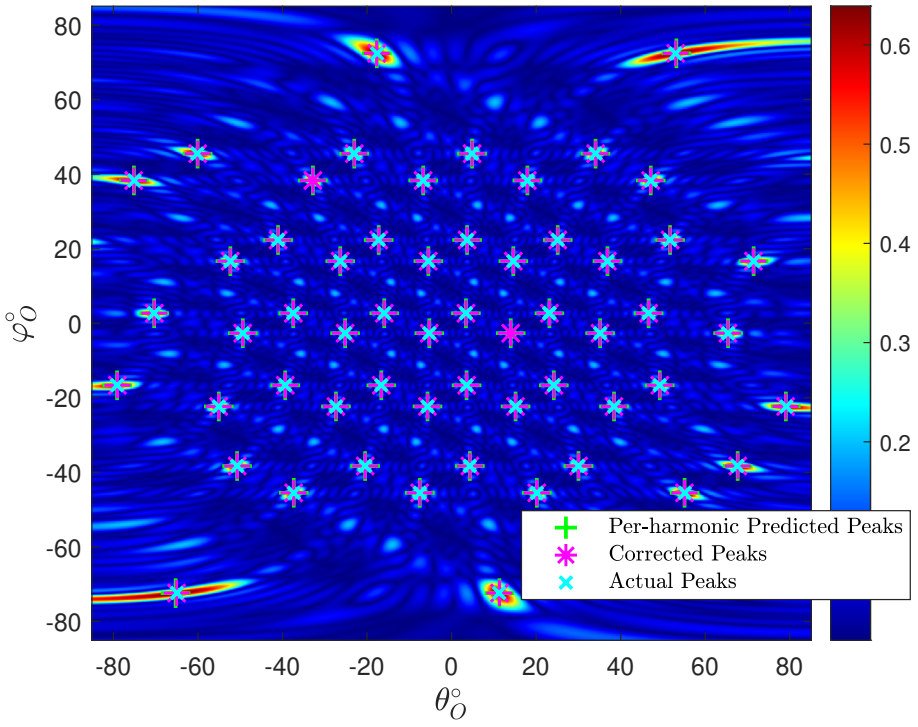


Figure 6.5: Quantized RIS beampattern under frequency mismatch for $\rho = 0.5$, $\alpha_y = \alpha_z = 1.5$, $N_y = 12$, $N_z = 12$, and $b = 1$, with incidence $(\varphi_I, \theta_I) = (-30°, -10°)$ and design direction $(\varphi_D, \theta_D) = (-24°, 44°)$.

point, the correction requires evaluating $O(|\mathcal{K}_c|)$ harmonic contributions. Moreover, each harmonic term admits a separable 2D form (a product of a length-N_z sum and a length-N_y sum), so its field/derivative evaluation in Appendix D can be computed using two 1D sums, leading to $O(N_z + N_y)$ operations per harmonic (rather than $O(N_y N_z)$). Consequently, the correction complexity scales as:

$$\mathsf{Cost}_{\text{corr}} = O\Big(|\mathcal{K}_\eta| \left(\tfrac{\alpha_z}{\rho} + 1\right) \left(\tfrac{\alpha_y}{\rho} + 1\right) |\mathcal{K}_c| \, (N_z + N_y)\Big), \tag{6.42}$$

which avoids the cost of exhaustive search by limiting evaluation to a small candidate

set indexed by feasible integer pairs.

6.6 Use-Case Study: Beam-Split Aware Scheduling in Multi-RIS, Multi-User, Multi-band Network

In this section, to demonstrate the effectiveness of the obtained closed-form per-harmonic beam-squint/split angles in (6.28)–(6.29), we consider a two-BS, two-RIS, two-UE downlink transmission scenario in an MBN, as illustrated in Fig. 6.1. User Equipment (UE) 1 is served on frequency band f_1 and UE 2 on frequency band f_2, with $\rho_{12} = \frac{f_1}{f_2}$, $\rho_{21} = \frac{f_2}{f_1}$, and $\lambda_k = \frac{c}{f_k},\ k \in \{1, 2\}$. RIS 1 is configured at f_1 and RIS 2 at f_2. UE 1 is considered as the *home* UE of RIS 1 and the *cross* UE of RIS 2, and vice versa. Specifically, we assume that the direct paths between the BSs and the users are blocked, and we denote the diagonal phase responses of RIS 1 and RIS 2 by $\mathbf{\Psi}_1$ and $\mathbf{\Psi}_2$, respectively. Let us define the cascaded channel gain between the k-th BS and the i-th UE through the r-th RIS as $h_{k,i}(\mathbf{\Psi}_r)$, where $\{k, i, r\} \in \{1, 2\}$. Then, considering the orthogonality of the different frequency bands, the sum-rate is computed as follows:

$$R = \log_2\left(1 + \frac{P_1|h_{1,1}(\mathbf{\Psi}_1) + h_{1,1}(\mathbf{\Psi}_2)|^2}{\sigma_1}\right) + \log_2\left(1 + \frac{P_2|h_{2,2}(\mathbf{\Psi}_2) + h_{2,2}(\mathbf{\Psi}_1)|^2}{\sigma_2}\right), \tag{6.43}$$

where P_1 and P_2 denote the transmit powers at BS 1 and BS 2, respectively, and σ_1 and σ_2 are the additive noise powers at UE 1 and UE 2, respectively. Our objective is to maximize the sum-rate while exploiting each RIS to serve both UEs. In particular, the sum-rate depends on $|h_1(\mathbf{\Psi}_1, \mathbf{\Psi}_2)|^2 = |h_{1,1}(\mathbf{\Psi}_1) + h_{1,1}(\mathbf{\Psi}_2)|^2$ for UE 1 and $|h_2(\mathbf{\Psi}_1, \mathbf{\Psi}_2)|^2 = |h_{2,2}(\mathbf{\Psi}_2) + h_{2,2}(\mathbf{\Psi}_1)|^2$ for UE 2. Therefore, improving the *cross*

terms $h_{1,1}(\mathbf{\Psi}_2)$ and $h_{2,2}(\mathbf{\Psi}_1)$ (mismatched-band contribution) can directly increase the received power by enabling more constructive combining in the coherent sum of the cascaded channels through both RISs.

While (6.4) can be instantiated for any continuous steering direction (φ_D, θ_D), practical beam management relies on selecting a profile from a finite RIS codebook to enable codeword-based beam training with limited overhead [143]. We therefore discretize the direction-cosine domain into a grid of resolution $Q_{\min}$, which controls the tradeoff between angular snapping error and the number of codewords evaluated. This direction discretization is distinct from the b-bit element-level phase quantization in (6.18).

In what follows, we describe the design of a beam-split-aware RIS phase-shift codebook with a given resolution $Q_{\min}$ and the corresponding codeword selection. We then present the associated simulation results in the next section. Moreover, for ease of exposition, we adopt generic notation for the codebook design and codeword selection method for a given RIS.

6.6.1 Quantized codebooks at $\rho = 1$ (home and cross incidence)

For each RIS, we construct a finite set of b-bit phase profiles $\{\mathbf{\Psi}_q\}$ at $\rho = 1$ by instantiating the continuous-phase design in (6.4) for a grid of *design* directions (φ_D, θ_D) and then applying the elementwise quantizer in (6.18). To do so, we first parameterize directions using the direction-cosines implied by (7.11), $s_{z,D} = \sin\varphi_D$, $s_{y,D} = \cos\varphi_D \sin\theta_D$, and sample $(s_{z,D}, s_{y,D})$ on a $Q_{\min} \times Q_{\min}$ uniform lattice, where $Q_{\min} \in \mathbb{Z}_+$ determines the resolution of sampling.

Hence, for $|s_{z,D}| \leq 1$ and $|s_{y,D}| \leq \sqrt{1 - s_{z,D}^2}$, we have $\varphi_D = \arcsin(s_{z,D})$ and

$\theta_D = \arcsin\left(\frac{s_{y,D}}{\cos\varphi_D}\right)$. For each (φ_D, θ_D), we plug (φ_I, θ_I) and (φ_D, θ_D) into the phase-alignment rule (6.4), quantize using (6.18), and form the unit-modulus profile $\mathbf{\Psi}_q = \exp(j\boldsymbol{\psi}_q)$, which corresponds to the q-th codeword.

We then precompute *two* codebooks per RIS: a *home* codebook built using the home BS→RIS incidence angles, and a *cross* codebook built using the cross BS→RIS incidence angles. This separation is necessary because the incidence angles appear inside the phase-slope terms in (6.4), so changing the BS incidence changes the entire quantized phase pattern associated with a given (φ_D, θ_D). As a result, the home and cross-incidence cases generally induce different discrete mappings $(\varphi_D, \theta_D) \mapsto \mathbf{\Psi}_q$, and both are required by the two-RIS two-BS geometry.

6.6.2 Beam-split-aware Codeword Selection

For each RIS, the cross-assistance goal is to find a *small* set of quantized phase profiles (designed at $\rho = 1$ using the cross-incidence codebook) whose *mismatched* dominant lobe aligns with the cross-user direction at the other band. We avoid exhaustive (φ, θ) sweeps at $\rho \neq 1$ by using the per-harmonic peak-family structure in (6.28)–(6.29). Consider one RIS with known incidence angles (φ_I, θ_I) for the cross link (other BS→RIS). Let (φ_O, θ_O) denote the *desired* cross-user direction (RIS→cross UE). For each tuple (ℓ, m_z, m_y) with $\ell \in \mathcal{K}_\eta$, $m_z \in \mathcal{M}_z(\ell)$ and $m_y \in \mathcal{M}_y(\ell)$, we compute a *design* direction (φ_D, θ_D) at $\rho = 1$ such that the ℓ-th harmonic peak at mismatch ρ coincides with (φ_O, θ_O). Then, assuming $\rho = \rho_{12}$ for RIS 1 evaluated at f_2 (or $\rho = \rho_{21}$ for RIS 2 evaluated at f_1), using the peak-family equations used to obtain

(6.28)–(6.29) and solving for (φ_D, θ_D) gives

$$\sin\varphi_D(\ell, m_z) = \frac{1}{A_\ell}\Big(\frac{\zeta_{I,O}}{\rho} - \frac{m_z}{\alpha_z}\Big) - \sin\varphi_I, \tag{6.44}$$

$$\sin\theta_D(\ell, m_z, m_y) = \frac{1}{\cos\varphi_D}\Bigg(\frac{1}{A_\ell}\Big(\frac{\xi_{I,O}}{\rho} - \frac{m_y}{\alpha_y}\Big) - \sin\theta_I\cos\varphi_I\Bigg), \tag{6.45}$$

where $A_\ell = 1 + \ell B$, $\varphi_D = \arcsin(\sin\varphi_D)$ and $\theta_D = \arcsin(\sin\theta_D)$ are accepted only when $|\sin\varphi_D| \le 1$, $\cos\varphi_D > 0$, and $|\sin\theta_D| \le 1$ hold. Because the RIS can only realize codebook profiles, each feasible (φ_D, θ_D) is mapped to the nearest codeword index in the precomputed *cross* codebook (built at $\rho = 1$ with incidence (φ_I, θ_I)). Using the direction cosines already used for codebook construction, i.e., $s_{z,D}$ and $s_{y,D}$, we select

$$q = \arg\min_{q'} \Big[(s_{z,D} - s_z(q'))^2 + (s_{y,D} - s_y(q'))^2\Big], \tag{6.46}$$

where $(s_z(q'), s_y(q'))$ are the stored direction cosines of codeword q' in the cross codebook. Since multiple tuples (ℓ, m_z, m_y) can map to the same q, we keep only one representative tuple per q by re-evaluating the *mismatched* peak that the selected codeword produces. More specifically, codeword q at design angles $(\varphi_D(q), \theta_D(q))$ produces a peak at (φ_P, θ_P) at mismatch ρ. We then measure the residual misalignment to the desired cross-user direction (φ_O, θ_O) via $d_{\text{cross}}(q, \ell, m_z, m_y) \triangleq d\big((\varphi_P, \theta_P), (\varphi_O, \theta_O)\big)$, where $d(\cdot,\cdot)$ is the angular distance. Next, we must prune to a *small* cross-aware candidate set of size L_{keep} so that the final two-RIS joint selection remains tractable in terms of time complexity. If we rank candidates only by

$$d^\star_{\text{cross}}(q) \triangleq \min_{\ell, m_z, m_y} d_{\text{cross}}(q, \ell, m_z, m_y),$$

then for moderate $Q_{\min}$ and small L_{keep} the shortlist can become dominated by profiles that are excellent for the cross peak but correspond to design directions far from the home steering direction. To bias the shortlist toward candidates that remain compatible with home steering while still enabling cross assistance, we rank codewords using the following weighted angular distance:

$$\mathcal{L}(q) \triangleq d^{\star}_{\text{cross}}(q) + w_{\text{home}}\, d_{\text{home}}(q), \tag{6.47}$$

where $d_{\text{home}}(q) \triangleq d\big((\varphi_D(q), \theta_D(q)), (\varphi_{O,\text{home}}, \theta_{O,\text{home}})\big)$, and $w_{\text{home}} \in [0, 1]$ is the home-priority weight. We then keep the L_{keep} smallest $\mathcal{L}(q)$ codewords as the cross-aware candidate set. Finally, we also include the home-maximizing codeword (q_{home}) from the agnostic selection, which only considers the home UE per RIS, as a robust fall-back. Algorithm 5 summarizes the per-RIS beam-split-aware candidate generation and pruning used to build a small cross-assistance shortlist without exhaustive angular sweeps at $\rho \neq 1$. We run Algorithm 5 independently at RIS 1 and RIS 2 (with the appropriate ρ for cross-band evaluation). Then, we select the operating pair that maximizes the sum-rate in (6.43) over the candidate sets of codewords generated in Algorithm 5.

6.7 Numerical Results and Discussions

In this section, we validate the closed-form expressions and the proposed correction by comparing predicted peak angles with exhaustive 2D angular search, and by quantifying split probabilities and rate gains in the MBN use case. Unless stated otherwise, angles are drawn uniformly in the direction-cosine domain over $\varphi, \theta \in [-80°, 80°]$ (to avoid degenerate cases), and Monte-Carlo averaging is performed with 5×10^3

Algorithm 5 Beam-split-aware Codeword Selection in Multi-band RIS-assisted Networks

Define $\zeta_{I,O}$ and $\xi_{I,O}$ as in (7.11) using $O = (\varphi_O, \theta_O)$.
Initialize $d^{\star}_{\text{cross}}(q) \leftarrow +\infty$ for all q.
for $\ell \in \mathcal{K}_\eta$, and $(m_z, m_y) \in \mathcal{M}_z(\ell) \times \mathcal{M}_y(\ell)$ **do**
    Compute (φ_D, θ_D) via (6.44)–(6.45).
    $(s_{z,D}, s_{y,D}) \leftarrow (\sin\varphi_D,\ \cos\varphi_D \sin\theta_D)$ and snap $q \leftarrow \arg\min_{q'}(s_{z,D} - s_z(q'))^2 + (s_{y,D} - s_y(q'))^2$.
    Predict (φ_P, θ_P) for codeword q at mismatch ρ using (6.28)–(6.29).
    $d^{\star}_{\text{cross}}(q) \leftarrow \min\Big\{ d^{\star}_{\text{cross}}(q),\ d\big((\varphi_P, \theta_P), (\varphi_O, \theta_O)\big) \Big\}$.
end for
$d_{\text{home}}(q) \leftarrow d\big((\varphi_D(q), \theta_D(q)), (\varphi_{O,\text{home}}, \theta_{O,\text{home}})\big)$.
$\mathcal{L}(q) \leftarrow d^{\star}_{\text{cross}}(q) + w_{\text{home}}\, d_{\text{home}}(q)$ and $\mathcal{Q} \leftarrow$ indices of the L_{keep} smallest $\mathcal{L}(q)$.
$\mathcal{Q} \leftarrow \mathcal{Q} \cup \{q_{\text{home}}\}$.

trials per point. For quantized results, nearest-neighbor b-bit phase quantization is applied, and only the dominant harmonics in $\mathcal{K}_\eta$ ((6.31) with $\eta > \frac{1}{3}$) are used for peak characterization. Moreover, in the benchmarks, "Cont" refers to the continuous RIS phase profile case, and "b-bit no corr" and "b-bit corr" refer to RIS phase profiles with b-bit quantization, without and with applying the proposed correction method, respectively.

6.7.1 Beam-Squint/Split Analytical Validation

Fig. 6.6 shows the mean angular error between the predicted peak locations and the true peaks obtained by exhaustive search as a function of the frequency ratio ρ, for two element spacings $\alpha = \alpha_z = \alpha_y \in \{0.5, 0.85\}$. In the continuous-phase case, the error is essentially zero for all ρ because the RIS array sum has an affine phase progression across the UPA, and the derived integer-slope peak conditions are necessary and sufficient; thus, the closed-form peak families coincide with the maximizers found by the exhaustive search.

Under 1-bit quantization, exact phase affinity is lost and the beampattern becomes a superposition of harmonic array responses. Consequently, the maximizer of the total pattern can deviate from the per-harmonic predicted peak locations, yielding a non-zero error that generally increases with ρ. This trend is consistent with the reduction of dominant-harmonic curvature as ρ grows, which increases sensitivity to interference from the remaining harmonics. The proposed correction alleviates this by applying a perturbation-based shift that accounts for the residual-harmonic gradient while leveraging the dominant-harmonic curvature, resulting in consistently smaller errors than the uncorrected estimates.

Increasing the spacing from $\alpha = 0.5$ to $\alpha = 0.85$ reduces the error for both quantized predictors by increasing the electrical aperture at f_I, which sharpens the mainlobe and increases curvature near the maximum, making the peak location less sensitive to multi-harmonic perturbations.

6.7.2 Beam-split Existence (Elevation and Azimuth)

Figs. 6.7–6.8 verify the feasibility-based split criteria in Section III. As shown in Fig. 6.7, elevation split occurs when the feasible-integer set in (6.11) contains at least two integers, i.e., $|\mathcal{M}_z| \geq 2$. It can be observed that the split probability is essentially zero below $\frac{2\alpha_z}{\rho} = 1$, becomes one above $\frac{2\alpha_z}{\rho} = 2$, and transitions smoothly in the intermediate region $1 \leq \frac{2\alpha_z}{\rho} < 2$ where the outcome depends on the random interval center induced by (φ_I, φ_D). Fig. 6.8 shows that, conditioned on a valid elevation peak, azimuth split occurs when the feasible-integer set in (6.15) satisfies $|\mathcal{M}_y| \geq 2$. In contrast to elevation, the azimuth feasibility-interval length scales with $A_O = \cos\varphi_O^*$, so the effective length is reduced by the elevation projection and varies

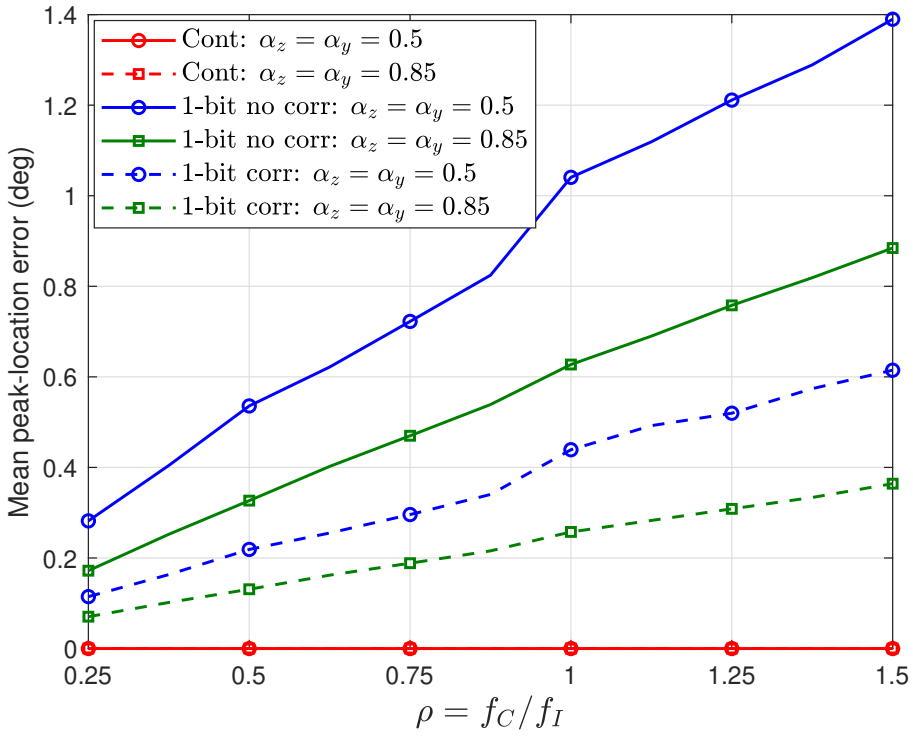


Figure 6.6: Mean error between the predicted peaks and the actual peaks for both continuous and quantization cases as a function of frequency ratio ρ for different element spacing, and $N_z = N_y = 10$.

across trials. This shifts the high-split region to a larger α_y. The overlaid contours indicate where $\frac{2\alpha_y}{\rho} A_O$ exceeds the boundaries 1 and 2 for the feasibility interval in Section III.C with probability one (dashed) and in expectation (solid), under the same conditioning.

6.7.3 Number of RIS Elements in b-bit Quantization

Fig. 6.9 plots the mean peak-location error versus the RIS size $N = N_y N_z$ for 1-bit and 2-bit phase shift, with and without the proposed correction, at $\rho \in \{0.75, 1\}$. For all quantized cases, the error decreases with N because enlarging the aperture

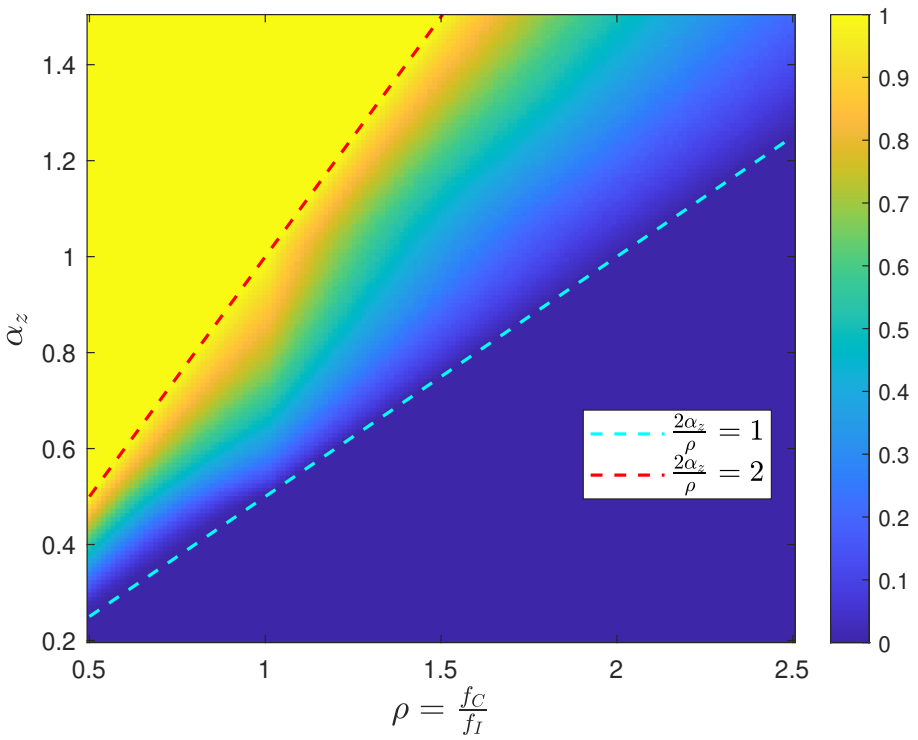


Figure 6.7: Elevation split probability $\mathbb{P}(|\mathcal{M}_z| \geq 2)$ versus (ρ, α_z), where $\mathcal{M}_z$ is the feasible-integer set in (6.11); dashed lines mark $\frac{2\alpha_z}{\rho} = 1$ and $\frac{2\alpha_z}{\rho} = 2$.

sharpens the array factor around each peak (higher curvature), so the true maximizer of the *total* quantized pattern becomes less displaced from the analytically predicted peak locations, consistent with the discussion in Section V. It is observed that the uncorrected predictor exhibits larger errors for 1-bit than for 2-bit, since coarser quantization yields stronger non-dominant harmonics (larger $|\widehat{g}[\ell]|$ for $\ell \neq 0$), which increases multi-harmonic peak pulling. The proposed correction reduces the error for both bit-depths by adjusting each nominal per-harmonic peak using the proposed perturbation-based method, with the largest gains at small N where the mainlobe is widest. Finally, for a fixed N, the $\rho = 1$ curves are generally above $\rho = 0.75$, indicating stronger harmonic superposition when evaluating at the design frequency.

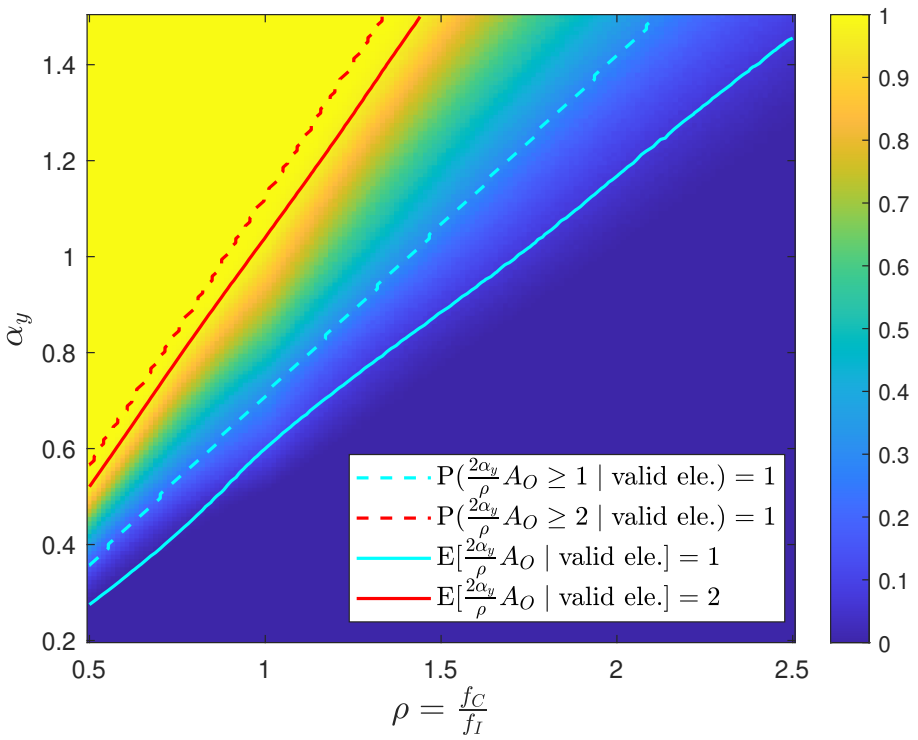


Figure 6.8: Azimuth split probability $\mathbb{P}(|\mathcal{M}_y| \geq 2 \mid \text{valid elevation})$ versus (ρ, α_y), where $\mathcal{M}_y$ is the feasible-integer set in (6.15); contours summarize thresholds of $\frac{2\alpha_y}{\rho} A_O$ under the same conditioning.

6.7.4 Beam-split-aware Scheduling

Fig. 6.10 reports the average two-user sum-rate as the RIS size grows ($N = N_y N_z$), comparing the proposed beam-split-aware selection against a beam-split-agnostic baseline that configures each RIS solely by maximizing its *home* UE gain at its *own* band. All curves increase with $N_y = N_z$ with diminishing returns due to the coherent field combining in (7.11) and the logarithmic rate.

Across all settings, the beam-split-aware method achieves a higher sum-rate because it explicitly searches, via the peak-family structure in (6.28)–(6.29), for cross-incidence codewords whose *mismatched* dominant lobe falls close to the cross-user

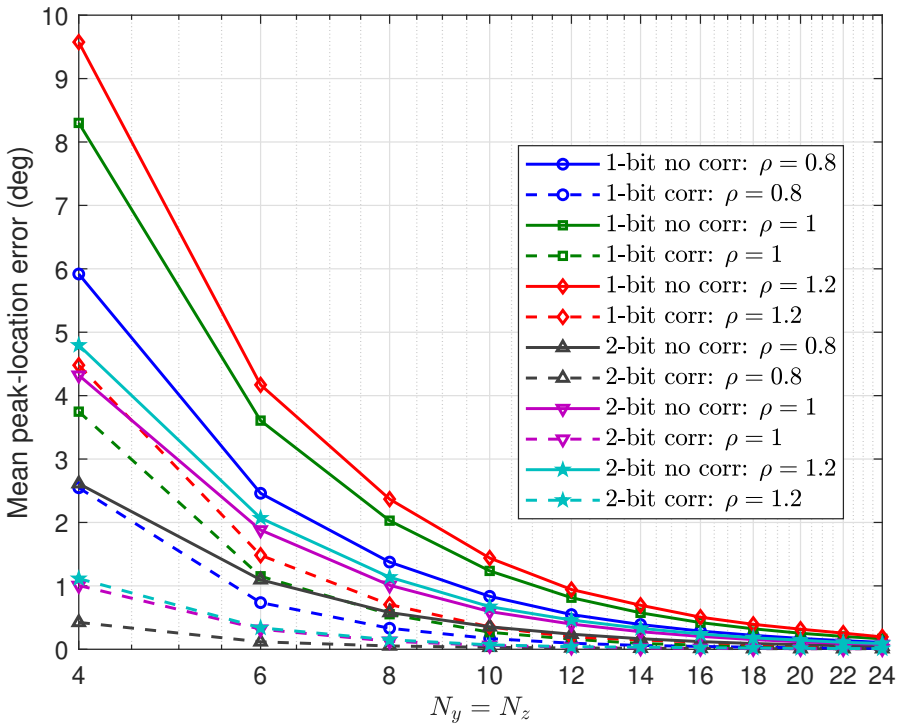


Figure 6.9: Mean error between the predicted peaks and the actual peaks for per-harmonic and corrected peaks versus the number of RIS elements, for different frequency ratio ρ and $\alpha_y = \alpha_z = 0.5$.

direction at the other band; this makes the cross term $h_{1,1}(\mathbf{\Psi}_2)$ and $h_{2,2}(\mathbf{\Psi}_1)$ more reliably strong than in the agnostic baseline, where the cross-band response is incidental. The gain often becomes more pronounced for larger RISs because increasing N narrows the beampattern, making the cross-band field more sensitive to angular misalignment. The beam-split-aware method selects codewords whose mismatched dominant lobe is predicted to lie near the cross-UE direction, whereas the agnostic method optimizes only the home link and does not target the cross-UE, so its cross-band response is more likely to be off-peak as N grows. Increasing the codebook resolution from $Q_{\min} = 8$ to $Q_{\min} = 16$ improves both schemes by reducing the

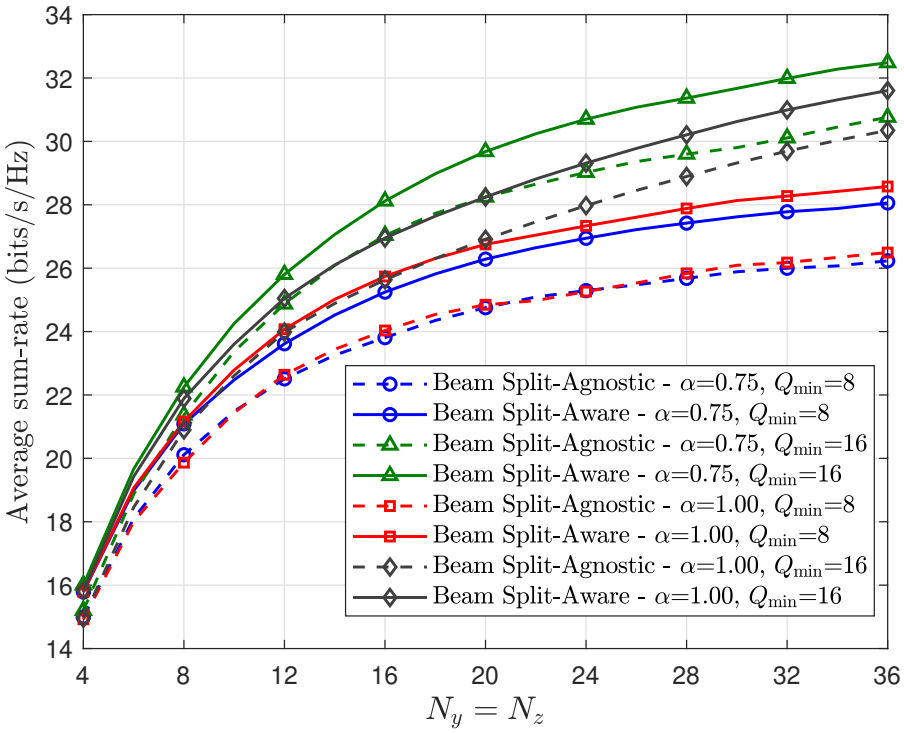


Figure 6.10: Average two-user sum-rate versus RIS size ($N_y = N_z$) in the two-RIS two-UE multi-band network, comparing beam-split-aware and beam-split-agnostic codeword selection for 1-bit quantization, i.e., $\mathcal{K}_\eta = \{-1, 0\}$, $\alpha \in \{0.75, 1\}$ and $Q_{\min} \in \{8, 16\}$.

snapping error when mapping a desired (φ_D, θ_D) to the nearest quantized codeword. Finally, the element spacing factor α changes the achievable rates and the gap by altering the peak/split structure and how effectively the cross-band dominant lobe can be aligned under quantization.

6.8 Complete Proof and Derivations

6.8.1 Proof of Lemma 6.1

The necessity part of Lemma 6.1 follows directly from $\phi_{n_y,n_z} \equiv \phi_{0,0} \pmod{2\pi}$, $\forall n_y, n_z$, since the condition holds for all elements. For the sufficiency, assume there exists fixed integers m_y and m_z such that $\phi_{n_y+1,n_z} - \phi_{n_y,n_z} = 2\pi m_y$ and $\phi_{n_y,n_z+1} - \phi_{n_y,n_z} = 2\pi m_z$ hold for all (n_y, n_z). First, fix $n_z = 0$. The relation for unit-step increments along the y-direction becomes: $\phi_{n_y+1,0} = \phi_{n_y,0} + 2\pi m_y$. Applying this relation repeatedly from $n_y = 0$, after n_y steps we obtain: $\phi_{n_y,0} = \phi_{0,0} + n_y(2\pi m_y)$. Now, fix n_y and start from $n_z = 0$. Applying the unit-step increment along the z-direction yields:

$$\phi_{n_y,n_z} = \phi_{n_y,0} + n_z(2\pi m_z). \tag{6.48}$$

Substituting $\phi_{n_y,0}$ into (6.48), we obtain:

$$\phi_{n_y,n_z} = \phi_{0,0} + 2\pi(n_y m_y + n_z m_z). \tag{6.49}$$

Since $n_y m_y + n_z m_z$ is always an integer, it follows $\phi_{n_y,n_z} \equiv \phi_{0,0} \pmod{2\pi}$, thus proving the sufficiency part of Lemma 6.1.

6.8.2 Proof of Lemma 6.2

Periodicity of $g(x)$: For the round function $\lfloor x \rceil$ and any integer k, we have $\lfloor x \rceil = k$ for $x \in [k - \frac{1}{2}, k + \frac{1}{2})$, with a fixed (but arbitrary) convention at the endpoints $x = k \pm \frac{1}{2}$.

This implies that $\lfloor x+1 \rceil = \lfloor x \rceil + 1$ for all x. Hence,

$$\varepsilon(x+1) = \Delta(\lfloor x+1 \rceil - (x+1)) = \Delta(\lfloor x \rceil - x) = \varepsilon(x),$$

which shows that $\varepsilon(x)$ is 1-periodic, and so is $g(x) = e^{j\varepsilon(x)}$.

Bounded variation of $g(x)$ on one period: Let us define total variation and bounded variation:

Definition 6.1. *The variation $V_{[a,b]}(f)$ of a function $f : [a,b] \to \mathbb{C}$ is $V_{[a,b]}(f) = \sup_{a=x_0<\cdots<x_m=b} \sum_{i=1}^{m} |f(x_i) - f(x_{i-1})|$. A function f is said to be of bounded variation on $[a,b]$ if $V_{[a,b]}(f) < \infty$ [144].*

Now, we show that $g(x)$ is of bounded variation on the period $[0,1]$. We use the following identity, which follows from trigonometric identities and the inequality $|\sin x| \leq |x|$:

$$\left|e^{ju} - e^{jv}\right| = 2\left|\sin\left(\frac{u-v}{2}\right)\right| \leq |u-v|. \tag{6.50}$$

Partition the interval $[0,1]$ at $x = \frac{1}{2}$. For $x \in [0, \frac{1}{2})$, we have $g(x) = e^{-j\Delta x}$, so

$$\begin{aligned}\sum_{x_i \in [0,\frac{1}{2})} |g(x_i) - g(x_{i-1})| &\overset{(a)}{\leq} \sum_{x_i \in [0,\frac{1}{2})} |-\Delta(x_i - x_{i-1})| \\ &\overset{(b)}{=} \left|-\Delta\left(\tfrac{1}{2}^- - 0\right)\right| = \tfrac{1}{2}|\Delta|, \end{aligned} \tag{6.51}$$

where (a) follows from (6.50) and (b) holds because $-\Delta x$ is monotonic. For $x \in (\frac{1}{2}, 1]$, we have $g(x) = e^{j\Delta(1-x)}$, and similarly, $\sum_{x_i \in (\frac{1}{2},1]} |g(x_i) - g(x_{i-1})| \leq \frac{1}{2}|\Delta|$. At the jump point $x = \frac{1}{2}$, we have $g(\frac{1}{2}^-) = e^{-j\frac{\Delta}{2}}$ and $g(\frac{1}{2}^+) = e^{j\frac{\Delta}{2}}$, so $\left|g(\frac{1}{2}^+) - g(\frac{1}{2}^-)\right| = 2\left|\sin\frac{\Delta}{2}\right|$. Summing across all partitions, we obtain: $\sum_{i=1}^{m} |g(x_i) - g(x_{i-1})| \leq |\Delta| + 2\left|\sin\frac{\Delta}{2}\right| \leq$

$2|\Delta| < \infty$. Therefore, $g(x)$ is of bounded variation over the period $[0, 1]$.

Convergence of the Fourier series: To show that the Fourier series of $g(x)$ converges, we use the Dirichlet–Jordan test:

Theorem 6.1 (Dirichlet–Jordan Test)**.** *If a 2π-periodic function $f : \mathbb{R} \to \mathbb{C}$ is of bounded variation over $[0, 2\pi]$, then its Fourier series converges at every point x to $\frac{1}{2}(f(x^+) + f(x^-))$ [145].*

Define 2π-periodic function $G(y) = g\left(\frac{y}{2\pi}\right)$. Then, by applying the Dirichlet–Jordan test to $G(y)$ and re-scaling to $g(x)$, we conclude that the Fourier series of $g(x)$ converges pointwise to $\tilde{g}(x) = \frac{1}{2}(g(x^+) + g(x^-))$. At continuity points, $\tilde{g}(x) = g(x)$, and at the jump points $x = \frac{1}{2} + k$, the sum equals $\tilde{g}(x) = \cos\frac{\Delta}{2}$. This completes the proof of **Lemma** 6.2.

6.8.3 Proof of Lemma 6.3

On $x \in [0, 1]$, we can reformulate $g(x)$ as:

$$g(x) = \begin{cases} e^{-j\Delta x}, & 0 \le x < \frac{1}{2}, \\ e^{j\Delta}\, e^{-j\Delta x}, & \frac{1}{2} \le x \le 1. \end{cases}$$

Fix $\ell \in \mathbb{Z}$ and set $a = \Delta + 2\pi\ell \neq 0$. Then the Fourier series coefficient of $g(x)$ is obtained by:

$$\widehat{g}[\ell] = \int_0^{\frac{1}{2}} e^{-jax} dx + e^{j\Delta} \int_{\frac{1}{2}}^{1} e^{-jax} dx = \frac{1 - e^{-j\frac{a}{2}}}{ja} + e^{j\Delta} \frac{e^{-j\frac{a}{2}} - e^{-ja}}{ja}. \tag{6.52}$$

Since $e^{-ja} = e^{-j(\Delta+2\pi\ell)} = e^{-j\Delta}$,

$$\widehat{g}[\ell] = \frac{e^{-j\frac{a}{2}}\big(e^{j\Delta}-1\big)}{ja} = \frac{2je^{-j\frac{\Delta+2\pi\ell-\Delta}{2}}\sin\frac{\Delta}{2}}{ja} = \frac{2(-1)^{\ell}\sin\frac{\Delta}{2}}{\Delta+2\pi\ell}.$$

Then, using $\widehat{g}[\ell]$, $g(x)$ can be represented by its Fourier series $g(x) \sim \sum_{\ell\in\mathbb{Z}} \widehat{g}[\ell]\; e^{j2\pi\ell x}$. Moreover, followed by **Lemma 6.2**, this Fourier series converges pointwise to $\frac{1}{2}\big(g(x^-)+ g(x^+)\big)$.

6.8.4 Gradient and Hessian of J used in (6.40)

Let $f(\varphi,\theta) \in \mathbb{C}$ and define $J_f(\varphi,\theta) = |f(\varphi,\theta)|^2$. Then

$$J_{f,\varphi} = 2\Re\{f_\varphi f^*\}, \quad J_{f,\theta} = 2\Re\{f_\theta f^*\}.$$

The Hessian entries are: $J_{f,\varphi\varphi} = 2\Re\{f_{\varphi\varphi}f^* + f_\varphi f_\varphi^*\}$, $J_{f,\theta\theta} = 2\Re\{f_{\theta\theta}f^* + f_\theta f_\theta^*\}$, and $J_{f,\varphi\theta} = 2\Re\{f_{\varphi\theta}f^* + f_\varphi f_\theta^*\}$ with $J_{f,\theta\varphi} = J_{f,\varphi\theta}$. Hence, $\boldsymbol{H}_{J_f}(\varphi,\theta) = \begin{bmatrix} J_{f,\varphi\varphi} & J_{f,\varphi\theta} \\ J_{f,\varphi\theta} & J_{f,\theta\theta} \end{bmatrix}$.

With $J = |S|^2$ and $J_\ell = |S_\ell|^2$, we have:

$$\nabla J = 2\begin{bmatrix} \Re\{S_\varphi S^*\} \\ \Re\{S_\theta S^*\} \end{bmatrix}, \qquad \nabla J_\ell = 2\begin{bmatrix} \Re\{S_{\ell,\varphi} S_\ell^*\} \\ \Re\{S_{\ell,\theta} S_\ell^*\} \end{bmatrix},$$

and

$$\boldsymbol{H}_J = 2\begin{bmatrix} \Re\{S_{\varphi\varphi}S^* + S_\varphi S_\varphi^*\} & \Re\{S_{\varphi\theta}S^* + S_\varphi S_\theta^*\} \\ \Re\{S_{\varphi\theta}S^* + S_\varphi S_\theta^*\} & \Re\{S_{\theta\theta}S^* + S_\theta S_\theta^*\} \end{bmatrix},$$

$$\boldsymbol{H}_{J_\ell} = 2\begin{bmatrix} \Re\{S_{\ell,\varphi\varphi}S_\ell^* + S_{\ell,\varphi}S_{\ell,\varphi}^*\} & \Re\{S_{\ell,\varphi\theta}S_\ell^* + S_{\ell,\varphi}S_{\ell,\theta}^*\} \\ \Re\{S_{\ell,\varphi\theta}S_\ell^* + S_{\ell,\varphi}S_{\ell,\theta}^*\} & \Re\{S_{\ell,\theta\theta}S_\ell^* + S_{\ell,\theta}S_{\ell,\theta}^*\} \end{bmatrix}.$$

Using $\nabla R_\ell(\boldsymbol{p}) = \nabla J(\boldsymbol{p})$, we obtain $\nabla R_\ell(\boldsymbol{p})$ using $\nabla R_\ell = \nabla J - \nabla J_\ell$. In particular, at the per-harmonic nominal peak $\boldsymbol{p}$ of J_ℓ, $\nabla J_\ell(\boldsymbol{p}) = \boldsymbol{0}$, and therefore $\nabla R_\ell(\boldsymbol{p}) = \nabla J(\boldsymbol{p})$. The first derivatives of the phase slopes are: $\beta_{z,\varphi} = \frac{\alpha_z}{\rho}\cos\varphi_O$, $\beta_{z,\theta} = 0$, $\beta_{y,\varphi} = -\frac{\alpha_y}{\rho}\sin\theta_O\sin\varphi_O$, and $\beta_{y,\theta} = \frac{\alpha_y}{\rho}\cos\theta_O\cos\varphi_O$. The second derivatives are: $\beta_{z,\varphi\varphi} = -\frac{\alpha_z}{\rho}\sin\varphi_O$, $\beta_{z,\theta\theta} = 0$, $\beta_{z,\varphi\theta} = 0$, $\beta_{y,\varphi\varphi} = -\frac{\alpha_y}{\rho}\sin\theta_O\cos\varphi_O$, $\beta_{y,\theta\theta} = -\frac{\alpha_y}{\rho}\sin\theta_O\cos\varphi_O$, and $\beta_{y,\varphi\theta} = -\frac{\alpha_y}{\rho}\cos\theta_O\sin\varphi_O$. Thus,

$$\Phi_{\ell,\varphi} = 2\pi\big(n_z\beta_{z,\varphi} + n_y\beta_{y,\varphi}\big)$$

$$\Phi_{\ell,\theta} = 2\pi\big(n_z\beta_{z,\theta} + n_y\beta_{y,\theta}\big)$$

$$\Phi_{\ell,\varphi\varphi} = 2\pi\big(n_z\beta_{z,\varphi\varphi} + n_y\beta_{y,\varphi\varphi}\big)$$

$$\Phi_{\ell,\theta\theta} = 2\pi\big(n_z\beta_{z,\theta\theta} + n_y\beta_{y,\theta\theta}\big)$$

and

$$\Phi_{\ell,\varphi\theta} = 2\pi\big(n_z\beta_{z,\varphi\theta} + n_y\beta_{y,\varphi\theta}\big)$$

Using $\partial_x e^{j\Phi} = j\Phi_x e^{j\Phi}$ and $\partial_{xx}e^{j\Phi} = (j\Phi_{xx} - \Phi_x^2)e^{j\Phi}$, $\partial_{xy}e^{j\Phi} = (j\Phi_{xy} - \Phi_x\Phi_y)e^{j\Phi}$,

we obtain:

$$S_{\ell,\varphi} = \widehat{g}[\ell] \sum_{n_z=0}^{N_z-1} \sum_{n_y=0}^{N_y-1} \big(j\Phi_{\ell,\varphi}\big) e^{j\Phi_\ell},$$

$$S_{\ell,\theta} = \widehat{g}[\ell] \sum_{n_z=0}^{N_z-1} \sum_{n_y=0}^{N_y-1} \big(j\Phi_{\ell,\theta}\big) e^{j\Phi_\ell},$$

$$S_{\ell,\varphi\varphi} = \widehat{g}[\ell] \sum_{n_z=0}^{N_z-1} \sum_{n_y=0}^{N_y-1} \big(j\Phi_{\ell,\varphi\varphi} - \Phi_{\ell,\varphi}^2\big) e^{j\Phi_\ell},$$

$$S_{\ell,\theta\theta} = \widehat{g}[\ell] \sum_{n_z=0}^{N_z-1} \sum_{n_y=0}^{N_y-1} \big(j\Phi_{\ell,\theta\theta} - \Phi_{\ell,\theta}^2\big) e^{j\Phi_\ell},$$

$$S_{\ell,\varphi\theta} = \widehat{g}[\ell] \sum_{n_z=0}^{N_z-1} \sum_{n_y=0}^{N_y-1} \big(j\Phi_{\ell,\varphi\theta} - \Phi_{\ell,\varphi}\Phi_{\ell,\theta}\big) e^{j\Phi_\ell}.$$

6.9 Summary

This chapter characterized beam-squint and beam-split in UPA-based RISs operating under the configuration–incident frequency mismatch. For continuous phase, we derived necessary and sufficient peak conditions and explicit elevation–azimuth peak families, together with feasibility criteria governed by frequency ratio, element spacing, and incident–design angle pairs. For practical b-bit quantization, we reformulated the beampattern via a Fourier-series decomposition of the quantization error phasor, yielding closed-form per-harmonic peak families and existence conditions. To bridge isolated per-harmonic characterizations to the maximizer of the superposed pattern, we developed a perturbation-based angle correction that exploits dominant-harmonic curvature and residual-harmonic gradients. Numerical results validated the peak/split characterizations, quantified the impact of frequency ratio, element spacing, RIS size,

and quantization depth on angular displacement, and show that beam-split-aware codeword selection improves sum-rate in a two-RIS two-UE MBN.

Chapter 7

Circuit-aware Performance Optimization in RIS-Assisted Sub-6 GHz and Mid-band Multi-Band Networks

7.1 Overview

Chapter 6 studied RIS-assisted multi-frequency networks from an array-level perspective. In that chapter, the RIS elements were modeled as frequency-independent unit-modulus phase shifters in order to isolate the beam-squint, beam-split, and quantization-induced effects caused by the fixed RIS aperture and element spacing. However, practical RIS elements are also frequency dependent at the hardware level. Their reflection coefficients and achievable phase shifts depend on the incident frequency, the circuit parameters, and the feasible tuning range of the RIS elements.

This chapter complements Chapter 6 by focusing on this element-level frequency dependence. In particular, we study RIS-assisted sub-6 GHz and upper mid-band (UMB) multi-band networks using a practical circuit-based RIS model. Each RIS element is modeled by two inductors, a resistor, and a tunable varactor. This model

allows us to characterize how the feasible capacitance window determines the supported phase-shift range at different incident frequencies. Such characterization is important because, outside the useful tuning region, changing the capacitance may provide little phase control, and the RIS may behave almost like a fixed reflector.

The objective of this chapter is to understand how practical circuit constraints affect RIS-assisted multi-frequency and multi-band operation and to design the RIS capacitances accordingly. Instead of assigning an independent phase shift at each frequency, we optimize a single capacitance vector that jointly serves multiple frequencies. This setting captures the practical constraint that the same RIS hardware configuration must operate over a set of frequencies within the desired band, and can also be evaluated under cross-band operation.

The main contributions of this chapter can be summarized as follows:

• We analytically characterize the supported phase-shift range induced by the feasible capacitance window. This characterization identifies when capacitance tuning provides useful phase control and when the RIS response becomes effectively fixed.

• We formulate a multi-frequency sum array-gain maximization problem, where a single capacitance vector is optimized over a set of incident frequencies for native-band and cross-band operation.

• We develop a projected adaptive moment estimation (ADAM)-based algorithm with an adaptive zero-phase-based initialization that exploits the circuit response of the RIS elements.

• Numerical results show that the proposed adaptive zero-phase-based initialization achieves higher gain than random-initialized ADAM and a genetic-algorithm baseline in the considered native-band cases. The results also show that native-band

and cross-band performance are governed by whether the feasible capacitance window covers the useful phase-shift range of the incident band, and that larger element spacing increases frequency-dependent array mismatch in native-band multi-frequency operation.

The rest of this chapter is organized as follows. Section 7.2 presents the frequency-dependent array response, the circuit-based reflection model, and the supported phase-shift range of the RIS elements. Section 7.3 formulates the multi-frequency RIS capacitance optimization problem and presents the proposed projected ADAM-based solution. Section 7.4 provides numerical results and discussions on convergence, solver comparison, feasible capacitance window, and element spacing. Finally, Section 7.5 summarizes the chapter.

7.2 RIS Model Characterization

In this section, we first present the frequency-dependent array response and the circuit-based reflection model of the RIS elements, and then characterize the supported phase-shift range induced by the feasible capacitance window. We consider two band-specific RIS designs operating in the sub-6 GHz and mid-band ranges. Let $r \in \{\mathrm{S}, \mathrm{M}\}$ denote the frequency band, where $r = \mathrm{S}$ corresponds to the sub-6 GHz and $r = \mathrm{M}$ corresponds to the mid-band. The RIS is tuned to the design frequency f_r. Also, for $b \in \{\mathrm{S}, \mathrm{M}\}$, let $\mathcal{I}_b$ denote the index set of the incident frequencies. Thus, the native-band case corresponds to $b = r$, whereas the cross-band case corresponds to $b \neq r$, i.e., the same RIS hardware is optimized over the frequency set of the other band while its circuit parameters and element spacing remain unchanged.

7.2.1 Array Response and Circuit-Based Reflection Model

7.2.1.1 Array Response

Assuming that the transmitter(s) and receiver(s) are located in the far field of the RIS, a plane wave signal with frequency $f_i, \forall i \in \mathcal{I}_b$, impinges on the uniform planar array (UPA)-based RIS from angular directions characterized by an elevation angle $\varphi_i \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ and an azimuth angle $\theta_i \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. The corresponding array response vector, capturing the respective phase shifts across the RIS elements, is given by:

$$\boldsymbol{a}(\varphi_i, \theta_i, f_i, f_r) = [1, \ldots, e^{-j2\pi(N_z-1)d_r \frac{\sin(\varphi_i)}{\lambda_i}}]^T \otimes [1, \ldots, e^{-j2\pi(N_y-1)d_r \frac{\sin(\theta_i)\cos(\varphi_i)}{\lambda_i}}]^T, \forall i \tag{7.1}$$

where N_z and N_y are the number of vertical and horizontal elements in the UPA, respectively, and $N = N_z N_y$ is the total number of RIS elements. $\lambda_i = \frac{c}{f_i}$ is the wavelength of the incident signal, c is the speed of light. The element spacing is $d_r = \alpha_r \lambda_r = \frac{\alpha_r c}{f_r}$, where α_r is a factor controlling the element spacing.

7.2.1.2 Frequency Dependency of RIS Elements

Each RIS element functions as a resonant circuit consisting of two inductors ($L_{r,1}, L_{r,2}$), a resistor (R_r), and a varactor with an adjustable capacitance $C_{r,n}$ for the n-th element of the RIS tuned at f_r. Using circuit theory, the frequency-dependent reflection coefficient of the n-th element of the RIS at frequency f_i is given by [58]:

$$\psi(\nu_i, C_{r,n}) = \frac{Z(\nu_i, C_{r,n}) - Z_0}{Z(\nu_i, C_{r,n}) + Z_0}, \tag{7.2}$$

where $\nu_i = 2\pi f_i$, Z_0 is the free space impedance, and $Z(\nu_i, C_{r,n})$ is the self-impedance of the n-th element of the RIS operating at f_r and is given by:

$$Z\left(\nu_i, C_{r,n}\right) = \frac{j\nu_i L_{r,1}\left(j\nu_i L_{r,2} + R_r + \frac{1}{j\nu_i C_{r,n}}\right)}{j\nu_i L_{r,1} + j\nu_i L_{r,2} + R_r + \frac{1}{j\nu_i C_{r,n}}}. \tag{7.3}$$

Using (7.3), further simplified formulation of (7.2) can be obtained as:

$$\psi(\nu_i, C_{r,n}) = \frac{\rho_r(\nu_i)C_{r,n} + \tau_r(\nu_i)}{\mu_r(\nu_i)C_{r,n} + \tau_r^*(\nu_i)}, \tag{7.4}$$

where

$$\rho_r(\nu_i) = -\nu_i^2 L_{r,1}L_{r,2} - Z_0 R_r + j\nu_i(R_r L_{r,1} - Z_0(L_{r,1} + L_{r,2})),$$

$$\tau_r(\nu_i) = L_{r,1} + j\frac{Z_0}{\nu_i},$$

$$\mu_r(\nu_i) = -\nu_i^2 L_{r,1}L_{r,2} + Z_0 R_r + j\nu_i(R_r L_{r,1} + Z_0(L_{r,1} + L_{r,2})).$$

7.2.2 RIS Supported Phase-Shift Range

Due to the frequency dependency of RIS elements, for a given value of inductors that determine the main operating frequency of an RIS, the range of phase shifts imposed by an RIS element on the incident signal varies with both the range of the adjustable capacitor and the frequency of the incident signal. Therefore, for further analysis and effective utilization of RISs, it is necessary to identify this range.

The phase response of each RIS element is given by:

$$\angle\psi(f_i, C_{r,n}) = \text{atan2}\left(2Z_0 X_{r,z}, R_{r,z}^2 - Z_0^2 + X_{r,z}^2\right), \tag{7.5}$$

where

$$R_{r,z} = \frac{(\nu_i L_{r,1})^2 R_r}{R_r^2 + (\nu_i L_{r,1} + X_{r,s})^2}, \tag{7.6}$$

$$X_{r,z} = \frac{\nu_i L_{r,1} \left(R_r^2 + X_{r,s}(\nu_i L_{r,1} + X_{r,s})\right)}{R_r^2 + (\nu_i L_{r,1} + X_{r,s})^2}, \tag{7.7}$$

$X_{r,s} = \nu_i L_{r,2} - \frac{1}{\nu_i C_{r,n}}$ is the impedance of the $L_2 - R - C$ series branch of the circuit, and atan2 is the four-quadrant inverse tangent.

Define $g(f_i, C_{r,n}) = \angle\psi(f_i, C_{r,n})$. It can be shown that $g(f_i, C_{r,n})$ is a decreasing function of $C_{r,n}$ up to a certain point, after which it abruptly jumps from $-\pi$ to π and then varies slightly. Moreover, the imaginary part of the reflection coefficient becomes zero at two capacitance values. The smaller root gives us the capacitance $C_r^+(f_i)$ at which $g(f_i, C_r^+(f_i))$ becomes zero and the larger root yields capacitance $C_r^-(f_i)$, where the phase experiences a jump from $-\pi$ to π. These two capacitance values are obtained by:

$$C_r^{\pm}(f_i) = \frac{2}{\nu_i \left(\nu_i(L_{r,1} + 2L_{r,2}) \pm \sqrt{\nu_i^2 L_{r,1}^2 - 4R_r^2}\right)}, \tag{7.8}$$

where $\nu_i^2 L_{r,1}^2 \geq 4R_r^2$.

Next, we define $C_r^{\min}$ and $C_r^{\max}$ as the minimum and maximum capacitances of the varactors for frequency band r. Using the decreasing behavior of $g(f_i, C_{r,n})$ and $\lim_{C_{r,n} \to (C_r^-(f_i))^-} g(f_i, C_{r,n}) = -\pi$, the range of phase shifts supported by the n-th

element of the RIS tuned to f_r can be approximated by:

$$\mathcal{R}(f_i, C_r^{\min}, C_r^{\max}) = \begin{cases} [g(f_i, C_r^{\max}), g(f_i, C_r^{\min})], & \text{if } C_r^{\min} < C_r^{\max} < C_r^{-}(f_i), \\ [-\pi, g(f_i, C_r^{\min})], & \text{if } C_r^{\min} < C_r^{-}(f_i) < C_r^{\max}, \\ \emptyset, & \text{if } C_r^{-}(f_i) < C_r^{\min} < C_r^{\max}. \end{cases} \quad (7.9)$$

In (7.9), we interpret $\emptyset$ as an empty controllable phase interval, not as zero reflection. In particular, in the region $C_r^{-}(f_i) < C_r^{\min} \leq C_{r,n}$, the phase remains approximately fixed near π. Thus, if the incident frequency satisfies $C_r^{-}(f_i) < C_r^{\min}$, the capacitance tuning provides negligible phase-control capability. In this scenario, the RIS loses controllable phase support and behaves in a mirror-like manner. We refer to such frequencies as effectively out-of-band with respect to controllable phase tuning for the RIS hardware designed at f_r.

7.3 Problem Formulation and Proposed Solution

In this section, we formulate the design of the RIS capacitances for multi-frequency operation. For a given RIS tuned to the design frequency f_r, the objective is to determine a single capacitance vector $\boldsymbol{C}_r = [C_{r,1}, \ldots, C_{r,N}]^T$ that provides high reflected array gains over a set of incident frequencies, which may correspond to either native-band or cross-band operation. To this end, we first define the frequency-dependent array gain of the RIS and then formulate a multi-frequency sum gain maximization problem.

Consider the diagonal reflection matrix of the RIS as

$$\boldsymbol{\Psi}(f_i, \boldsymbol{C}_r) = \mathrm{diag}(\psi(f_i, C_{r,1}), \ldots, \psi(f_i, C_{r,N}))\,, \tag{7.10}$$

where $\boldsymbol{C}_r$ is the capacitance vector of the RIS operating in band r. Then, the array gain at observation (desired reflection) elevation angle of $\tilde{\varphi}_i$, and azimuth angle of $\tilde{\theta}_i$ for frequency f_i is given by:

$$u(f_i, \boldsymbol{C}_r) = \frac{1}{N}\left|\boldsymbol{a}^T(\tilde{\varphi}_i, \tilde{\theta}_i, f_i, f_r)\boldsymbol{\Psi}(f_i, \boldsymbol{C}_r)\boldsymbol{a}(\varphi_i, \theta_i, f_i, f_r)\right|^2 = \frac{1}{N}\left|\sum_{n_y=0}^{N_y-1}\sum_{n_z=0}^{N_z-1} e^{-j\frac{2\pi\alpha_r\lambda_r}{\lambda_i}(n_z\zeta_i+n_y\xi_i)}\psi(f_i, C_{r,n})\right|^2, \tag{7.11}$$

where $\zeta_i = \sin(\varphi_i)+\sin(\tilde{\varphi}_i)$, $\xi_i = \sin(\theta_i)\cos(\varphi_i)+\sin(\tilde{\theta}_i)\cos(\tilde{\varphi}_i)$, and $n = n_y+n_zN_y+1$.

7.3.1 Sum Array-Gain Maximization

For a fixed element spacing α_r, we seek a single capacitance vector associated with the RIS operating in band r that jointly serves all incident frequencies in the set $\mathcal{I}_b$. Accordingly, the design problem is formulated as:

$$\begin{aligned} \mathcal{P}_1: \quad & \max_{\boldsymbol{C}_r} \ \sum_{i\in\mathcal{I}_b} u(f_i, \boldsymbol{C}_r) \\ \text{s.t.} \quad & C_r^{\min} \le C_{r,n} \le C_r^{\max}, \quad \forall n. \end{aligned} \tag{7.12}$$

Problem $\mathcal{P}_1$ captures the practical requirement that one fixed RIS configuration must operate efficiently over multiple frequencies within the desired band. The main challenge is that the array gain depends nonlinearly on the capacitance vector through

the frequency-dependent reflection coefficients of the RIS elements, which makes $\mathcal{P}_1$ non-convex.

7.3.1.1 Projected ADAM-Based Solution

To derive an efficient first-order method, we rewrite the objective of $\mathcal{P}_1$ in a compact form. Define

$$\hat{\boldsymbol{a}}_i = \boldsymbol{a}^T(\tilde{\varphi}_i, \tilde{\theta}_i, f_i, f_r)\text{diag}(\boldsymbol{a}(\varphi_i, \theta_i, f_i, f_r))\,, \tag{7.13}$$

and

$$\hat{\boldsymbol{\psi}}_i = [\psi(f_i, C_{r,1}), \ldots, \psi(f_i, C_{r,N})]^T. \tag{7.14}$$

Then, the objective function in (7.12) can be written as:

$$U(\boldsymbol{C}_r) = \frac{1}{N}\sum_{i\in\mathcal{I}_b}\left|\hat{\boldsymbol{a}}_i\hat{\boldsymbol{\psi}}_i\right|^2. \tag{7.15}$$

Using (7.4), the gradient of $U(\boldsymbol{C}_r)$ is obtained by:

$$\nabla_{\boldsymbol{C}_r}U = \left[\frac{\partial U}{\partial C_{r,1}}, \ldots, \frac{\partial U}{\partial C_{r,N}}\right]^T,$$

where

$$\frac{\partial U}{\partial C_{r,n}} = \frac{2}{N}\sum_{i\in\mathcal{I}_b}\Re\left\{\hat{\boldsymbol{a}}_{i,n}\,\frac{\rho_r(\nu_i)\tau_r^*(\nu_i) - \mu_r(\nu_i)\tau_r(\nu_i)}{\left(\mu_r(\nu_i)C_{r,n} + \tau_r^*(\nu_i)\right)^2}\,\left(\hat{\boldsymbol{a}}_i\hat{\boldsymbol{\psi}}_i\right)^*\right\}, \tag{7.16}$$

and $\hat{\boldsymbol{a}}_{i,n}$ denotes the n-th element of $\hat{\boldsymbol{a}}_i$.

Starting from an initial point $\boldsymbol{C}_r^{(0)}$, we solve $\mathcal{P}_1$ using projected ADAM. In each iteration, the capacitance vector is updated along the gradient direction with adaptive

first- and second-moment normalization, and then projected onto the feasible interval $[C_r^{\min}, C_r^{\max}]$. The procedure is summarized in Algorithm 1.

Algorithm 6 Projected ADAM for Solving $\mathcal{P}_1$

Require: Initial $\boldsymbol{C}_r^{(0)} \in [C_r^{\min}, C_r^{\max}]^N$; step size η; moments β_1, β_2; numerical constant ε; tolerance `tol`; maximum number of iterations `maxIter`.

$\mathbf{m}^{(0)} \leftarrow \mathbf{0}, \quad \mathbf{v}^{(0)} \leftarrow \mathbf{0}$

$U^{(0)} \leftarrow U\left(\boldsymbol{C}_r^{(0)}\right)$

for $t = 1$ to `maxIter` **do**

  $\mathbf{g}^{(t)} \leftarrow \nabla_{\boldsymbol{C}_r} U\left(\boldsymbol{C}_r^{(t-1)}\right)$

  $\mathbf{m}^{(t)} \leftarrow \beta_1 \mathbf{m}^{(t-1)} + (1-\beta_1)\mathbf{g}^{(t)}$

  $\mathbf{v}^{(t)} \leftarrow \beta_2 \mathbf{v}^{(t-1)} + (1-\beta_2)\left(\mathbf{g}^{(t)} \odot \mathbf{g}^{(t)}\right)$

  $\widehat{\mathbf{m}}^{(t)} \leftarrow \dfrac{\mathbf{m}^{(t)}}{1-\beta_1^t}, \ \widehat{\mathbf{v}}^{(t)} \leftarrow \dfrac{\mathbf{v}^{(t)}}{1-\beta_2^t}$

  $\boldsymbol{C}_r^{(t)} \leftarrow \boldsymbol{C}_r^{(t-1)} + \eta \dfrac{\widehat{\mathbf{m}}^{(t)}}{\sqrt{\widehat{\mathbf{v}}^{(t)}} + \varepsilon}$

  $\boldsymbol{C}_r^{(t)} \leftarrow \max\left\{\min\left\{\boldsymbol{C}_r^{(t)}, C_r^{\max}\right\}, C_r^{\min}\right\}$

  $U^{(t)} \leftarrow U\left(\boldsymbol{C}_r^{(t)}\right)$

  if $\left|U^{(t)} - U^{(t-1)}\right| \le$ `tol` **then**

    return $\boldsymbol{C}_r^{(t)}$

  end if

end for

return $\boldsymbol{C}_r^{(\texttt{maxIter})}$

Algorithm 1 is a projected first-order heuristic that empirically converges to a stationary or locally stable solution. Because problem $\mathcal{P}_1$ is non-convex, the solution depends on initialization $\boldsymbol{C}_r^{(0)}$. A physically meaningful initialization can be obtained

from the zero-phase capacitance in (7.8), i.e., $C_r^+(f_i)$, at a given frequency f_i, provided that $g(f_i, C_r^+(f_i))$ lies within the feasible phase-support range $\mathcal{R}(f_i, C_r^{\min}, C_r^{\max})$. Motivated by this observation, we initialize the capacitance vector by averaging the zero-phase capacitances over the frequency set $\mathcal{I}_b$, namely

$$C_{r,n}^{(0)} = \frac{\epsilon}{|\mathcal{I}_b|} \sum_{i \in \mathcal{I}_b} C_r^+(f_i), \forall n \tag{7.17}$$

where $\epsilon \in [\epsilon_{\min}, \epsilon_{\max}]$ is a scalar search parameter. For different values of ϵ, Algorithm 1 is executed and the solution achieving the largest sum array gain is selected. This adaptive initialization procedure is summarized in Algorithm 2.

It is worth noting that, if a particular frequency is prioritized, its corresponding zero-phase capacitance can be directly used as the initialization point.

Algorithm 7 Adaptive Initialization for Solving $\mathcal{P}_1$

Require: Search interval $[\epsilon_{\min}, \epsilon_{\max}]$

1: **for** $\epsilon \in [\epsilon_{\min}, \epsilon_{\max}]$ **do**

2: $\quad \boldsymbol{C}_r^{(0,\epsilon)} \leftarrow \frac{\epsilon}{|\mathcal{I}_b|} \sum_{i \in \mathcal{I}_b} C_r^+(f_i) \mathbf{1}$

3: $\quad \boldsymbol{C}_r^{(0,\epsilon)} \leftarrow \max\Big\{\min\Big\{\boldsymbol{C}_r^{(0,\epsilon)}, C_r^{\max}\Big\}, C_r^{\min}\Big\}$

4: $\quad \boldsymbol{C}_r^{(\epsilon)} \leftarrow$ Solve $\mathcal{P}_1$ with Algorithm 1 and initialization $\boldsymbol{C}_r^{(0,\epsilon)}$

5: **end for**

6: $\boldsymbol{C}_r^{\text{opt}} \leftarrow \arg\max_{\epsilon} U\Big(\boldsymbol{C}_r^{(\epsilon)}\Big)$

7: **return** $\boldsymbol{C}_r^{\text{opt}}$

7.3.2 Complexity Analysis

Let $K = |\mathcal{I}_b|$ denote the number of incident frequencies, T the maximum number of projected ADAM iterations in Algorithm 1, and Q the number of tested initialization parameters ϵ in Algorithm 2. For a given capacitance vector, evaluating the objective and its gradient requires computing the RIS reflection coefficient and derivative for all N elements over all K frequencies. Hence, the per-iteration complexity of Algorithm 1 is $\mathcal{O}(KN)$. Since Algorithm 1 runs for at most T iterations, its complexity is $\mathcal{O}(TKN)$. Algorithm 2 executes Algorithm 1 for each value of ϵ. Therefore, the total complexity of the adaptive initialization procedure is: $\mathcal{O}(QTKN)$. The additional cost of computing the zero-phase capacitances $C_r^+(f_i)$ over all frequencies is only $\mathcal{O}(K)$, which is negligible compared with the iterative optimization cost.

7.4 Numerical Results and Discussions

In this section, we evaluate the proposed RIS capacitance optimization framework for sub-6 GHz and mid-band operation. We first compare the proposed optimization method with two baseline solvers for native-band RIS design, and then study the impact of the feasible capacitance window and the element spacing factor on the achievable multi-frequency array gain. The RIS design frequencies are $f_{\mathrm{S}} = 2.5$ GHz and $f_{\mathrm{M}} = 8$ GHz. The sub-6 frequency set is $[2, 3.5]$ GHz and the mid-band frequency set is $[6, 10]$ GHz, with $K = 7$ equally spaced frequencies in each band. Unless otherwise stated, each RIS has $N_y = N_z = 8$ elements, i.e., $N = 64$, and $\alpha_r = 0.5$.

To avoid mirror angles, the incident and observation (desired reflection) angles are fixed as $\varphi_i = 30°, \quad \theta_i = 35°, \quad \tilde{\varphi}_i = 5°, \quad \tilde{\theta}_i = -10°$. The circuit parameters are $L_{\mathrm{S},1} = 2.5$ nH, $\quad L_{\mathrm{S},2} = 0.7$ nH, and $L_{\mathrm{M},1} = 0.83$ nH, $\quad L_{\mathrm{M},2} = 0.16$ nH. For

both RISs, we set $R_r = 1\ \Omega$, $Z_0 = 377\ \Omega$. The performance metric is the normalized sum array gain $\bar{U}_r = \frac{1}{N}\sum_{i\in\mathcal{I}_b} u(f_i, \boldsymbol{C}_r)$, where $u(f_i, \boldsymbol{C}_r)$ is the array gain defined in (7.11). Since $u(f_i, \boldsymbol{C}_r) \leq N$ under passive reflection, the additional factor $1/N$ normalizes the contribution of each frequency to at most one. For Section IV.A and IV.C, the fixed capacitance windows are $[C_r^{\min}, C_r^{\max}] = [0.05, 1.60]$ pF for the sub-6 RIS and $[C_r^{\min}, C_r^{\max}] = [0.05, 0.70]$ pF for the mid-band RIS. The projected ADAM parameters are $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\eta = 0.015$, and $\varepsilon = 10^{-8}$, and the adaptive initialization factor is $\epsilon \in [0.75, 1.25]$ with 10 values within the interval.

7.4.1 Convergence and Solver Comparison

Before studying the physical effects of the capacitance window and element spacing, we first compare the convergence behavior of the proposed optimization method with two baselines. The first baseline is Algorithm 1 with random initialization, where the projected ADAM updates are kept unchanged but the initial capacitance vectors are sampled uniformly from the feasible interval. This baseline isolates the benefit of the proposed zero-phase-based initialization in Algorithm 2. The plotted curves for the proposed method and the random-initialized ADAM baseline correspond to the best final run out of 10 trials. The second baseline is a genetic algorithm (GA), which uses random population initialization. The plotted GA curve likewise corresponds to the best run out of 10 independent trials.

Fig. 7.1 shows the convergence behavior of the best run for each method in the native-band sub-6 and mid-band RIS cases. In Fig. 7.1(a), the proposed method converges rapidly to the highest final gain for both RISs. This is because its initialization is based on the average zero-phase capacitance $C_r^{+}(f_i)$ over the considered frequency

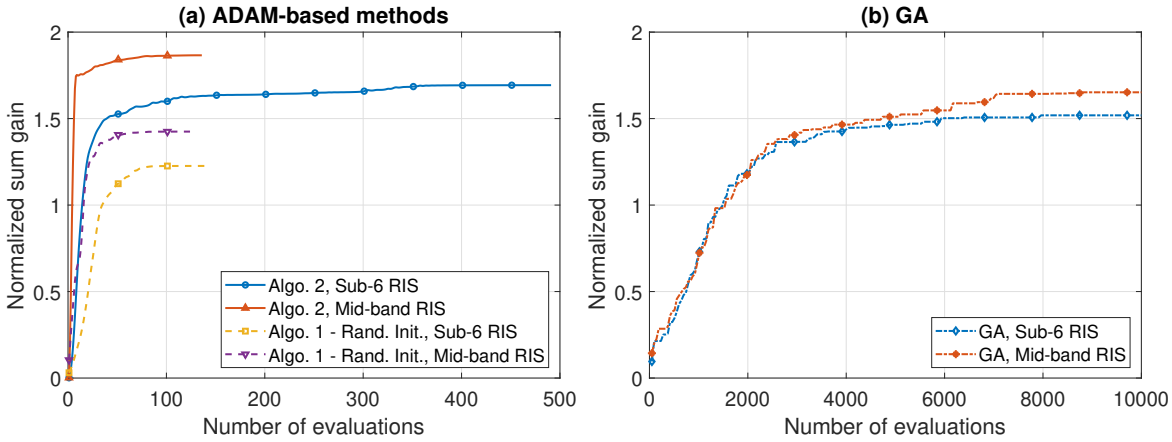


Figure 7.1: Convergence behavior of the best run for each method in native-band RIS optimization: (a) ADAM-based methods and (b) GA.

set, which places the initial point near a physically meaningful phase-shift range. In contrast, the random-initialized ADAM baseline also converges quickly, but to a lower local optimum, showing the sensitivity of the non-convex problem to the initialization point. In Fig. 7.1(b), the GA improves more gradually and requires substantially more evaluations, but its best run can still outperform the random-initialized ADAM baseline while remaining below the proposed method. These results show that the proposed method benefits from both the analytical gradient and the circuit-aware adaptive initialization.

7.4.2 Impact of Feasible Capacitance Window

Figs. 7.2–7.5 illustrate the impact of the feasible capacitance window on the normalized sum array gain. For each RIS design and target band, the feasible varactor range is swept over 0.05 pF $\leq C_r^{\min} < C_r^{\max} \leq 3$ pF and Algorithm 2 is used to solve the problem. The dashed line indicates the boundary $C_r^{\min} = C_r^{\max}$. Fig. 7.2 shows that the sub-6 RIS achieves strong native-band performance over a broad range of

capacitance windows, as long as the feasible interval contains sufficiently small capacitance values and extends into the useful tuning region for the sub-6 frequency set. In contrast, Fig. 7.3 shows that when the same RIS serves the mid-band, high gain is obtained only when $C_r^{\min}$ is very small, since higher incident frequencies shift the useful phase-shift range toward smaller capacitances. When $C_r^{\min}$ is too large, the low-gain region is consistent with the onset of out-of-band in (7.9), where controllable phase support is lost.

Fig. 7.4 shows that the mid-band RIS achieves strong native-band performance when the capacitance window includes the low-pF region associated with the 8 GHz design. However, Fig. 7.5 shows that when the mid-band RIS serves the sub-6 band, the normalized sum array gain improves mainly as $C_r^{\max}$ increases, indicating that lower incident frequencies require larger capacitance values to recover useful phase tunability. The strongest region appears near the upper end of the tested $C_r^{\max}$ range, suggesting that a larger maximum capacitance could further improve performance. Overall, these results show that native-band and cross-band RIS operation are governed jointly by the circuit parameters and the feasible capacitance window. The heatmaps are also consistent with the phase-support analysis in Section II: strong performance is obtained when the feasible capacitance window covers the useful phase-shift range of the incident band, whereas insufficient coverage leads to weak tunability and reduced multi-frequency array gain.

7.4.3 Impact of Element Spacing

Fig. 7.6 illustrates the impact of the element spacing factor α_r on the normalized sum array gain. For both the sub-6 and mid-band RISs, the proposed multi-frequency

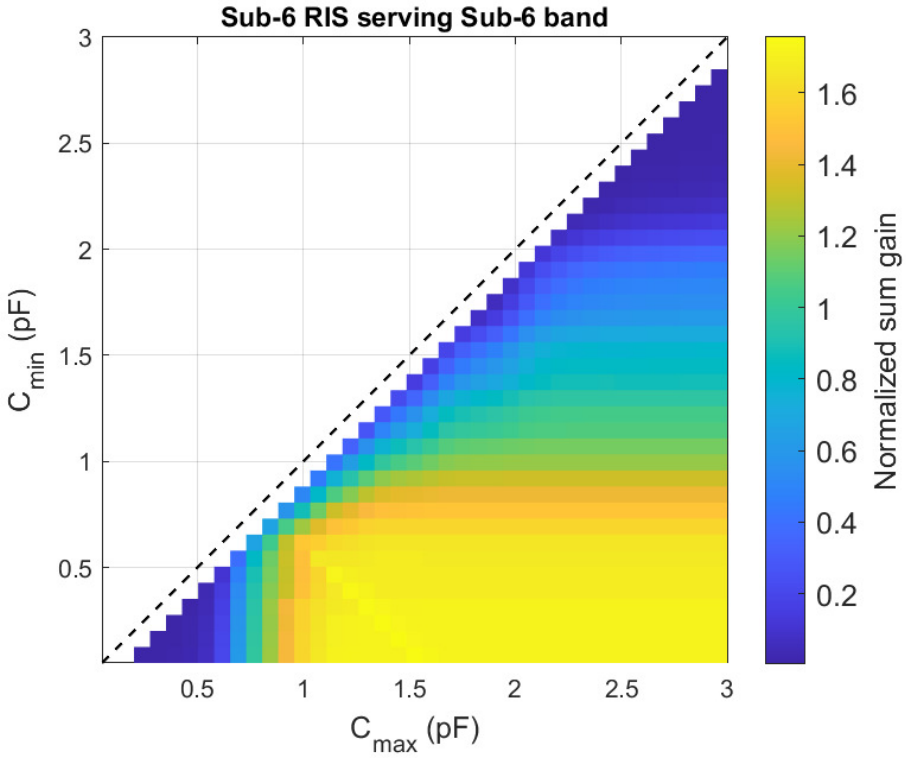


Figure 7.2: Impact of the feasible capacitance window on the normalized sum array gain for the sub-6 RIS serving the sub-6 frequencies.

adaptive initialization consistently outperforms the f_r-baseline. In the f_r-baseline, the RIS capacitances are optimized only at the RIS operating frequency f_r, using the zero-phase capacitance $C_r^+(f_r)$ as the initialization, and the resulting capacitance vector is then evaluated over the entire corresponding band r. As α_r increases, the electrical spacing satisfies $\frac{d_r}{\lambda_i} = \frac{\alpha_r \lambda_r}{\lambda_i} = \frac{\alpha_r f_i}{f_r}$, which increases the frequency-dependent phase variation across the RIS aperture. Consequently, a single capacitance vector becomes less effective at jointly aligning the reflected beams over the full band, and the normalized sum array gain decreases. These results confirm that the element spacing plays a key role in the severity of frequency-dependent array mismatch and that the proposed multi-frequency optimization is more robust than tuning the RIS

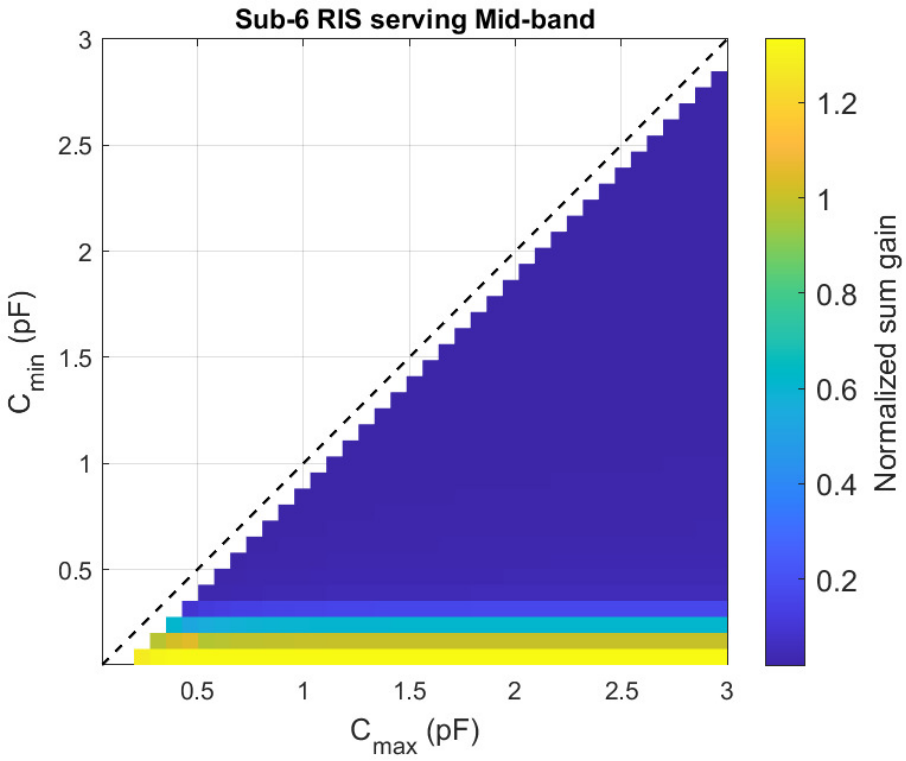


Figure 7.3: Impact of the feasible capacitance window on the normalized sum array gain for the sub-6 RIS serving the mid-band frequencies.

only at f_r.

7.5 Summary

This chapter studied performance optimization for RIS-assisted multi-band networks using a practical circuit-based RIS model. We analytically characterized the frequency-dependent phase-shift range of RIS elements in terms of the circuit parameters, incident frequency, and feasible capacitance window. We then formulated a multi-frequency sum array gain maximization problem and solved it using projected ADAM with an adaptive zero-phase-based initialization. Numerical results showed that the proposed circuit-aware initialization achieves higher gain than random-initialized

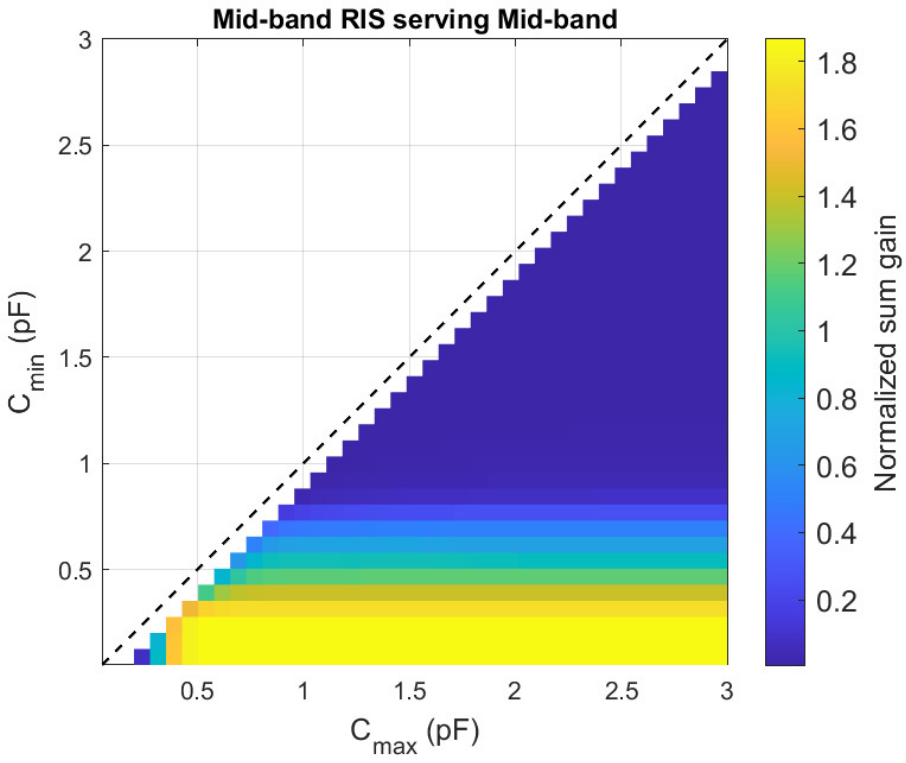


Figure 7.4: Impact of the feasible capacitance window on the normalized sum array gain for the mid-band RIS serving the mid-band frequencies.

ADAM and a GA baseline in the considered native-band cases. They also showed that native-band and cross-band performance are strongly governed by the feasible capacitance window, while larger element spacing increases frequency-dependent array mismatch over the native frequency band.

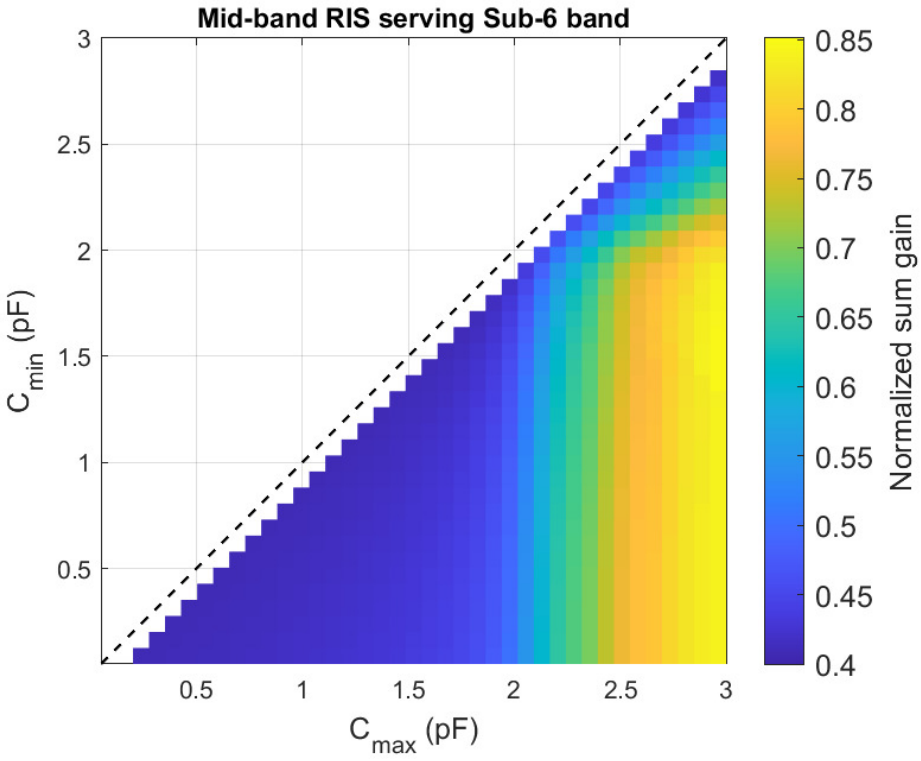


Figure 7.5: Impact of the feasible capacitance window on the normalized sum array gain for the mid-band RIS serving the sub-6 frequencies.

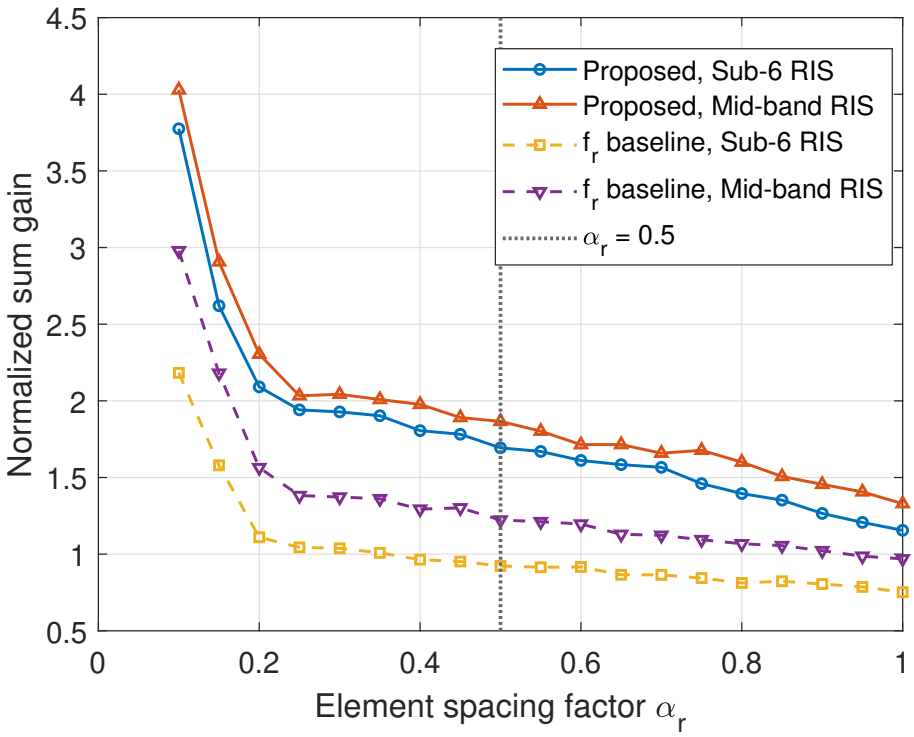


Figure 7.6: Impact of the element spacing factor α_r on the normalized sum array gain for the sub-6 and mid-band RISs.

Chapter 8

Conclusion and Future Work

In this chapter, we summarize the major research contributions and point out future research directions.

8.1 Conclusion

This dissertation investigated resource and mobility management for 6G wireless networks operating in high-frequency, multi-band, and multi-frequency settings. The work begins with THz resource allocation and then extends to broader MBN architectures, mobility-aware cooperative UMB/THz networking, and RIS-assisted multi-frequency systems. Across these settings, the dissertation addresses how frequency-dependent propagation, blockage, beamforming, handovers, multi-connectivity, and practical hardware constraints affect network design and optimization. The main contributions of the dissertation are summarized as follows.

- **Molecular-absorption-aware resource allocation in dense THz networks:** We developed a low-complexity framework for sub-band partitioning, joint user association, sub-band assignment, and power allocation in dense THz

networks with multi-connectivity. The framework accounts for molecular absorption variations, molecular absorption noise, blockage, interference, beam squint, hardware impairments, and imperfect CSI. A key contribution is that the joint user association and sub-band assignment problem was transformed into a linear program by proving total unimodularity of the constraint matrix, which enables solving the relaxed problem without losing optimality. For power allocation, we proposed both a centralized FP-based method and a lightweight distributed ADMM-based method. The results show that the proposed algorithms improve the network sum-rate compared with the considered benchmarks, while highlighting the impact of BS density, multi-connectivity, molecular absorption, blockage, hardware impairments, and imperfect CSI on THz network performance.

- **Characterization of multi-band network architectures:** We characterized MBNs and compared stand-alone and integrated MBN architectures. We discussed key challenges related to multi-band transceivers, antenna systems, channel modeling, deployment and planning, user offloading, mobility management, and resource allocation. We also proposed a molecular-absorption-aware user offloading metric for RF/THz MBNs. The results show that the proposed metric can achieve performance close to the maximum-rate association rule, that Int-MBN can require fewer BSs than SA-MBN for the same target rate when equal RF/THz densities are used, and that SA-MBN can approach Int-MBN performance when unequal RF and THz BS densities are allowed.

- **Mobility-aware resource allocation in cooperative UMB/THz MBNs:** We developed a cooperative UMB/THz MBN framework where BSs in each

band cooperatively serve users using ELAAs and hybrid beamforming. The framework incorporates near-field and velocity-aware channel modeling, blockage, QoS constraints, and user mobility. For stationary users, we derived feasible analog beamformers, reformulated the power constraint, and solved the joint user association and digital beamforming problem using big-M transformation, FP, quadratic transformation, and MM. A key theoretical step was proving the equivalence of the quadratic-transformed FP problem with logarithmic fractional objective and QoS constraints. For mobile users, we developed two HO-aware resource allocation methods. The first uses the structure of the HO-aware rate, proves its log-concavity after quadratic transformation, and applies a convex approximation using Jensen's inequality and logarithmic properties. The second balances sum-rate maximization and HO minimization through a weighted-objective formulation. The results show that, for both stationary and mobile users, cooperative UMB/THz MBNs improve performance compared with THz-only networks, that near-field beamforming is important for ELAAs, and that larger cooperative clusters are not always beneficial under mobility due to the HO trade-off.

- **Beam-squint and beam-split characterization in RIS-assisted multi-frequency networks:** We developed a unified analytical framework to characterize beam-squint and beam-split effects in UPA-based RIS-assisted multi-frequency systems under both continuous and finite-resolution phase profiles. For continuous RIS phase control, we derived necessary and sufficient conditions for maximum beamforming gain, obtained closed-form elevation–azimuth squint/split angle pairs, and established feasibility conditions that determine

when beam squint or beam split occurs as a function of frequency ratio, element spacing, and incidence/configuration angles. For practical b-bit phase quantization, we represented the quantized beampattern as a superposition of harmonic array responses using Fourier expansion of the quantization-error phasor, which enabled per-harmonic peak characterization, feasibility analysis, and dominance ordering. We also developed a perturbation-based correction method to improve the predicted peak locations under harmonic superposition. The results validate the analytical characterization, quantify the dependence of beam-split probability on frequency ratio and element spacing, and show that beam-split-aware codeword selection can improve the sum-rate in multi-RIS, multi-user, multi-band networks.

- **Circuit-aware RIS optimization for sub-6 GHz and UMB multi-band networks:** We studied RIS-assisted multi-band networks with practical frequency-dependent RIS elements and used a circuit-based model to characterize the supported phase-shift range as a function of incident frequency, circuit parameters, and feasible capacitance window. We then formulated a multi-frequency sum array-gain maximization problem and solved it using a projected ADAM-based algorithm with an adaptive zero-phase-based initialization that exploits the RIS circuit response. The results show that this initialization improves the achieved gain in the considered native-band cases, while the feasible capacitance window determines the availability of useful phase control and the element spacing affects the robustness of multi-frequency RIS operation.

Overall, the dissertation shows that multi-band and multi-frequency 6G networks

require resource and mobility management methods that are aware of both network-level and hardware-level frequency dependence. At the network level, propagation loss, molecular absorption, blockage, mobility, and handovers affect user association, spectrum allocation, power allocation, and cooperative transmission. At the surface and array level, frequency mismatch affects RIS beam directions, quantized beampatterns, feasible phase control, and multi-frequency array gain. These findings support the broader conclusion that future 6G systems should be designed using frequency-aware and mobility-aware optimization methods rather than treating different bands or different propagation-control components independently.

The results should also be interpreted together with practical deployment limitations. The proposed algorithms reduce several difficult resource allocation problems to tractable forms, but real networks will still require further complexity reduction, lower CSI acquisition overhead, and more distributed control. This motivates future work on hierarchical, distributed, proactive, and learning-assisted resource allocation methods that can reduce computational burden and avoid relying only on instantaneous full CSI. Similarly, THz links and RISs offer important practical benefits, but they are not universal solutions that can be deployed everywhere. THz communication is especially useful for data-hungry applications such as extended reality, digital twins, high-capacity indoor access, wireless backhauling, and short-range high-speed connectivity. Due to limited coverage and blockage sensitivity, THz links are most suitable for indoor hotspots, factories, transportation hubs, data centers, fixed point-to-point links, and carefully planned outdoor deployments. In outdoor networks, THz is more practical when integrated with lower-frequency bands in MBNs, where lower-frequency links provide coverage, control signaling, and mobility support, while THz

links provide high-rate transmissions under favorable propagation conditions. This also makes THz-enabled MBNs relevant for rural or low-blockage deployments and satellite or non-terrestrial communication scenarios where highly directional links and large bandwidths can be exploited. RIS deployment also faces practical constraints related to site availability, ownership, zoning or building regulations, maintenance, and coexistence with other operators. In particular, RISs may create unintended reflections and interference for neighboring networks or operators, which motivates future research on deployment-aware, regulation-aware, and coexistence-aware RIS-assisted MBN optimization. These practical considerations motivate the following future research directions.

8.2 Future Work

This dissertation opens several research directions, as discussed below.

- *Predictive and CSI-light resource allocation:* The resource allocation frameworks developed in this dissertation mainly rely on channel, location, and mobility information available at the optimization time. In practical multi-band networks, acquiring full instantaneous CSI across numerous BSs, RISs, antennas, users, and frequency bands incurs significant pilot overhead, feedback overhead, and latency. A promising future direction is therefore to develop predictive and CSI-light resource allocation frameworks that combine partial or statistical CSI with forecasting models capable of anticipating user mobility, traffic demand, channel quality, blockage events, interference, and handover requirements. By exploiting predicted network dynamics rather than reacting only to the current network state, these frameworks can proactively optimize user association,

spectrum allocation, power control, beamforming, RIS configuration, and multi-connectivity before severe blockage, beam misalignment, or unnecessary handovers occur. Furthermore, incorporating prediction uncertainty into robust or risk-aware optimization frameworks [146, 147] can improve the reliability of predictive resource allocation under imperfect forecasts. Such approaches are particularly attractive for highly dynamic 6G scenarios, including vehicular, railway, fixed wireless access, and dense hotspot deployments, where network conditions exhibit predictable spatio-temporal patterns.

- *Hierarchical and distributed low-complexity resource allocation:* Although this dissertation proposed several tractable and low-complexity solutions, further complexity reduction is needed for real-time dense MBN deployments. Future work can investigate hierarchical and distributed resource allocation methods, where large-scale problems are decomposed into smaller subproblems across network entities, frequency bands, or time scales. For example, slow-timescale decisions such as BS/RIS clustering and band selection can be separated from fast-timescale decisions such as power allocation, beam updates, and handover response. Such approaches can reduce computational complexity and signaling overhead while making the proposed frameworks more practical for dense 6G deployments.

- *Learning-to-optimize and surrogate solvers for real-time operation:* Another way to reduce complexity is to replace repeated execution of iterative solvers with trained surrogate models. Learning-to-optimize approaches can provide fast initializations or approximate solutions for user association, power allocation, beamforming, cooperative clustering, and RIS configuration. A practical design

can combine learning-based inference with a small number of optimization-based correction steps to enforce feasibility, QoS, power-budget, and interference constraints.

- *Joint resource allocation across more diverse frequency bands:* This dissertation focused mainly on THz, UMB/THz, RF/THz, and RIS-assisted sub-6 GHz/UMB settings. Future MBNs may jointly use sub-6 GHz, UMB, mmWave, THz, VLC, FSO, and satellite links. Extending the proposed methods to such heterogeneous settings requires models that can handle very different propagation mechanisms, blockage behaviors, latency constraints, device capabilities, and backhaul requirements. A particularly important direction is the development of unified resource allocation methods that can decide when to use high-bandwidth but fragile links and when to rely on more stable lower-frequency or non-terrestrial links.

- *Joint sensing, communication, and beam-squint exploitation:* The beam-squint and beam-split characterization in Chapter 6 can be used for more than communication beam management. Since frequency-dependent angular shifts contain information about direction, frequency, and geometry, the proposed analytical framework can be extended to RIS-assisted sensing and integrated sensing and communication. In this direction, beam split should not only be treated as an impairment. Instead, the additional frequency-dependent dominant lobes may be exploited to improve user localization, environment sensing, and target detection, especially in wideband and multi-band RIS-assisted systems.

- *Network-level optimization using beam-split-aware RIS models:* Chapter 6 demonstrated the usefulness of beam-split-aware codeword selection in a multi-RIS, multi-user, multi-band example. A broader future direction is to embed the analytical beam-squint and beam-split model into full network-level optimization. This includes joint RIS selection, BS association, codeword selection, frequency allocation, scheduling, and power control. Such a framework would allow the network to decide whether a split beam should be avoided, mitigated, or intentionally exploited to serve users at different frequencies or directions.

- *Practical RIS deployment, regulation, and multi-operator coexistence:* RISs also introduce practical deployment challenges that are not fully captured by mathematical propagation models. In practice, RIS placement depends on site access, ownership of walls and poles, zoning or building regulations, aesthetics, maintenance, power and control connections, and coordination with existing infrastructure. Therefore, RISs cannot be assumed to be deployable at arbitrary locations. Moreover, in multi-operator environments, a surface configured to improve one operator's links may unintentionally reflect interference toward users or BSs belonging to another operator. Future work should incorporate deployment constraints, ownership models, inter-operator interference, pilot contamination, and coordination mechanisms into RIS-assisted MBN optimization [148].

Bibliography

[1] International Telecommunication Union Radiocommunication Sector (ITU-R), "Future technology trends of terrestrial international mobile telecommunications systems towards 2030 and beyond," International Telecommunication Union, ITU-R Report M.2516-0, November 2022, accessed: 2025-07-14. [Online]. Available: https://www.itu.int/dms_pub/itu-r/opb/rep/R-REP-M.2516-2022-PDF-E.pdf

[2] M. Rasti, S. K. Taskou, H. Tabassum, and E. Hossain, "Evolution toward 6G multi-band wireless networks: A resource management perspective," *IEEE Wirel. Commun.*, vol. 29, no. 4, pp. 118–125, Aug. 2022.

[3] S. Kang, M. Mezzavilla, S. Rangan, A. Madanayake, S. B. Venkatakrishnan, G. Hellbourg, M. Ghosh, H. Rahmani, and A. Dhananjay, "Cellular wireless networks in the upper mid-band," *IEEE Open Journal of the Communications Society*, 2024.

[4] E. Björnson, F. Kara, N. Kolomvakis, A. Kosasih, P. Ramezani, and M. B. Salman, "Enabling 6g performance in the upper mid-band through gigantic mimo," *arXiv preprint arXiv:2407.05630*, 2024.

[5] D. M. Slocum, E. J. Slingerland, R. H. Giles, and T. M. Goyette, "Atmospheric absorption of terahertz radiation and water vapor continuum effects," *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 127, pp. 49–63, 2013.

[6] C. Slezak, V. Semkin, S. Andreev, Y. Koucheryavy, and S. Rangan, "Empirical effects of dynamic human-body blockage in 60 GHz communications," *IEEE Commun. Mag.*, vol. 56, no. 12, pp. 60–66, Dec 2018.

[7] C. Han and I. F. Akyildiz, "Distance-aware bandwidth-adaptive resource allocation for wireless systems in the terahertz band," *IEEE Trans. Terahertz Sci. Technol.*, vol. 6, no. 4, pp. 541–553, Jul. 2016.

[8] L. Xu, M. Chen, M. Chen, Z. Yang, C. Chaccour, W. Saad, and C. S. Hong, "Joint location, bandwidth and power optimization for THz-Enabled UAV communications," *IEEE Commun. Lett.*, vol. 25, no. 6, pp. 1984–1988, Jun. 2021.

[9] C. Jeong, C.-J. Chun, W.-Y. Shin, and I.-M. Kim, "Transport capacity optimization for resource allocation in Tera-IoT networks," *IEEE Internet Things J.*, vol. 9, no. 16, pp. 15 270–15 284, Aug. 2022.

[10] A. Shafie, N. Yang, C. Han, and J. M. Jornet, "Novel spectrum allocation among multiple transmission windows for terahertz communication systems," *IEEE Trans. Veh. Technol.*, vol. 71, no. 12, pp. 13 415–13 421, Dec. 2022.

[11] A. Shafie, N. Yang, S. A. Alvi, C. Han, S. Durrani, and J. M. Jornet, "Spectrum allocation with adaptive sub-band bandwidth for terahertz communication systems," *IEEE Trans. on Commun.*, vol. 70, no. 2, pp. 1407–1422, Feb. 2021.

[12] H. Zhang, Y. Duan, K. Long, and V. C. M. Leung, "Energy efficient resource allocation in terahertz downlink NOMA systems," *IEEE Trans. on Commun.*, vol. 69, no. 2, pp. 1375–1384, Feb. 2021.

[13] Y. Zhang, H. Zhang, and K. Long, "Energy efficient resource allocation in cache based terahertz vehicular networks: A mean-field game approach," *IEEE Trans. Veh. Technol.*, vol. 70, no. 6, pp. 5275–5285, Jun. 2021.

[14] A. Moldovan, P. Karunakaran, I. F. Akyildiz, and W. H. Gerstacker, "Coverage and achievable rate analysis for indoor terahertz wireless networks," in *2017 IEEE International Conference on Communications (ICC)*, Paris, France, May 2017, pp. 1–7.

[15] H. Zhang, H. Zhang, W. Liu, K. Long, J. Dong, and V. C. M. Leung, "Energy efficient user clustering, hybrid precoding and power optimization in terahertz MIMO-NOMA systems," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 9, pp. 2074–2085, Sep 2020.

[16] N. Hassan, M. T. Hossan, and H. Tabassum, "User association in coexisting RF and terahertz networks in 6G," in *Proc. IEEE Can. Conf. Elect. Comput. Eng.*, London, ON, Canada, Aug. 2020, pp. 1–5.

[17] Z. Yan and H. Tabassum, "Reinforcement learning for joint V2I network selection and autonomous driving policies," in *Proc. IEEE Global Commun. Conf.*, Dec. Rio de Janeiro, Brazil, 2022, pp. 1241–1246.

[18] M. Amjad, H. K. Qureshi, S. A. Hassan, A. Ahmad, and S. Jangsher, "Optimization of MAC frame slots and power in hybrid VLC/RF networks," *IEEE Access*, vol. 8, pp. 21 653–21 664, Jan 2020.

[19] M. Kashef, M. Ismail, M. Abdallah, K. A. Qaraqe, and E. Serpedin, "Energy efficient resource allocation for mixed RF/VLC heterogeneous wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 883–893, Apr 2016.

[20] S. Aboagye, A. Ibrahim, T. M. N. Ngatched, A. R. Ndjiongue, and O. A. Dobre, "Design of energy efficient hybrid VLC/RF/PLC communication system for indoor networks," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 143–147, Feb 2020.

[21] C. Chaieb, F. Abdelkefi, and W. Ajib, "On the sum-rate maximization in multi-access and multi-band wireless networks," in *ICC 2022 - IEEE International Conference on Communications*, Seoul, Korea, Republic of, May 2022, pp. 4902–4907.

[22] H. Z. Khan, M. Ali, M. Naeem, I. Rashid, A. N. Akhtar, and F. Akram, "Joint DL/UL decouple user association in microwave and mmWave enabled beyond 5G heterogeneous networks," *IEEE Access*, vol. 9, pp. 134 703–134 715, Oct. 2021.

[23] S. Enayati and H. Saeedi, "Deployment of hybrid FSO/RF links in backhaul of relay-based rural area cellular networks: Advantages and performance analysis," *IEEE Commun. Lett.*, vol. 20, no. 9, pp. 1824–1827, Sep. 2016.

[24] S. Bayati, M. Darabi, A. Mostafa, and L. Lampe, "Resource allocation in C-RAN with hybrid RF/FSO and full-duplex self-backhauling radio units," in *ICC 2021 - IEEE International Conference on Communications*, Montreal, QC, Canada, Jun. 2021, pp. 1–7.

[25] X. L. et al., "Multi-user modular XL-MIMO communications: Near-field beam focusing pattern and user grouping," *IEEE Trans. Wirel. Commun.*, vol. 23, no. 10, Part 2, pp. 13 766–13 781, Oct. 2024.

[26] H. Z. et al., "Beam focusing for near-field multiuser MIMO communications," *IEEE Trans. Wirel. Commun.*, vol. 21, no. 9, pp. 7476–7490, Sep. 2022.

[27] Y. Zhang and C. You, "SWIPT in mixed near- and far-field channels: Joint beam scheduling and power allocation," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 6, pp. 1683–1698, Jun. 2024.

[28] Z. T. et al., "Dynamic precoding for near-field secure communications: Implementation and performance analysis," *IEEE Internet Things J.*, May 2025.

[29] M. Jafri, S. Srivastava, N. K. Venkategowda, A. K. Jagannatham, and L. Hanzo, "Cooperative hybrid transmit beamforming in cell-free mmwave mimo networks," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 5, pp. 6023–6038, 2023.

[30] C. Fang, B. Makki, J. Li, and T. Svensson, "Hybrid precoding in cooperative millimeter wave networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 8, pp. 5373–5388, 2021.

[31] Z. Zhu, L. Wang, X. Wang, B. Tan, and S. Jin, "Cell-free terahertz networks: A spatial-spectral approach," *IEEE Wireless Communications Letters*, 2023.

[32] Y. Modami, H. Beiranvand, and M. T. Dabiri, "Cell-free infrastructure-to-vehicle communication: Empowering with hybrid fso/thz backhaul," *IEEE Wireless Communications Letters*, 2024.

[33] M. A. Saeidi, H. Shoaib, and H. Tabassum, "A tractable handoff-aware rate outage approximation with applications to thz-enabled vehicular network optimization," in *GLOBECOM 2023 - 2023 IEEE Global Communications Conference*, 2023, pp. 5092–5097.

[34] H. A. Ammar, R. Adve, S. Shahbazpanahi, G. Boudreau, and K. V. Srinivas, "Handoffs in user-centric cell-free mimo networks: A pomdp framework," *IEEE Transactions on Wireless Communications*, 2024.

[35] W. Bao and B. Liang, "Optimizing cluster size through handoff analysis in user-centric cooperative wireless networks," *IEEE Trans. Wirel. Commun.*, vol. 17, no. 2, pp. 766–778, Feb. 2018.

[36] Z. Zeng, M. D. Soltani, Y. Wang, X. Wu, and H. Haas, "Realistic indoor hybrid wifi and ofdma-based lifi networks," *IEEE Transactions on Communications*, vol. 68, no. 5, pp. 2978–2991, 2020.

[37] F. Wang, Z. Wang, C. Qian, L. Dai, and Z. Yang, "Efficient vertical handover scheme for heterogeneous vlc-rf systems," *Journal of Optical Communications and Networking*, vol. 7, no. 12, pp. 1172–1180, 2015.

[38] Y. Wang, D. A. Basnayaka, X. Wu, and H. Haas, "Optimization of load balancing in hybrid lifi/rf networks," *IEEE Transactions on Communications*, vol. 65, no. 4, pp. 1708–1720, 2017.

[39] O. Semiari, W. Saad, M. Bennis, and B. Maham, "Caching meets millimeter wave communications for enhanced mobility management in 5g networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 2, pp. 779–793, 2018.

[40] M. A. Dastgheib, H. Beyranvand, J. A. Salehi, and M. Maier, "Mobility-aware resource allocation in vlc networks using t-step look-ahead policy," *J. Lightwave Technol.*, vol. 36, no. 23, pp. 5358–5370, Dec 2018. [Online]. Available: https://opg.optica.org/jlt/abstract.cfm?URI=jlt-36-23-5358

[41] Q. Ren, O. Abbasi, G. K. Kurt, H. Yanikomeroglu, and J. Chen, "Handoff-aware distributed computing in high altitude platform station (haps)–assisted vehicular networks," *IEEE Transactions on Wireless Communications*, vol. 22, no. 12, pp. 8814–8827, 2023.

[42] M. Gupta, S. Chinchali, P. Varkey, and J. G. Andrews, "Forecaster-aided user association and load balancing in multi-band mobile networks," *IEEE Transactions on Wireless Communications*, 2024.

[43] R. Su *et al.*, "Wideband precoding for RIS-aided THz communications," *IEEE Trans. Commun.*, vol. 71, no. 6, pp. 3592–3604, 2023.

[44] W. Yan *et al.*, "Beamforming analysis and design for wideband THz reconfigurable intelligent surface communications," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 8, pp. 2306–2320, 2023.

[45] G. Sun *et al.*, "Beamforming design for the distributed RISs-aided THz communications with double-layer true time delays," *IEEE Trans. Veh. Technol.*, vol. 73, no. 3, pp. 3886–3900, 2023.

[46] X. Su *et al.*, "Joint precoding for RIS-assisted wideband THz cell-free massive MIMO systems," *IEEE Internet Things J.*, vol. 11, no. 20, pp. 33 361–33 370, 2024.

[47] W. Yan *et al.*, "Wideband beamforming for STAR-RIS-assisted THz communications with three-side beam split," *IEEE Trans. Commun.*, vol. 73, no. 3, pp. 2133–2145, 2024.

[48] H. Sun *et al.*, "Time-delay unit based beam squint mitigation for RIS-aided communications," *IEEE Commun. Lett.*, vol. 26, no. 9, pp. 2220–2224, 2022.

[49] L. Yashvanth *et al.*, "Mitigating spatial-wideband and beam-split effects via distributed IRSs: Design and analysis," *IEEE Trans. Signal Process.*, 2025.

[50] J. Li *et al.*, "User sensing in RIS-aided wideband mmWave system with beam-squint and beam-split," *IEEE Trans. Commun.*, vol. 73, no. 2, pp. 1304–1319, 2024.

[51] H. Luo *et al.*, "Beam squint assisted user localization in near-field integrated sensing and communications systems," *IEEE Trans. Wireless Commun.*, vol. 23, no. 5, pp. 4504–4517, 2023.

[52] A. M. Elbir *et al.*, "Antenna selection with beam squint compensation for integrated sensing and communications," *IEEE J. Sel. Topics Signal Process.*, vol. 18, no. 5, pp. 857–870, 2024.

[53] ——, "Terahertz-band channel and beam split estimation via array perturbation model," *IEEE Open J. Commun. Soc.*, vol. 4, pp. 892–907, 2023.

[54] A. M. Elbir and S. Chatzinotas, "BSA-OMP: Beam-split-aware orthogonal matching pursuit for THz channel estimation," *IEEE Wireless Commun. Lett.*, vol. 12, no. 4, pp. 738–742, 2023.

[55] X. Su *et al.*, "Channel estimation for RIS assisted THz systems with beam split," *IEEE Commun. Lett.*, vol. 28, no. 3, pp. 637–641, 2024.

[56] ——, "Two-phase channel estimation for RIS-assisted THz systems with beam split," *IEEE Trans. Wireless Commun.*, 2025.

[57] E. Björnson *et al.*, "Reconfigurable intelligent surfaces: A signal processing perspective with wireless applications," *IEEE Signal Process. Mag.*, vol. 39, no. 2, pp. 135–158, 2022.

[58] S. Abeywickrama, R. Zhang, Q. Wu, and C. Yuen, "Intelligent reflecting surface: Practical phase shift model and beamforming optimization," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5849–5863, 2020.

[59] H. Li, W. Cai, Y. Liu, M. Li, Q. Liu, and Q. Wu, "Intelligent reflecting surface enhanced wideband MIMO-OFDM communications: From practical model to reflection optimization," *IEEE Trans. Commun.*, vol. 69, no. 7, pp. 4807–4820, 2021.

[60] W. Cai, H. Li, M. Li, and Q. Liu, "Practical modeling and beamforming for intelligent reflecting surface aided wideband systems," *IEEE Commun. Lett.*, vol. 24, no. 7, pp. 1568–1571, 2020.

[61] W. Cai, R. Liu, M. Li, Y. Liu, Q. Wu, and Q. Liu, "IRS-assisted multicell multiband systems: Practical reflection model and joint beamforming design," *IEEE Trans. Commun.*, vol. 70, no. 6, pp. 3897–3911, 2022.

[62] O. Abbas *et al.*, "Unit cell phase-frequency profile optimization in RIS-assisted wide-band OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 24, no. 1, pp. 735–748, 2024.

[63] J. C. Bezdek and R. J. Hathaway, "Some notes on alternating optimization," in *Advances in Soft Computing — AFSS 2002*, N. R. Pal and M. Sugeno, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 288–300.

[64] W. Dinkelbach, "On nonlinear fractional programming," *Zeitschrift f¨ur Operations Res.*, vol. 133, no. 7, pp. 492–498, Mar. 1967.

[65] K. Shen and W. Yu, "Fractional programming for communication systems—Part I: Power control and beamforming," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2616–2630, May 2018.

[66] ——, "Fractional programming for communication systems—Part II: Uplink scheduling via matching," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2631–2644, May 2018.

[67] M. Razaviyayn, "Successive convex approximation: Analysis and applications," Ph.D. dissertation, University of Minnesota, 2014.

[68] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, May/Jun. 2020.

[69] M. A. Saeidi, H. Shoaib, and H. Tabassum, "A tractable handoff-aware rate outage approximation with applications to THz-enabled vehicular network optimization," 2023.

[70] H. Zarini, N. Gholipoor, M. R. Mili, M. Rasti, H. Tabassum, and E. Hossain, "Resource management for multiplexing eMBB and URLLC services over RIS-Aided THz communication," *IEEE Trans. Commun.*, vol. 71, no. 2, pp. 1207–1225, Feb. 2023.

[71] B. Ning, Z. Tian, W. Mei, Z. Chen, C. Han, S. Li, J. Yuan, and R. Zhang, "Beamforming technologies for ultra-massive MIMO in terahertz communications," *IEEE Open Journal of the Communications Society*, vol. 4, pp. 614–658, 2023.

[72] S. Aboagye, M. Amin Saeidi, H. Tabassum, Y. Tayyar, E. Hossain, H.-C. Yang, and M.-S. Alouini, "Multi-band wireless communication networks: Fundamentals, challenges, and resource allocation," *IEEE Trans. Commun.*, vol. 72, no. 7, pp. 4333–4352, Jul. 2024.

[73] M. A. Saeidi, H. Tabassum, and M.-S. Alouini, "Multi-band wireless networks: Architectures, challenges, and comparative analysis," *IEEE Commun. Mag.*, pp. 1–7, Jul. 2023.

[74] Q. Yuan, H. Liu, M. Xu, Y. Wu, L. Xiao, and T. Jiang, "Deep learning-based hybrid precoding for terahertz massive MIMO communication with beam squint," *IEEE Commun. Lett.*, vol. 27, no. 1, pp. 175–179, Jan. 2023.

[75] B. Ning, W. Mei, L. Zhu, Z. Chen, and R. Zhang, "Max-min beamformer for THz wideband communications," in *2023 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2023, pp. 1747–1752.

[76] M. A. Saeidi, H. Tabassum, and M. Alizadeh, "Molecular absorption-aware user assignment, spectrum, and power allocation in dense thz networks with multi-connectivity," *IEEE Transactions on Wireless Communications*, 2024.

[77] S. Tarboush, H. Sarieddeen, H. Chen, M. H. Loukil, H. Jemaa, M.-S. Alouini, and T. Y. Al-Naffouri, "TeraMIMO: A channel simulator for wideband ultra-massive MIMO terahertz communications," *IEEE Trans. Veh. Technol.*, vol. 70, no. 12, pp. 12 325–12 341, Dec. 2021.

[78] S. B. Melhem and H. Tabassum, "User pairing and outage analysis in multi-carrier NOMA-THz networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 5, pp. 5546–5551, May 2022.

[79] M. Di Renzo, "Stochastic geometry modeling and analysis of multi-tier millimeter wave cellular networks," *IEEE Trans. on Wireless Commun.*, vol. 14, no. 9, pp. 5038–5057, Sep. 2015.

[80] M. T. Hossan and H. Tabassum, "Mobility-aware performance in hybrid RF and terahertz wireless networks," *IEEE Trans. on Commun.*, vol. 70, no. 2, pp. 1376–1390, Feb. 2021.

[81] W. Chen, L. Li, Z. Chen, and T. Q. S. Quek, "Coverage modeling and analysis for outdoor THz networks with blockage and molecular absorption," *IEEE Wireless Commun. Lett.*, vol. 10, no. 5, pp. 1028–1031, May 2021.

[82] D. Kumar, J. Kaleva, and A. Tölli, "Blockage-aware reliable mmWave access via coordinated multi-point connectivity," *IEEE Trans. on Wireless Commun.*, vol. 20, no. 7, pp. 4238–4252, Jul. 2021.

[83] "The HITRAN2020 molecular spectroscopic database," *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 277, p. 107949, 2022.

[84] A. Schrijver, *Theory of linear and integer programming.* John Wiley & Sons, 1998.

[85] J. Lee, *A first course in combinatorial optimization.* Cambridge University Press, 2004, vol. 36.

[86] M. ApS, "The optimizer for mixed-integer problems," 2023. [Online]. Available: https://docs.mosek.com/latest/cxxfusion/mip-optimizer.html

[87] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011. [Online]. Available: http://dx.doi.org/10.1561/2200000016

[88] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization.* Cambridge University Press, 2004.

[89] V. N. Moothedath and S. Bhashyam, "Distributed pareto optimal beamforming for the MISO multi-band multi-cell downlink," *IEEE Transactions on Wireless Communications*, vol. 19, no. 11, pp. 7196–7209, 2020.

[90] M. A. Saeidi, M. J. Emadi, H. Masoumi, M. R. Mili, D. W. K. Ng, and I. Krikidis, "Weighted sum-rate maximization for multi-IRS-assisted full-duplex systems with hardware impairments," *IEEE Transactions on Cognitive Communications and Networking*, vol. 7, no. 2, pp. 466–481, 2021.

[91] T. Mao, Q. Wang, and Z. Wang, "Spatial modulation for terahertz communication systems with hardware impairments," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 4, pp. 4553–4557, 2020.

[92] P. Yang, L. Yang, and S. Wang, "Performance analysis for RIS-aided wireless systems with imperfect CSI," *IEEE Wireless Communications Letters*, vol. 11, no. 3, pp. 588–592, 2022.

[93] M. A. Saeidi and H. Tabassum, "Resource allocation and performance analysis of hybrid RSMA-NOMA in the downlink," in *2023 IEEE 34th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2023, pp. 1–6.

[94] S. Aboagye, T. M. N. Ngatched, O. A. Dobre, and H. V. Poor, "Energy-efficient resource allocation for aggregated RF/VLC systems," *IEEE Transactions on Wireless Communications*, vol. 22, no. 10, pp. 6624–6640, 2023.

[95] "Achieving multi-band 5G NR NSA - 28 GHz 5G - Ericsson," Available: https://www.ericsson.com/en/5g/5g-technology-milestones/

global-mobile-industry-leaders-achieve-multi-band-5g-nr-nsa, Dec. 2017, accessed: 2023-01-25.

[96] Huawei. 5g-oriented site evolution. Accessed: 2023-01-30. [Online]. Available: https://carrier.huawei.com/en/trends-and-insights/emsite/5g-oriented-site-evolution

[97] M. Obeed, A. M. Salhab, M.-S. Alouini, and S. A. Zummo, "On optimizing VLC networks for downlink multi-user transmission: A survey," *IEEE Commun. Surveys & Tutorials*, vol. 21, no. 3, pp. 2947–2976, 2019.

[98] M. A. Saeidi, H. Tabassum, and M.-S. Alouini, "Multi-band wireless networks: Architectures, challenges, and comparative analysis," *IEEE Communications Magazine*, vol. 62, no. 1, pp. 80–86, 2024.

[99] I. F. Akyildiz, J. M. Jornet, and S. Nie, "A new cubesat design with reconfigurable multi-band radios for dynamic spectrum satellite communication networks," *Ad Hoc Networks*, vol. 86, pp. 166–178, 2019.

[100] P. K. Singya, B. Makki, A. D'Errico, and M.-S. Alouini, "Hybrid FSO/THz-based backhaul network for mmWave terrestrial communication," *arXiv preprint arXiv:2204.08357*, 2022.

[101] I. F. Akyildiz, A. Kak, and S. Nie, "6G and beyond: The future of wireless communications systems," *IEEE Access*, vol. 8, pp. 133 995–134 030, 2020.

[102] S. Tarboush, H. Sarieddeen, H. Chen, M. H. Loukil, H. Jemaa, M.-S. Alouini, and T. Y. Al-Naffouri, "Teramimo: A channel simulator for wideband ultra-massive MIMO terahertz communications," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 12, pp. 12 325–12 341, 2021.

[103] H. Shoaib and H. Tabassum, "Optimization of speed and network deployment for reliable V2I communication in the presence of handoffs and interference," *IEEE Wireless Commun. Lett.*, vol. 12, no. 6, pp. 1051–1055, 2023.

[104] M. Monemi and H. Tabassum, "Performance of uav-assisted d2d networks in the finite block-length regime," *IEEE Trans. on Commun.*, vol. 68, no. 11, pp. 7270–7285, 2020.

[105] S. Zarandi and H. Tabassum, "Delay minimization in sliced multi-cell mobile edge computing (mec) systems," *IEEE Communications Letters*, vol. 25, no. 6, pp. 1964–1968, 2021.

[106] M. A. Saeidi and H. Tabassum, "Resource allocation and performance analysis of hybrid RSMA-NOMA in the downlink," Toronto, ON, Canada, pp. 1–6, Sep 2023.

[107] F. Agnoletto *et al.* (Jan. 2021) The economics of mmwave 5G - GSMA intelligence. [Online]. Available: https://data.gsmaintelligence.com/api-web/v2/research-file-download?id=59768858&file=210121-Economics-of-mmWave.pdf

[108] "5G action plan review for europe: final report," Sep. 2020. [Online]. Available: https://www.qualcomm.com/content/dam/qcomm-martech/dm-assets/documents/5g_cost_benefit_analysis_for_europe_final_report_290121.pdf

[109] M. A. Saeidi *et al.*, "Molecular absorption-aware user assignment, spectrum, and power allocation in dense THz networks with multi-connectivity," *IEEE Trans. Wireless Commun.*, vol. 23, no. 11, pp. 16 404–16 420, 2024.

[110] A. Weissberger. (2023, Jul.) New ITU report in progress: Technical feasibility of IMT in bands above 100 GHz (92 GHz and 400 GHz). TechBlog - IEEE Communications Society. Blog post. [Online]. Available: https://techblog.comsoc.org/category/imt-above-100-ghz/

[111] S. Aboagye, T. M. Ngatched, O. A. Dobre, and H. V. Poor, "Energy-efficient resource allocation for aggregated rf/vlc systems," *IEEE Transactions on Wireless Communications*, vol. 22, no. 10, pp. 6624–6640, 2023.

[112] W. B. et al., "Photonics-assisted millimeter-wave multiband integrated sensing and communication system using coherent receiving," *IEEE J. Sel. Top. Quantum Electron.*, vol. 29, no. 6, pp. 1–11, Nov. 2023.

[113] I. F. e. a. Akyildiz, "A new cubesat design with reconfigurable multi-band radios for dynamic spectrum satellite communication networks," *Ad Hoc Networks*, vol. 86, pp. 166–178, 2019.

[114] M. A. Saeidi and H. Tabassum, "Resource allocation in cooperative mid-band/THz networks in the presence of mobility," *IEEE Trans. Wireless Commun.*, vol. 25, pp. 5046–5062, 2026.

[115] Y. Liu, Z. Wang, J. Xu, C. Ouyang, X. Mu, and R. Schober, "Near-field communications: A tutorial review," *IEEE Open Journal of the Communications Society*, vol. 4, pp. 1999–2049, 2023.

[116] A. Liao, Z. Gao, D. Wang, H. Wang, H. Yin, D. W. K. Ng, and M.-S. Alouini, "Terahertz ultra-massive mimo-based aeronautical communications in space-air-ground integrated networks," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 6, pp. 1741–1767, 2021.

[117] F. P. et al., "A novel mobility induced channel prediction mechanism for vehicular communications," *IEEE Trans. Wirel. Commun.*, vol. 22, no. 5, pp. 3488–3502, May 2022.

[118] W. M. et al., "UAV-assisted communications in SAGIN-ISAC: Mobile user tracking and robust beamforming," *IEEE J. Sel. Areas Commun.*, vol. 43, no. 1, pp. 186–200, Jan. 2025.

[119] J. F. et al., "Mobility-aware predictive beamforming design in sensing-assisted SWIPT systems," *IEEE Wireless Commun. Lett.*, vol. 14, no. 5, pp. 1476–1480, May 2025.

[120] D. Y. et al., "Temporal-frequency domain channel prediction for LEO satellite communication system: A novel TFformer structure," *IEEE Trans. Wirel. Commun.*, 2025, early Access.

[121] B. Barahimi, H. Singh, H. Tabassum, O. Waqar, and M. Omer, "Rsc-net: Dynamic csi compression for cloud-based wifi sensing," *arXiv preprint arXiv:2402.04888*, 2024.

[122] "The HITRAN2020 molecular spectroscopic database," *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 277, p. 107949, 2022.

[123] L. Yan, C. Han, and J. Yuan, "Energy-efficient dynamic-subarray with fixed true-time-delay design for terahertz wideband hybrid beamforming," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 10, pp. 2840–2854, 2022.

[124] M. T. Hossan and H. Tabassum, "Mobility-aware performance in hybrid rf and terahertz wireless networks," *IEEE Transactions on Communications*, vol. 70, no. 2, pp. 1376–1390, 2021.

[125] J. Ye, S. Dang, B. Shihada, and M.-S. Alouini, "Modeling co-channel interference in the thz band," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 7, pp. 6319–6334, 2021.

[126] A. Pradhan, M. A. Abd-Elmagid, H. S. Dhillon, and A. F. Molisch, "Robust optimization of ris in terahertz under extreme molecular re-radiation manifestations," *IEEE Transactions on Wireless Communications*, vol. 23, no. 3, pp. 1783–1797, 2023.

[127] F. Sohrabi and W. Yu, "Hybrid analog and digital beamforming for mmwave OFDM large-scale antenna arrays," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 7, pp. 1432–1443, 2017.

[128] M. Cococcioni and L. Fiaschi, "The big-m method with the numerical infinite m," *Optimization Letters*, vol. 15, no. 7, pp. 2455–2468, 2021.

[129] M. Grant and S. Boyd, "CVX: Matlab Software for disciplined convex programming, version 2.1," Available at http://cvxr.com/cvx, Mar. 2014.

[130] Y. S. et al., "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Feb. 2017.

[131] M. A. Saeidi and H. Tabassum, "Resource allocation and performance analysis of hybrid rsma-noma in the downlink," in *2023 IEEE 34th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2023, pp. 1–6.

[132] M. e. a. Razaviyayn, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 1126–1153, 2013.

[133] Y. Ye, *Interior point algorithms: theory and analysis*. John Wiley & Sons, 2011.

[134] E. D. e. a. Andersen, "On implementing a primal-dual interior-point method for conic quadratic optimization," *Mathematical Programming*, vol. 95, pp. 249–277, 2003.

[135] W. S. et al., "Movement aware CoMP handover in heterogeneous ultra-dense networks," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 340–352, Jan. 2021.

[136] R. T. Marler and J. S. Arora, "The weighted sum method for multi-objective optimization: new insights," *Structural and multidisciplinary optimization*, vol. 41, pp. 853–862, 2010.

[137] G. Interdonato, M. Karlsson, E. Björnson, and E. G. Larsson, "Local partial zero-forcing precoding for cell-free massive mimo," *IEEE Transactions on Wireless Communications*, vol. 19, no. 7, pp. 4758–4774, 2020.

[138] N. Deshpande, M. R. Castellanos, S. R. Khosravirad, J. Du, H. Viswanathan, and R. W. Heath, "A wideband generalization of the near-field region for extremely large phased-arrays," *IEEE Wireless Commun. Lett.*, vol. 12, no. 3, pp. 515–519, Mar. 2023.

[139] X. M. et al., "Cooperative beamforming for RIS-aided cell-free massive MIMO networks," *IEEE Trans. Wirel. Commun.*, vol. 22, no. 11, pp. 7243–7258, Nov. 2023.

[140] G. Z. et al, "A framework of robust transmission design for IRS-aided MISO communications with imperfect cascaded channels," *IEEE Trans. Signal Process.*, vol. 68, no. 17, pp. 5092–5106, Sep. 2020.

[141] M. Alizadeh and H. Tabassum, "Power control with QoS guarantees: A differentiable projection-based unsupervised learning framework," *IEEE Trans. Commun.*, vol. 71, no. 8, pp. 4605–4619, Aug. 2023.

[142] C. Edwards, *Advanced Calculus of Several Variables*, ser. Dover Books on Mathematics. Dover Publications, 2012. [Online]. Available: https://books.google.ca/books?id=sZIFcJ8DJAIC

[143] S. Lv *et al.*, "RIS-aided near-field MIMO communications: Codebook and beam training design," *IEEE Trans. Wireless Commun.*, vol. 23, no. 9, pp. 12 531–12 546, 2024.

[144] L. C. Evans and R. F. Gariepy, *Measure Theory and Fine Properties of Functions, Revised Edition*, 1st ed. Chapman and Hall/CRC, 2015. [Online]. Available: https://doi.org/10.1201/b18333

[145] C. Berg, "Fourier analysis," 2009. [Online]. Available: https://noter.math.ku.dk/indentryk/fouan12.pdf

[146] Z. Yan, Y. Huang, J. Pei, H. Tabassum, and L. Chiaraviglio, "Emfusion: Uncertainty-aware conditional diffusion model for multivariate narrow-band exposure forecasting," *IEEE Transactions on Network Science and Engineering*, pp. 1–19, 2026.

[147] X. Mootoo, H. Tabassum, and L. Chiaraviglio, "Emforecaster: A deep learning framework for time series forecasting in wireless networks with distribution-free uncertainty quantification," *IEEE Transactions on Network Science and Engineering*, vol. 13, pp. 1207–1225, 2026.

[148] S. Bani Melhem, H. Tabassum, L. Chiaraviglio, and U. Trang Nguyen, "Joint coverage and emf compliance analysis in ris-assisted terahertz networks with optimized ris deployment," *IEEE Open Journal of the Communications Society*, vol. 6, pp. 8816–8834, 2025.