

Resource Allocation for Intelligent Reflecting Surface Aided Wireless Networks

Joshua Jalali



Supervisors **Prof. dr. Jeroen Famaey — Prof. dr. Rafael Berkvens**

Thesis submitted in fulfilment of the requirements for the degree of doctor in applied engineering
Faculty of Applied Engineering — Antwerp, 2024



**University
of Antwerp**



Faculty of Applied Engineering
doctor in applied engineering

Resource Allocation for Intelligent Reflecting Surface Aided Wireless Networks

Thesis submitted in fulfilment of the requirements for the degree of
doctor in applied engineering
at University of Antwerp

Joshua Jalali

Antwerp, 2024

Supervisors
Prof. dr. Jeroen Famaey
Prof. dr. Rafael Berkvens

Jury

Chairman

Prof. dr. Maarten Weyn, University of Antwerp, Belgium

Supervisors

Prof. dr. Jeroen Famaey, University of Antwerp, Belgium

Prof. dr. Rafael Berkvens, University of Antwerp, Belgium

Members

Prof. dr. Erik Mannens, University of Antwerp, Belgium

Prof. dr. David Plets, Ghent University, Belgium

Dr. Sergi Abadal, Universitat Politècnica de Catalunya, Spain

Dr. Gerhard Schoenthal, Virginia Diodes, Inc., USA

Contact

Joshua Jalali

University of Antwerp

Faculty of Applied Engineering — Electronics and ICT Engineering Technology

IDLab Research Group

Sint-Pietersvliet 7, 2020 Antwerp, Belgium

Email: jala.jalali@uantwerpen.be

josh@juliaspace.com

© 2024 Joshua Jalali

All rights reserved.

ISBN 978-90-57288-555

Wettelijk depot D/2024/12.293/21



Abstract

In today's world, staying connected is more important than ever, but achieving reliable wireless communication everywhere can be a challenge. This dissertation introduces a cutting-edge technology known as Intelligent Reflecting Surfaces (IRSs) that promises to revolutionize how we connect. Imagine a smart, invisible "mirror" that can bend and direct wireless signals precisely where needed, overcoming obstacles and ensuring your device always gets a strong connection. That is what the IRS does.

IRS, at its core, is a sophisticated planar array, composed of numerous passive or active elements capable of individually manipulating electromagnetic waves to reshape the wireless signal propagation environment. By smartly adjusting the phase and amplitude of these elements, an IRS can seamlessly steer signals toward intended receivers, effectively creating optimized communication paths even in scenarios where direct Line-of-Sight (LoS) is obstructed. This ability to mold the propagation environment on demand, without additional energy for signal transmission, enables the IRS to enhance connectivity in diverse environments, from densely built urban areas to indoor spaces. Furthermore, the ability of the IRS to operate without the need for active power amplification allows for a significant reduction in energy consumption, making it an eco-friendly solution for extending and improving wireless network coverage.

In this dissertation, IRS is presented as a key enabler for a myriad of advanced technologies, unlocking new potentials across various high-tech fields by enhancing their performance and efficiency. By strategically manipulating electromagnetic waves, IRS provides a solution to enhance power efficiency in multi-user Simultaneous Wireless Information and Power Transfer (SWIPT) networks. This capability allows for a steady flow of information and power transfer, illustrating the dual capability of the IRS to support energy harvesting and data transmission. Furthermore, the integration of IRS into Ultra-Reliable Low-Latency Communication (URLLC) and Machine Type Communication (MTC) systems emerges as a game-changer, significantly reducing latency and increasing reliability. IRS can significantly benefit Virtual Reality (VR) users facing considerable path loss or blockages, ensuring immersive experiences without latency or loss of quality.

IRS also enhances Mobile Edge Computing (MEC) by optimizing signal delivery for efficient edge data processing. These improvements are essential for critical applications requiring instantaneous feedback and high levels of data integrity, such as autonomous vehicles and industrial automation, underpinning the role of the IRS in facilitating the next wave of communication needs. This work delves into the strategic deployment of IRS across a broad frequency spectrum, from Frequency Range 1 (FR1) to Frequency Range 2 (FR2), extending into the higher frequency domains of millimeter-Wave (mmWave) and TeraHertz (THz) frequencies, illustrating its profound impact on the future of telecommunications.

In order to investigate the performance of IRS-assisted networks, this dissertation defines a range of Key Performance Indicators (KPIs), such as data rate, power efficiency, energy efficiency, Signal-to-Interference-plus-Noise Ratio (SINR), transmit signal power budget, and received power strength. These KPIs serve as metrics to assess and optimize the network's performance based on designing an efficient resource allocation policy. Non-linear, non-convex, and Mixed Integer Nonlinear Programming (MINLP) problems arise when addressing the resource allocation optimization problem. These problems are Non-deterministic Polynomial time (NP)-hard due to the complex relationship between variables and the system's constraints. Given the complexity of these optimization problems, different strategies are used to simplify and approach their solution. By relaxing the objective function (the NPs) and constraints that are non-convex to a more tractable format, the problems became more manageable. This relaxation approach often involved transforming the optimization problem into its convex equivalent or utilizing approximation techniques to linearize or convexify non-convex terms.

Algorithms are developed that are capable of solving the main problem either globally or sub-optimally but sufficiently close to the global optimum. These solutions employ optimization solvers and computer simulations, exploiting advanced mathematical tools and techniques such as the big-M method for linearizing product terms involving binary variables and Successive Convex Approximation (SCA) to obtain convex approximations of non-convex terms. The iterative nature of these solutions allowed for step-by-step refinement, gradually moving towards an optimal configuration of a resource allocation design despite the initial problem's complexity.

Through exhaustive simulations, this dissertation unveils the diverse performance improvements achievable through resource allocation in IRS-assisted networks, providing rich insights into how IRS technology can improve wireless communication systems. These simulations serve as a critical bridge, connecting theoretical predictions with empirical evidence and validating the practical feasibility of the proposed IRS-enhanced network. By exploring various IRS configurations — examining both passive and active types and varying the number of reflective elements — and their implementation in different environments and settings, this study not only confirms the theoretical models' accuracy but also explains the conditions under which IRS deployments yield maximal performance gains, manifesting the IRS versatility in adapting new technologies.

Collectively, this dissertation studies the impact of IRS across a broad range of technologies. By enhancing the performance of SWIPT networks, facilitating URLLC and MTC, enabling MEC, and revolutionizing VR, mmWave, and THz applications, IRS stands at the forefront of wireless communication innovation. This work demonstrates the diverse applications of IRS technology and lays the foundation for future research aimed at utilizing IRS to tackle the dynamic challenges of modern wireless networks. It charts a path toward the development of more robust, efficient, and engaging communication ecosystems.

Samenvatting

In de wereld van vandaag is constant verbonden zijn essentieel, maar overal betrouwbare draadloze communicatie bieden is complex. Deze dissertatie onthult een vooruitstrevende technologie, Intelligent Reflecting Surface (IRS), die draadloze connectiviteit zal transformeren. Denk aan een slimme, onzichtbare "spiegel" die draadloze signalen buigt en richt naar waar nodig, obstakels omzeilt en een sterke verbinding garandeert. Dat is wat IRS doet.

IRS, in zijn kern, is een geavanceerde vlakke antennematrix, bestaande uit talrijke passieve of actieve elementen die in staat zijn om elektromagnetische golven individueel te manipuleren om de draadloze signaalpropagatieomgeving te hervormen. Door de fase en amplitude van deze elementen slim aan te passen, kan een IRS signalen naadloos naar bedoelde ontvangers sturen, waardoor geoptimaliseerde communicatiepaden worden gecreëerd, zelfs in scenario's waar het directe zicht (LoS) wordt belemmerd. Dit vermogen om de propagatieomgeving naar behoefte te vormen, zonder extra energie voor signaaltransmissie, stelt de IRS in staat de connectiviteit in diverse omgevingen te verbeteren, van dichtbebouwde stedelijke gebieden tot binnenshuis. Bovendien maakt het vermogen van de IRS om te werken zonder de noodzaak van actieve vermogensversterking een aanzienlijke vermindering van het energieverbruik mogelijk, waardoor het een milieuvriendelijke oplossing is voor het uitbreiden en verbeteren van de draadloze netwerkdekking.

In deze dissertatie wordt de IRS gepresenteerd als een sleutelfactor voor een breed scala aan geavanceerde technologieën, waardoor nieuwe mogelijkheden worden ontsloten in diverse high-tech velden door hun prestaties en efficiëntie te verbeteren. Door elektromagnetische golven strategisch te manipuleren, biedt IRS een oplossing om de energie-efficiëntie in multi-gebruiker Simultane Draadloze Informatie- en Energieoverdracht (SWIPT)-netwerken te verbeteren. Deze mogelijkheid zorgt voor een constante stroom van informatie- en energieoverdracht, waarbij de dubbele capaciteit van de IRS wordt geïllustreerd om energieoogst en datatransmissie te ondersteunen. Bovendien biedt de integratie van IRS in Ultra-Reliable Low-Latency Communication (URLLC) en Machine Type Communication (MTC)-systemen unieke voordelen, waardoor de vertraging aanzienlijk wordt verminderd en de betrouwbaarheid wordt verhoogd. IRS kan Virtual Reality (VR)-gebruikers die aanzienlijke padverliezen of blokkages tegenkomen aanzienlijk ten goede komen, waarbij realistische virtuele ervaringen zonder vertragingvertraging of kwaliteitsverlies worden gegarandeerd.

IRS verbetert ook Mobile Edge Computing (MEC) door de signaallevering te optimaliseren voor efficiënte gegevensverwerking aan de rand van het netwerk. Deze verbeteringen zijn essentieel voor kritieke toepassingen die onmiddellijke feedback en hoge niveaus van gegevensintegriteit vereisen, zoals autonome voertuigen en industriële automatisering, waardoor de rol van IRS wordt onderstreept bij het faciliteren van de volgende golf van communicatiebehoeften. Dit werk verdiept zich in de strategische inzet van IRS over een breed frequentiespectrum, van Frequentiebereik 1 (FR1) tot Frequentiebereik 2 (FR2), en breidt uit naar de hogere frequentiedomeinen van millimetergolf (mmWaves) en terahertz

(THz)-frequenties, waarbij de diepgaande impact op de toekomst van telecommunicatie wordt geïllustreerd.

Om de prestaties van IRS-ondersteunde netwerken te onderzoeken, definieert deze dissertatie een reeks Kernprestatie-indicatoren (KPI's), zoals datasnelheid, energie-efficiëntie, vermogensefficiëntie, SINR (Signaal tot Interferentie plus Ruisverhouding), zendvermogensbudget en ontvangen vermogenssterkte. Deze KPI's dienen als meetwaarden om de prestaties van het netwerk te beoordelen en te optimaliseren op basis van het ontwerpen van een efficiënt resource allocatiebeleid. Niet-lineaire, niet-convexe en Gemengde Gehele Getallen Niet-lineaire Programmering (MINLP)-problemen ontstaan bij het aanpakken van het resource allocatie optimalisatieprobleem. Deze problemen zijn niet-deterministisch polynomiale tijd (NP)-hard vanwege de complexe relatie tussen variabelen en de systeembeperkingen. Gezien de complexiteit van deze optimalisatieproblemen worden verschillende strategieën gebruikt om hun oplossing te vereenvoudigen en te benaderen. Door de doelfunctie (de KPI's) en beperkingen die niet-convex zijn te ontspannen naar een beter hanteerbaar formaat, werden de problemen beheersbaarder. Deze ontspanningsbenadering omvatte vaak het transformeren van het optimalisatieprobleem naar zijn convexe equivalent of het gebruik van benaderingstechnieken om niet-convexe termen te lineariseren of convex te maken.

Algoritmen worden ontwikkeld die in staat zijn het hoofdprobleem op te lossen, hetzij globaal of suboptimaal, maar voldoende dicht bij het globale optimum. Deze oplossingen maken gebruik van optimalisatie-oplossers en computersimulaties, waarbij geavanceerde wiskundige hulpmiddelen en technieken zoals de grote-M-methode voor het lineariseren van producttermen met binaire variabelen en opeenvolgende convexe benadering (SCA) worden ingezet om convexe benaderingen van niet-convexe termen te verkrijgen. De iteratieve aard van deze oplossingen maakte stapsgewijze verfijning mogelijk, waardoor geleidelijk naar een optimale configuratie van een resource allocatieontwerp werd bewogen ondanks de complexiteit van het initiële probleem.

Door uitputtende simulaties onthult deze dissertatie de diverse prestatieverbeteringen die haalbaar zijn door resource allocatie in IRS-ondersteunde netwerken, waarbij rijke inzichten worden geboden in hoe IRS-technologie draadloze communicatiesystemen kan verbeteren. Deze simulaties fungeren als een cruciale brug, die theoretische voorspellingen verbindt met empirisch bewijs en de praktische haalbaarheid van het voorgestelde IRS-versterkte netwerk valideert. Door verschillende IRS-configuraties te verkennen - zowel passieve als actieve typen onderzoeken en het aantal reflecterende elementen variëren - en hun implementatie in verschillende omgevingen en instellingen, bevestigt deze studie niet alleen de nauwkeurigheid van de theoretische modellen, maar verklaart ook de voorwaarden waaronder IRS-implementaties maximale prestatieverbeteringen opleveren, waarbij de veelzijdigheid van IRS in het aanpassen van nieuwe technologieën wordt getoond.

Deze dissertatie bestudeert de impact van IRS op een breed scala aan technologieën. Door de prestaties van SWIPT-netwerken te verbeteren, URLLC en MTC te faciliteren, MEC mogelijk te maken en VR, mmWave en THz-toepassingen te revolutioneren, staat IRS aan de voorhoede van draadloze communicatie-innovatie. Dit werk toont de diverse toepassingen van IRS-technologie en legt de basis voor toekomstig onderzoek gericht op het gebruik van IRS om de dynamische uitdagingen van moderne draadloze netwerken aan te pakken. Het wijst de weg naar de ontwikkeling van robuustere, efficiëntere en boeiendere communicatie-ecosystemen.

Acknowledgements

MY adventure began a decade ago, with a dream in my heart and a determination to dive into cutting-edge research — a feat that seemed out of reach in my homeland. So, I packed my bags for an exciting chapter at Gent University in Belgium, aiming for a Master's degree. Picture this: juggling a full-time job while wrestling with my Master's coursework. Crazy, right? But I thrived on the madness. After crossing the graduation stage, I took a leap of faith — I waved goodbye to my full-time job, ready to chase the dream that initially brought me to Belgium. Embarking on a PhD was my true calling.

I owe a huge debt of gratitude to many people who turned my dream into reality, especially Prof. Dr. Maarten Weyn and Prof. Dr. Rafael Berkvens. Imagine a charismatic scientist with a long beard, who's also an entrepreneur; that's Maarten for you. He did something pretty unexpected. He hired me, and then, believe it or not, he let me go! Sounds wild, right? But it was actually the best thing ever. I once showed him a paper I wrote in my spare time, completely off-topic from my research. He encouraged me to chase my curiosity, to research what I loved, not just what I was paid to do. That's how I got introduced to Prof. Dr. Jeroen Famaey by Maarten to change a new PhD topic. Huge thanks to Maarten. You're one of a kind!

Raf was there from the start, guiding me through thick and thin, while Jeroen became my main supervisor for the thrilling three-year that followed. Both of them are simply incredible: understanding, easy to collaborate with, and just a blast to work alongside. Despite their packed schedules, they never once turned down my requests, whether it was for a meeting, a letter, or even something offbeat and not strictly research-related. They've supported me in every way, professionally and personally.

A special thank you to Jeroen for going above and beyond. He penned countless letters of recommendation, supported my applications for grants — travel, student, conference, you name it — and even an FWO (V478223N), which I snagged with a jaw-dropping 100% success rate. Thanks to him, my success rate is off the charts! Guess it's hard to top that, huh?

Working alongside so many brilliant scientists has been a highlight of my journey, and Prof. Dr. Hina Tabassum from York University deserves a special shout-out. In my last year of PhD, she opened the doors to a research visit that enriched my experience immensely. Despite her hectic schedule, Prof. Tabassum generously made time for our meetings, even on Saturdays. Her dedication and support have been incredible. A huge thank you to her for going the extra mile!

Ata Khalili deserves a massive thanks as well. There's a saying that if you've been friends with someone for more than 7 years, you become family. Well, we've crossed that mark and

then some. Together, we've worked, met, and traveled, weaving countless memories along the way. Ata wasn't just a co-author on papers; he's a true friend-author. His collaboration was instrumental in publishing so many high-quality works. If it weren't for him, those achievements would have been out of reach. Here's to you, Ata, for being more than a friend and an integral part of this journey.

Immense gratitude to all my co-authors for making this journey so rewarding; Dr. Filip Lemic, for introducing me to the fascinating world of virtual reality, and Prof. Dr. Walid Saad, whose collaboration turned the pages of my research into chapters of groundbreaking work. Warm thanks to Dr. Jakob Struye, Prof. Dr. Xavier Costa Perez, and other co-authors for their crucial contributions. The teams at IDLab have been foundational, offering support and insights that turned challenges into opportunities and successes into moments of shared joy. Special mentions to everyone there, though I can't (possibly remember and) name you all, your friendship extended beyond work. Deep appreciation goes to my doctoral committee members — Prof. Dr. Maarten Weyn, Prof. Dr. Erik Mannens, Dr. Sergi Abadal, Prof. Dr. David Plets, and Dr. Gerhard Schoenthal — for their essential feedback and guidance. Each of you has significantly shaped this journey, for which I am eternally grateful.

This book exists because of the incredible support from my friends and family. They might not have always understood exactly what I was doing, and honestly, sometimes I didn't either, but their encouragement meant everything. A big thank you to my friends for being the perfect distraction when I needed it the most. And to my parents, John (Nouri) and Anna (Afsaneh), for everything you've given me to get here. I'm forever grateful. I regret not letting you be at my bachelor's and master's graduations (I didn't attend myself either); I was too focused on my goals and thought those weren't my real graduation! But I truly wished for you to be there at my PhD graduation, to celebrate the biggest milestone of them all together.

Finally, to my incredible wife, Grace, you are something else. Simply, amazing! I can hardly believe how lucky I am to have you. I can't quite put into words how much you mean to me. Thank you for your patience and your trust. Thank you for letting me chase my dream. Thank you for giving me two beautiful daughters and a family. Thank you for being my only one. Thank you for everything.

*Joshua Jalali
June 2024, Menen, Belgium*

Contents

Abstract	i
Samenvatting	iii
Acknowledgements	v
Contents	vii
List of Publications	xiii
List of Figures	xvii
List of Tables	xxi
List of Acronyms	xxiii
1 Introduction	1
1.1 Motivation	2
1.2 Contributions	4
1.3 Outline	7
2 IRS-Aided Wireless Communications: A Brief Overview	9
2.1 Intelligent Reflecting Surfaces	11
2.2 Architecture of an IRS: From Concept to Implementation	11
2.3 IRS vs. Other Technologies: Key Advantages and Differences	13
2.4 Holographic Massive MIMO Surfaces	14
2.5 Optimization of IRSs	15

2.6	Applications of IRSs	16
2.6.1	Practical Scenarios: From Blockage Mitigation to Advance Usage .	16
2.6.2	Indoor Environments: A Factory Setting	19
2.7	IRS-aided System Design: Dissertation Objectives	20
3	Power Efficient Multi-User SWIPT Networks	23
3.1	Introduction	25
3.2	Generalized AS versus other SWIPT Architectures	29
3.3	System Model and Problem Formulation	32
3.4	A Two-layer Optimal Solution Design	37
3.4.1	Antenna Selection	37
3.4.2	Beamforming Design	38
3.5	Simulation Results	43
3.6	Conclusion	48
4	Energy Efficient IRS-aided Resource Allocation	51
4.1	Introduction	53
4.2	System Model of an IRS-assisted URLLC Network	56
4.3	Resource Allocation Problem Formulation	59
4.4	Proposed Solution	60
4.4.1	First-stage: Optimizing \mathbf{w}_k and ϵ_k with Fixed Θ	61
4.4.2	Second-stage: Optimizing Θ with Fixed \mathbf{w}_k and ϵ_k	66
4.5	Numerical Results	69
4.6	Conclusion	73
5	Energy Efficient Admission Control in IRS Networks	75
5.1	Introduction	77
5.2	System Model of an MTC-enabled IRS-assisted network	81
5.3	Practical Overhead and CSI Estimation	84

5.3.1	Channel Estimation for an IRS-Assisted IoT Users with Finite Block Length	87
5.3.2	Discrete IRS Phase Shifts	89
5.4	Multi-Objective Problem Formulation of EE and Admission Control	90
5.5	Proposed AO Solution	92
5.5.1	First-stage: Optimizing \mathbf{w}_k with Fixed Θ	93
5.5.2	Second-stage: Optimizing Θ	101
5.6	Computational Complexity and Convergence Analysis	104
5.7	Numerical Results for the MTC-enabled IRS-aided Network	105
5.7.1	Tradeoff between EE and the Average Number of Admitted Users .	107
5.7.2	User Admissibility vs. User Serviceability	109
5.8	Insights and Practical Implications of MTC-enabled services for IRS-aided Networks	112
5.8.1	Ultra-Reliable Low-Latency Communication (URLLC) in IRS-aided Networks	113
5.9	URLLC-enabled System Model and Problem Formulation	114
5.10	Proposed Solution: A Novel Rank Relaxation Method	117
5.10.1	Step One: Optimizing $\Omega_{k,l}$ with Fixed Ψ_l	120
5.10.2	Step Two: Optimizing Ψ_l	124
5.11	Numerical Results for the URLLC-enabled IRS-assisted Network	127
5.12	Conclusion	130
6	Resource Allocation in Uplink MEC-IRS Networks	133
6.1	Introduction	135
6.2	System Model for a MEC-enabled Uplink IRS Network	138
6.2.1	Signal and Channel Models	138
6.2.2	Achievable Rate with Short Packet Transmission	140
6.2.3	Offloading decision	141
6.3	Problem Formulation	142
6.4	Solution of the Optimization Problem	143

6.5	Performance Evaluation	150
6.5.1	Performance Bound and Benchmark Schemes	151
6.5.2	Simulation Results	152
6.6	Conclusions	155
7	Active IRS for mmWave Wireless Networks	157
7.1	Introduction to IRS-assisted mmWave Networks	159
7.2	System Model of an IRS-assisted mmWave Network	162
7.2.1	A Simplified mmWave Channel Model	164
7.3	Sum Data Rate Problem Formulation for the IRS-assisted mmWave Network	165
7.3.1	Optimization with respect to the Transmit Beamforming	167
7.3.2	Optimization with respect to the Active IRS Parameters	170
7.4	Simulation Results for the IRS-assisted mmWave Network	172
7.5	Insights and Practical Applications	175
7.5.1	Practical Use Case: IRS-aided mmWave Network for Virtual Reality	175
7.6	Location Optimization and Resource Allocation of IRS in a Multi-User Indoor mmWave VR Network	176
7.7	VR IRS-assisted System Model and Problem Formulation	178
7.8	Location Optimization and Resource Allocation of IRS in a VR Network	182
7.8.1	Step 1: AP Active (transmitter-side) Beamforming	183
7.8.2	Step 2: IRS Passive (receiver-side) Beamforming	184
7.8.3	Step 3: IRS Placement at Optimal Locations	185
7.8.4	Step 4: IRS Radiation Pattern Optimization	186
7.9	Analyzing Resource Allocation Complexity in IRS-aided VR Networks	187
7.10	Evaluation Setup and Simulations Results for the IRS-assisted VR Use Case	188
7.11	Conclusion	192
8	Energy Efficient THz Miniature UAV Networks	193
8.1	Introduction	195

8.2	System Model and Performance Metric	199
8.2.1	Phase One ⁽¹⁾ : Direct Transmission	200
8.2.2	Phase Two ⁽²⁾ : Cooperative Transmission	202
8.2.3	Energy Efficiency Metric in a Cooperative Network	204
8.3	EE Maximization Problem Formulation	205
8.4	A Two-Stage Solution to EE Problem	206
8.4.1	Stage-one: Optimizing PS ratio and Miniature UAV trajectory . . .	207
8.4.2	Stage-two: Optimizing power coefficients	213
8.5	Complexity Analysis	216
8.6	Simulation Results and Discussions	217
8.7	IRS-based UAV with Underlaid D2D Users in THz Networks	222
8.8	Conclusion	225
9	Conclusion	227
9.1	Main findings	229
9.2	Future Research Directions	231
A	Optimization Techniques	237
A.1	Convex Analysis	237
A.1.1	Definitions	238
A.2	Duality Theorem	238
A.2.1	Weak Duality and Duality Gap	239
A.2.2	Strong Duality and Slater condition	240
A.3	Lagrangian	240
A.3.1	Augmented Lagrangian Method	241
A.4	Complementary Slackness and KKT Optimality Conditions	243
A.5	Interior-Point Methods	244
A.6	MM Approach and DC Programming	246
A.7	Optimization Packages	248

Bibliography	251
---------------------	------------

List of Publications

Journal Articles [1, 2, 3, 4, 5, 6]

1. S. Bayat, **J. Jalali**, A. Khalili, M. R. Mili, S. Wittevrongel, and H. Steendam, "Optimal Multi-Objective Resource Allocation for D2D Underlaying Cellular Networks in Uplink Communications", *IEEE Access*, vol. 9, p. 114 153–114 166, 2021.
2. A. Rezaei, A. Khalili, **J. Jalali**, H. Shafiei, and Q. Wu, "Energy-Efficient Resource Allocation and Antenna Selection for IRS-Assisted Multicell Downlink Networks", *IEEE Wireless Communications Letters*, vol. 11, no. 6, pp. 1229-1233, June 2022
3. **J. Jalali**, A. Khalili, A. Rezaei, R. Berkvens, M. Weyn and J. Famaey, "IRS-Based Energy Efficiency and Admission Control Maximization for IoT Users With Short Packet Lengths", *IEEE Transactions on Vehicular Technology*, vol. 72, no. 9, pp. 12 379–12 384, Sept. 2023
4. **J. Jalali**, A. Khalili, H. Tabassum, R. Berkvens, J. Famaey and W. Saad, "Energy-Efficient THz NOMA for SWIPT-aided Miniature UAV Networks", *IEEE Communications Letters*, vol. 28, no. 5, pp. 1107-1111, May 2024.
5. **J. Jalali**, A. Khalili, H. Tabassum, J. Famaey, W. Saad, and Murat Uysal "Placement, Orientation, and Resource Allocation Optimization for Cell-Free OIRS-aided OWC Network (*Submitted*)", *IEEE Transactions on Vehicular Technology*, pp. 1–5, 2024.
6. **J. Jalali**, F. Lemic, H. Tabassum, and J. Famaey, "Is Geometric Polyhedron Approach Viable for Optimal Multi-IRS Placement in Upcoming 6G Networks? (*Preprint and under Submission*)", *IEEE Transactions on Vehicular Technology*, pp. 1–5, 2024.

Conference Proceedings [7, 8, 9, 10, 11, 12, 13]

1. **J. Jalali**, A. Kaya, M. Weyn, and R. Berkvens, "Activity Monitoring at an Intersection Using a Sub-GHz Wireless Sensor Network", in *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*, Norman, OK, USA, Sept. 2021, pp. 1–6.
2. **J. Jalali**, A. Rezaei, A. Khalili and J. Famaey, "Power-efficient Joint Resource Allocation and Decoding Error Probability for Multiuser Downlink MISO with Finite Block Length Codes", in *2022 25th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, Herning, Denmark, Oct. 2022, pp. 232–237.

3. **J. Jalali**, A. Khalili, A. Rezaei, J. Famaey and W. Saad, "Power-efficient Antenna Switching and Beamforming Design for Multi-User SWIPT with Non-Linear Energy Harvesting", in *2023 IEEE 20th Consumer Communications & Networking Conference (CCNC)*, Las Vegas, NV, USA, Jan. 2023, pp. 746–751.
4. **J. Jalali**, A. Khalili, A. Rezaei and J. Famaey, "Is Active IRS Useful for mmWave Wireless Networks or Not?", in *2023 International Conference on Computing, Networking and Communications (ICNC)*, Honolulu, HI, USA, Feb. 2023, pp. 377–382.
5. **J. Jalali**, A. Khalili, R. Berkvens and J. Famaey, "Joint Offloading Policy and Resource Allocation in IRS-aided MEC for IoT Users with Short Packet Transmission", in *2023 IEEE 98th Vehicular Technology Conference (VTC2023-Fall)*, Hong Kong, Hong Kong, Oct. 2023, pp. 1–7.
6. **J. Jalali**, F. Lemic, H. Tabassum, R. Berkvens, and J. Famaey, "Toward Energy Efficient Multiuser IRS-Assisted URLLC Systems: A Novel Rank Relaxation Method", in *GLOBECOM 2023 - 2023 IEEE Global Communications Conference - 6G Communication Workshop*, Kuala Lumpur, Malaysia, Dec. 2023, pp. 1–7.
7. **J. Jalali**, M. Bustamante, F. Lemic, H. Tabassum, J. Struye, J. Famaey, and X. Costa Pérez, "Location Optimization and Resource Allocation of IRS in a Multi-User Indoor mmWave VR Network", in *2024 IEEE Wireless Communications and Networking Conference (WCNC)*, Dubai, United Arab Emirates, Apr. 2024, pp. 1–7.

Book Chapter [14]

1. F. Lemic, **J. Jalali**, G. Calvo Bartra, A. Amat, J. Struye, J. Famaey, and X. Costa Perez, "Location-based Real-time Utilization of Reconfigurable Intelligent Surfaces for mmWave Communication and Sensing in Full-Immersive Multiuser VR (*Accepted for Publication*)", *Advanced Metaverse Wireless Communication Systems*, pp. 1–31, 2024.

List of Algorithms

1	Power Efficient Resource Allocation Algorithm for Beamforming in Multi-User SWIPT Networks	41
2	Iterative SCA Algorithm for Energy-efficient Resource Allocation Policy in IRS-aided URLLC Networks	66
3	Iterative Alternating Optimization (AO) algorithm for Energy-efficient Resource Allocation Policy in IRS-aided URLLC Networks	69
4	Proposed Algorithm for IRS-Based Energy Efficiency and Admission Control Maximization for Internet of Things (IoT) Users With Short Packet Lengths	103
5	Iterative SCA-based Resource Allocation Algorithm for Energy Efficient Multi-user IRS-Assisted URLLC Systems Based on a Novel Rank Relaxation Method	125
6	Proposed Iterative AO algorithm Toward Energy Efficient Multi-user IRS-Assisted URLLC Systems Based on a Novel Rank Relaxation Method	127
7	Proposed Iterative SCA Algorithm for Offloading and Resource Allocation in an Uplink IRS-assisted MEC Network	150
8	Iterative AO algorithm for a Multi-User Indoor mmWave VR Network	188
9	Iterative Resource Allocation Algorithm for Energy Efficiency (EE) maximization of THz-Non-Orthogonal Multiple Access (NOMA) SWIPT-aided Miniature Unmanned Aerial Vehicle (UAV) Networks	217
10	Interior-Point Method Algorithm with Barrier Function for LP	246
11	The Majorization Minimization (MM) (Majorization Minimization or Minorization Maximization) algorithm	248

List of Figures

1.1	Thesis outline	7
2.1	Architecture of an IRS.	12
2.2	Illustration of IRS applications in future wireless network.	17
2.3	Typical IRS applications in wireless network.	18
2.4	Application scenario for an IRS to mitigate blockage in an indoor factory.	19
3.1	Integrated receiver architecture designs for SWIPT.	30
3.2	Generalized Antenna-Switching (AS) approach to realize SWIPT architecture in a single-cell multi-user Multiple Input Multiple Output (MIMO) network.	32
3.3	Power efficiency versus maximum allowed transmit power in the downlink of SWIPT-aided multi-user single-cell network.	44
3.4	System performance tradeoff between power efficiency and EE for $p_{\max} = 40$ deciBel-milliwatts (dBm) in the downlink of SWIPT-aided multi-user single-cell network.	46
3.5	EE versus maximum allowed transmit power in the downlink of SWIPT-aided multi-user single-cell network.	47
4.1	Energy-Efficient Resource Allocation for a multi-user DownLink (DL) Multiple-Input Single-Output (MISO) URLLC-enabled IRS system.	57
4.2	Impact of decoding error, ϵ_k , on the average transmit power for a downlink of multi-user IRS-aided URLLC network.	71
4.3	Average EE vs. the number of transmit antennas and reflecting elements for a downlink of multi-user IRS-aided URLLC network.	72
5.1	Energy efficiency and admission control of IoT users in an MTC-enabled IRS-assisted network with finite blocklength transmission.	81
5.2	Channel estimation for an IRS-Assisted IoT users with finite block length.	88

5.3	EE and the average number of admitted users vs. p_{\max}	107
5.4	EE vs. the average number of admitted users for IoT users with short packet lengths in an IRS-assisted MTC network.	108
5.5	Jain's fairness index versus the total number of users in for IoT users with short packet lengths in an IRS-assisted MTC network.	111
5.6	Illustration of a multi-user IRS-assisted MISO downlink network comprising one Base Station (BS) and K URLLC users employing finite block length transmissions. \mathbf{h} is the channel matrix between the BS and the IRS, $\mathbf{h}_{\text{IRS},1}$ denotes the channel response vector from the IRS to user 1, and $\mathbf{h}_{\text{BS},1}$ indicates the channel between the BS and user 1.	115
5.7	Impact of tolerable decoding error, ϵ_k , on the EE in an URLLC-enabled IRS-assisted network.	128
5.8	Impact of the blocklength, m_d , on the EE in an URLLC-enabled IRS-assisted network.	129
6.1	Multi-user IRS-assisted URLLC MEC system with one Access Point (AP), and K IoT users with finite block-length transmission. The single-antenna MTC-enabled IoT users offload their tasks in the uplink to an MEC, directly or via IRS, using a single multi-antenna AP with an edge server.	139
6.2	Average power consumption [dBm] vs. packet decoding error probability in an IRS-aided network.	153
6.3	Average power consumption [dBm] vs. the task size [bits] in the uplink of an IRS-assisted Orthogonal Frequency Division Multiple Access (OFDMA) URLLC MEC network.	154
7.1	An IRS-assisted multi-user MISO mmWave wireless communication system. The AP is equipped with N_t -antennae and serves K single-antenna mobile users in the downlink. The active IRS has M reflecting elements in the AP's LoS signals.	162
7.2	Average sum data rate versus the number of reflecting elements.	173
7.3	Average sum data rate versus the number of iterations.	174
7.4	A multi-user IRS-assisted full-immersive VR scenario. The IRS is deployed on one of the walls, and a multi-antenna AP transmits data indirectly to a set of single-antenna Head-Mounted Device (HMD) via the IRS.	179
7.5	SNR variability enhancements due to the utilization of IRS resources at locations optimized by the proposed AO approach.	191

8.1	Miniature UAV-assisted cooperative THz NOMA-SWIPT network. The channel power gain between source-to-UAV and UAV-to-destination is denoted as $h_{sr}(t)$ and $h_{rd}(t)$, respectively. The miniature UAV acts as an energy-harvesting user in the first phase. In the second phase, the miniature UAV is a relay that uses the harvested energy from the previous phase.	199
8.2	Phase One: Miniature UAV-assisted cooperative THz NOMA-SWIPT network. The gray link means an inactive link. The source transmits the information to both the miniature UAV and destination node by exploiting power-domain NOMA. The miniature UAV has not only its own data in this phase but also the destination node's data. The miniature UAV decodes the data intended for the destination node but relays it in the next phase.	201
8.3	Phase Two: Miniature UAV-assisted cooperative THz NOMA-SWIPT network. The gray links mean inactive links. The miniature UAV relays the data intended to the destination node in this phase, using the harvested energy from the previous phase.	203
8.4	The effect of average network transmit power, \bar{p}_{sum} , on EE of THz-NOMA SWIPT-aided miniature UAV networks.	219
8.5	The trajectory of miniature UAV in the cooperative THz NOMA-SWIPT network.	220
8.6	The EE vs. the miniature UAV's operational time in the THz-enabled SWIPT network.	221
8.7	The EE vs. molecular absorption in the THz-enabled miniature UAV network.	222
8.8	Model of the UAV with underlaid device-to-device (Device-to-Device (D2D)) communications network. The system model of an uplink single macro cell OFDMA-based cellular network with one base station to serve M Cellular Userss (CUs) and K D2D Users (DUs). In this figure, the green arrow shows the cellular transmission link between the base station and the CUs, while the dotted line indicates the D2D links.	223
9.1	Illustration of the Optical IRS (OIRS)-supported Optical Wireless Communication (OWC) network, where the l -th Light-Emitting Diodes (LED) and its reflection are symmetrically positioned across the $x'y'z'$ plane. We consider the downlink of an OIRS-aided cell-free OWC system, where L LEDs serve K Photo-Diode (PD) users (or photodetectors), with an OIRS with N units enhancing communication.	232
9.2	Envisioning the Future of Connectivity: A cityscape where skyscrapers are embedded with Intelligent Reflecting Surfaces (IRS), enhancing the coverage and the boundaries of next-generation wireless networks. — Generated by OpenArtificial Intelligence (AI).	234

9.3 Illustration of a polyhedron \mathcal{P} as the intersection of five half-spaces, with outward normal vectors $\mathbf{a}_1, \dots, \mathbf{a}_5$. The red segments are the possible locations of IRSs in each half-space. 236

List of Tables

3.1	Summary of Our Main Notations in Power-efficiency Optimization of a Multi-user SWIPT Network.	33
3.2	Overview of Simulation Parameters for a Multi-Antenna, Multi-User SWIPT Network.	43
4.1	Simulation Parameters for Multi-user MISO URLLC-enabled IRS System.	70
5.1	Simulation Parameters for Multi-user MTC-enabled IRS Systems.	106
5.2	Simulation Parameters for Multi-user IRS-Assisted URLLC Systems.	127
6.1	Simulation Parameters for Offloading and Resource Allocation in an Uplink IRS-assisted MEC Network	152
7.1	Overview of Baseline Simulation Parameters for the IRS-assisted mmWave VR Network.	189
7.2	Summary of Achieved Results: Comparison of Different Approaches Based on Room Sizes, Average (Avg) Throughput, and Standard Deviation (Standard Deviation (SD)).	190
8.1	Simulation Parameters for EE maximization of THz-NOMA SWIPT-aided Miniature UAV Networks.	218

Acronyms

2D	two-Dimensional
3D	three-Dimensional
3GPP	3rd Generation Partnership Project
5G	fifth-Generation
6G	sixth-Generation
A-MPDU	Aggregated MAC Protocol Data Unit
A-MSDU	Aggregated MAC Service Data Unit
AF	Amplify-and-Forward
AI	Artificial Intelligence
ALM	Augmented Lagrangian Method
AO	Alternating Optimization
AP	Access Point
AR	Augmented Reality
AS	Antenna-Switching
AWGN	Additive White Gaussian Noise
B5G	Beyond 5G
BS	Base Station
CDF	Cumulative Distribution Function
cfm-MIMO	cell-free massive Multi-Input Multi-Output
COIN-OR	Common Optimization INterface for Operations Research
CPU	Central Processing Unit
CSCG	Circularly Symmetric Complex Gaussian
CSI	Channel State Information
CSMA/CA	Carrier Sense Multiple Access with Collision Avoidance
CU	Cellular Users
d-QoS	delay-Quality of Service
D2D	Device-to-Device
dB	deciBel
dBi	deciBel-isotropic
dBm	deciBel-milliwatts
DC	Difference of Convex
DEP	Decoding Error Probability
DF	Decode-and-Forward

DL	DownLink
DMG	Directional Multi-Gigabit
DoF	Degree of Freedom
DRL	Deep Reinforcement Learning
DU	D2D User
EB	EexaBytes
EE	Energy Efficiency
EEI	Energy Efficiency Indicator
EH	Energy Harvesting
EM	ElectroMagnetic
eMBB	enhanced Mobile BroadBand
FD	Full-Duplex
FDD	Frequency Division Duplex
FET	Field-Effect Transistor
FPGA	Field-Programmable Gate Array
FR1	Frequency Range 1
FR2	Frequency Range 2
FR3	Frequency Range 3
FRX	Frequency Range X
Gbps	Gigabits per second
GHz	GigaHertz
GLPK	GNU Linear Programming Kit
gNB	5G Node B
HMD	Head-Mounted Device
HMIMOS	Holographic Massive MIMO Surfaces
Hz	Hertz
ID	Information Decoding
IoE	Internet of Everything
IoT	Internet of Things
IoV	Internet of Vehicles
IRS	Intelligent Reflecting Surface
kHz	kiloHertz
KKT	Karush–Kuhn–Tucker
KPI	Key Performance Indicator
LED	Light-Emitting Diodes
LoS	Line-of-Sight
LP	Linear Programming
m-MIMO	massive Multiple Input Multiple Output
MAC	Medium Access Control
Mbps	Megabits per second

MCS	Modulation and Coding Scheme
MEC	Mobile Edge Computing
MEMS	MicroElectroMechanical System
MHz	MegaHertz
MILP	Mixed Integer Linear Programming
MIMO	Multiple Input Multiple Output
MINLP	Mixed Integer Nonlinear Programming
MIP	Mixed Integer Programming
MIQCP	Mixed Integer Quadratically Constrained Programming
MIQP	Mixed Integer Quadratic Programming
MISO	Multiple-Input Single-Output
mJoule	million Joule
ML	Machine Learning
MM	Majorization Minimization
MMSE	Minimum Mean Square Error
mMTC	massive Machine Type Communication
mmWave	millimeter-Wave
MOOP	Multi-Objective Optimization Problem
MRT	Maximum Ratio Transmission
MSE	Mean Square Error
MTC	Machine Type Communication
mW	milli Watt
NLoS	Non-Line-of-Sight
NOMA	Non-Orthogonal Multiple Access
NP	Non-deterministic Polynomial time
ns-3	discrete-event Network Simulator version-3
NTN	Non-Terrestrial Network
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OMA	Orthogonal Multiple Access
OSQP	Operator Splitting Quadratic Program
OWC	Optical Wireless Communication
P2P	Point-to-Point
PD	Photo-Diode
PHY	Physical
PIN	Positive-Intrinsic-Negative
PRU	Passive Reflection Unit
PS	Power-Splitting
PSD	Positive Semi-Definite
QAM	Quadrature Amplitude Modulation
QCQP	Quadratic Constrained Quadratic Programming
QKD	Quantum Key Distribution
QoS	Quality of Service
QP	Quadratic Programming

RF	Radio Frequency
RFID	Radio Frequency IDentification
RIS	Reconfigurable Intelligent Surfaces
RRH	Remote Radio Head
RSRP	Reference Signal Received Power
SC	Single Carrier
SCA	Successive Convex Approximation
SD	Standard Deviation
SDP	Semi-Definite Programming
SE	Spectral Efficiency
Sec	Second
SIC	Self-Interference Cancellation
SINR	Signal-to-Interference-plus-Noise Ratio
SISO	Single-Input Single-Output
SNR	Signal-to-Noise Ratio
SOCP	Second-Order Cone Programming
SOOP	Single-Objective Optimization Problem
SWIPT	Simultaneous Wireless Information and Power Transfer
Tbps	Terabits per second
TDD	Time Division Duplexing
TDMA	Time Division Multiple Access
THz	TeraHertz
TS	Time-Switching
UAV	Unmanned Aerial Vehicle
UDP	User Datagram Protocol
UE	User Equipment
UL	UpLink
ULA	Uniform Linear Array
uMTC	ultra-reliable MTC
UPA	Uniform Planar Array
URLLC	Ultra-Reliable Low-Latency Communication
VE	Virtual Experience
VLC	Visible Light Communication
VR	Virtual Reality
W	Watt
Wi-Fi	Wireless Fidelity
WiGig	60 GHz Wi-Fi
WMMSE	Weighted Minimum Mean Square Error
WPT	Wireless Power Transfer
WSMSE	Weighted Sum Mean Square Error

Chapter 1

Introduction

THE surge in innovative technologies like Artificial Intelligence (AI), Virtual Reality (VR), three-Dimensional (3D) media, and the Internet of Everything (IoE) has significantly increased data traffic worldwide. In the second quarter of 2023, global mobile data traffic was recorded at 133.86 ExaBytes (EB) per month, with projections suggesting a rise to 5016 EB per month by 2030 [15]. These figures suggest the critical need for advancements in communication technology as we move towards a future dominated by fully automated remote systems. Autonomous technologies are gaining ground across various sectors, such as manufacturing, healthcare, transportation, maritime, and space exploration. This trend is supported by the widespread integration of millions of sensors in urban areas, vehicles, homes, and industries, facilitating a smarter, automated lifestyle. As a result, there is an imminent need for communication networks that can handle high data rates and provide reliable connectivity to accommodate these evolving applications.

Despite their substantial advancements over current technologies, the fifth-Generation (5G) wireless networks are anticipated to fall short in delivering a fully automated, intelligent network capable of offering everything as a service and providing a fully immersive experience [16]. While 5G systems have marked a significant upgrade, they are not expected to meet the needs of the next wave of intelligent and automated systems in a few years' time [17]. 5G introduced numerous enhancements, including access to new frequency bands like millimeter-wave (mmWave) and optical spectra, improved spectrum usage and management, and the combination of licensed and unlicensed bands [18, 19]. Yet, the rapid expansion of data-driven and automated systems might surpass what 5G networks can handle. One of the critical areas where 5G may lag is in its integration of communication, intelligence, sensing, control, and computing capabilities, a fusion essential for future IoE applications. For instance, devices like VR headsets require Beyond 5G (B5G) capabilities, needing data rates of at least 10 Gigabits per second (Gbps) [20]. Therefore, as 5G approaches its capacity limits by 2030, discussions and research are already underway to outline the objectives for the subsequent generation of wireless communication technologies [21].

Items poised to necessitate the capabilities of a sixth-Generation (6G) system include (i) expansive interfaces for human and machine interaction, (ii) pervasive computing that bridges local devices with cloud services, (iii) the integration of sensory data to construct comprehensive multi-reality environments, and (iv) enhanced precision in sensing and communication for control over physical environments [22]. 6G networks aim to address the shortcomings of

5G by weaving together futuristic services, including ambient sensing intelligence and novel forms of interaction between humans and machines, as well as between humans themselves. This next step in network evolution will heavily incorporate AI and introduce cutting-edge technologies like terahertz (THz) communication, 3D networking, quantum communications, holographic beamforming, backscatter communications, Intelligent Reflecting Surface (IRS), and proactive caching [23]. The driving forces behind 6G involve a fusion of previous network advancements — network densification, superior throughput, utmost reliability, minimal energy consumption, and extensive connectivity — while also pushing forward with innovative services and technologies such as AI, smart wearables, implants, autonomous vehicles, augmented reality devices, sensing technologies, and 3D mapping [24]. A fundamental expectation from 6G wireless networks is their ability to manage enormous data volumes and provide exceptionally high data rates [25].

The ambition for the 6G communication system is to serve as a comprehensive global communication infrastructure, offering per-user bit rates of around 1 Terabits per second (Tbps) in many scenarios. This represents a connectivity capacity 1000 times greater than that of 5G, alongside ultra-long-distance communication capabilities with sub-millisecond latency [26]. A standout feature of 6G will be its full integration of AI to support autonomous systems, with video-type traffic expected to dominate data transmission in 6G networks. Leading technologies propelling 6G forward will include the THz spectrum, AI, Optical Wireless Communication (OWC), 3D networking, UAVs, IRS, and wireless power transfer, setting the stage for an unprecedented era of connectivity and technological integration [27].

In the transition from 5G to 6G wireless systems, there is an anticipated move away from the traditional massive Multiple Input Multiple Output (m-MIMO) configurations towards the adoption of IRS [28]. IRS represents a novel hardware innovation to significantly enhance energy efficiency in communication systems, often referred to as green communication. Also known as meta-surfaces, IRS comprises numerous reflective diode units capable of altering the phase of incoming electromagnetic signals in a controllable manner ¹. This capability positions IRS as a pivotal evolution in communication technology, sometimes referred to as “m-MIMO 2.0” for 6G networks [35]. These surfaces are expected to incorporate index modulation techniques, further boosting spectral efficiency and representing a key advancement in the infrastructure of 6G wireless communications.

1.1 Motivation

The 6G wireless network aims to surpass the capabilities of 5G by targeting more ambitious goals such as ultra-high data rates, superior energy efficiency, comprehensive global connectivity, and unmatched reliability and low latency. Achieving these objectives may be beyond the reach of current technology trends designed for 5G services, such as enhanced mobile broadband (eMBB), ultra-reliable and low latency communication (URLLC), and massive Machine Type Communication (mMTC), due to several challenges [36, 37, 38]:

¹An IRS can be a meta-surface (generally deep subwavelength) or a conventional passive reflectarray. Also, we acknowledge IRS can use other mechanisms than diode units to provide reconfigurability, e.g., [29, 30, 31, 32, 33, 34].

- Increasing the number of active nodes like Base Stations (BSs), Access Points (APs), relays, and distributed antennas/Remote Radio Heads (RRHs) to reduce communication distance and enhance network coverage and capacity, which leads to higher energy consumption, and increased deployment, backhaul, and maintenance costs, along with more complex network interference issues.
- Adding significantly more antennas to BSs/APs/relays to leverage the benefits of m-MIMO technology, necessitating greater hardware investment and energy expenditure, along with increased signal processing complexity.
- Shifting to higher frequency bands, such as mmWave and THz frequencies, to exploit their vast available bandwidth. This shift requires the deployment of additional active nodes and the installation of even more antennas (i.e., super MIMO) to offset the greater propagation loss associated with these higher frequencies.

Given these limitations, it is crucial to explore radically new and innovative technologies to ensure future wireless networks can grow sustainably, maintaining low costs, complexity, and energy consumption. Conversely, a primary obstacle in realizing ultra-reliable wireless communication stems from the dynamic nature of wireless channels, which fluctuate due to user movement. Traditional strategies to address this variability involve compensating for channel fading through a range of modulation, coding, and diversity techniques, or adjusting to it with adaptive power/rate control and beamforming methods [39, 40]. Yet, these approaches introduce extra overhead and provide only limited management of the inherently unpredictable wireless channels. This leaves the critical challenge of achieving both high-capacity and ultra-reliable wireless communications unresolved.

In this dissertation, IRSs are highlighted as a crucial technology that catalyzes advancements in a range of high-tech sectors by boosting their operational performance and energy efficiency. By adjusting the reflection of signals through numerous low-cost passive elements, IRS dynamically enhances communication performance, marking a significant step towards energy-efficient green communication. Known as meta-surfaces, these devices can modify the phase of impinging electromagnetic signals, positioning IRS as a pivotal innovation in 6G networks.

A notable breakthrough for IRS-aided wireless systems, especially in single-user scenarios, is the ability to form a “signal hot spot” through a combination of active beamforming at the BS/AP and passive beamforming at the IRS [41]. It has been demonstrated that IRS can achieve a significant asymptotic increase in Reference Signal Received Power (RSRP) or Signal-to-Noise Ratio (SNR), scaling quadratically in the order of $\mathcal{O}(N^2)$, where N represents the number of IRS reflecting elements [42]. This gain surpasses the linear enhancement $\mathcal{O}(N)$ seen with m-MIMO systems, attributable to IRS’s dual functionality, effectively doubling the gain compared to m-MIMO under the same total transmit power. Moreover, unlike traditional MIMO relays which, even in an ideal Full-Duplex (FD) mode with perfect Self-Interference Cancellation (SIC), only achieve a linear SNR improvement with the increase in active antennas due to relay noise, IRS benefits from a full-duplex, noise-free reflection

mechanism, thus providing a greater SNR boost [43]. An analysis comparing the performance between an IRS-assisted Single-Input Single-Output (SISO) system and a m-MIMO setup reveals the superior efficiency of IRS in enhancing wireless communications [44].

Furthermore, in multi-user systems supported by IRS, it is shown that IRS not only boosts the signal power/SNR at the user's receiver but also establishes an almost "interference-free" zone around it. This is achieved by utilizing the IRS's capability to nullify spatial interference, enabling users close to the IRS to withstand more interference from the BS/AP than those outside the IRS coverage. This advantage allows for more versatile transmit precoding strategies at the BS/AP for users located outside the IRS's influence, thereby enhancing the overall signal-to-interference-plus-noise ratio (SINR) for all users in the network. This dissertation also explores the integration of active and passive beamforming in various system designs [45, 46, 47], including physical layer security, simultaneous wireless information and power transfer (SWIPT) [48, 49, 50, 51], and NOMA [52, 53], underscoring the transformative impact of IRS across multiple wireless communication paradigms [54].

This dissertation explores the potential of the IRS in revolutionizing next-generation wireless networks. Below, we summarize the key contributions of this work.

1.2 Contributions

In this section, we detail the key contributions of our research, emphasizing the notable progress made in understanding and implementing IRS in wireless communication networks. Drawing from the 3rd Generation Partnership Project (3GPP) technical reports (38.101 and 38.101, cf. [55, 56, 57]) which examine channel models for frequencies ranging from 0.5 to 100 GHz and categorize the frequency spectrum into two distinct ranges — below 7.125 GHz as Frequency Range 1 (FR1) and above it as Frequency Range 2 (FR2)² — the primary contributions of this dissertation are divided into two main areas:

Frequency Range 1 (FR1) Contributions (Chapter 3, 4, 5, 6)

1. Establishing the baseline with SWIPT network optimization. (Chapter 3)

This dissertation's initial contribution lays the groundwork by enhancing power efficiency in single-cell networks with multi-antenna and multi-user setups, specifically through SWIPT. By aiming to maximize energy harvesting while minimizing power consumption, we present an optimization framework that simplifies beamforming and antenna selection. This foundational work provides the theoretical basis and establishes a baseline for advancing energy efficiency and operational effectiveness in future wireless networks, paving the way for subsequent studies on the impact of IRS.

²Due to the lack of established channel models for THz frequencies in the standardization community, we provisionally place THz frequencies into FR2. It is possible that this could lead to the creation of a Frequency Range X (FRX) in the future, although a consensus on this classification has not yet been reached. While it may seem counterintuitive, Frequency Range 3 (FR3) is an unofficial term that refers to the spectrum between 7.125 and 24.25 GHz, which lies between FR1 and FR2 [58]. In this dissertation, we loosely classify THz frequencies within FR2, despite anticipating that a more appropriate frequency range will soon be assigned to such high frequencies.

2. Optimizing URLLC through IRS-enhanced beamforming. (Chapter 4)

The second key contribution of this work is the integration of IRS in multi-user MISO systems, aiming to improve URLLC services. This is achieved by minimizing the overall transmission power via the concurrent optimization of both active and passive beamforming, facilitated by a sophisticated Alternating Optimization (AO) algorithm. Our method sheds light on the complex interplay between active and passive beamforming, demonstrating the significant potential of IRS to advance URLLC systems. Empirical evidence from simulation studies supports the effectiveness of our proposed solutions, representing an essential advancement towards fully leveraging IRS in the evolution of wireless communications.

3. Enhancing MTC in IoT networks with IRS for better energy efficiency and serviceability. (Chapter 5)

This study investigates the incorporation of Machine-Type Communication (MTC) within a multi-user MISO setup. Our focus was on enhancing the network's overall energy efficiency and improving the capacity to support an extensive range of IoT users by finely tuning both active and passive beamforming strategies. The adoption of a new AO algorithm served to demonstrate the effect of IRS on network performance, particularly in terms of energy savings and the ability to cater to a broader IoT user base.

4. Improving edge computing efficiency in multi-user MTC networks with IRS integration. (Chapter 6)

This portion of the dissertation investigates the integration of Mobile Edge Computing (MEC) with MTC in settings that involve multiple users. It emphasizes the crucial role played by IRS in boosting computational offloading, thereby enhancing both latency and reliability for MTC devices. Through the strategic optimization of radio resource allocation and decisions regarding edge offloading in networks supported by IRS, a groundbreaking method to increase the efficacy of edge computing in MTC scenarios is introduced.

Frequency Range 2 (FR2) Contributions (Chapter 7, 8)

6. Exploring FR2 with active and passive IRS in mmWave networks (Chapter 7)

This contribution is an exploration of FR2, diving into the domain of mmWave wireless networks. By introducing an active IRS-enhanced MISO system operating at mmWave frequencies, we tackled the optimization of the system's sum rate. This involved leveraging the unique benefits of active IRS configurations. Through a comprehensive optimization framework and the introduction of two algorithms, we highlighted active IRS's capabilities to boost network performance and established a new benchmark for applying IRS technology in mmWave spectra, paving the way for novel research and development avenues.

6. Exploring FR2 in THz miniature UAV networks (Chapter 8)

For the last contribution of this dissertation, we investigate the cutting-edge domain of THz band communication. A framework is developed to optimize energy efficiency through the strategic deployment of a miniature UAV trajectory alongside a refined

network resource allocation strategy. Our investigation focuses on the pivotal influences of miniature UAV mobility, NOMA power allocations, and SWIPT power-splitting ratios on system-wide performance.

The above contributions are described in detail in the following publications:

1. **J. Jalali**, A. Khalili, A. Rezaei, J. Famaey and W. Saad, "Power-efficient Antenna Switching and Beamforming Design for Multi-User SWIPT with Non-Linear Energy Harvesting", in *2023 IEEE 20th Consumer Communications & Networking Conference (CCNC)*, Las Vegas, NV, USA, Jan. 2023, pp. 746–751. (**IEEE Student Travel Grant**) (*Chapter 3*)
2. **J. Jalali**, A. Rezaei, A. Khalili and J. Famaey, "Power-efficient Joint Resource Allocation and Decoding Error Probability for Multiuser Downlink MISO with Finite Block Length Codes", in *2022 25th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, Herning, Denmark, Oct. 2022, pp. 232–237. (**Best Paper Award**) (*Chapter 4*)
3. **J. Jalali**, A. Khalili, A. Rezaei, R. Berkvens, M. Weyn and J. Famaey, "IRS-Based Energy Efficiency and Admission Control Maximization for IoT Users With Short Packet Lengths", *IEEE Transactions on Vehicular Technology*, vol. 72, no. 9, pp. 12379–12384, Sept. 2023. (*Chapter 5*)
4. **J. Jalali**, F. Lemic, H. Tabassum, R. Berkvens, and J. Famaey, "Toward Energy Efficient Multiuser IRS-Assisted URLLC Systems: A Novel Rank Relaxation Method", in *GLOBECOM 2023 - 2023 IEEE Global Communications Conference - 6G Communication Workshop*, Kuala Lumpur, Malaysia, Dec. 2023, pp. 1–7. (*Chapter 5*)
5. **J. Jalali**, A. Khalili, R. Berkvens and J. Famaey, "Joint Offloading Policy and Resource Allocation in IRS-aided MEC for IoT Users with Short Packet Transmission", in *2023 IEEE 98th Vehicular Technology Conference (VTC2023-Fall)*, Hong Kong, Hong Kong, Oct. 2023, pp. 1–7. (**VTS Student Travel Grant**) (*Chapter 6*)
6. **J. Jalali**, A. Khalili, A. Rezaei and J. Famaey, "Is Active IRS Useful for mmWave Wireless Networks or Not?", in *2023 International Conference on Computing, Networking and Communications (ICNC)*, Honolulu, HI, USA, Feb. 2023, pp. 377–382. (*Chapter 7*)
7. **J. Jalali**, M. Bustamante, F. Lemic, H. Tabassum, J. Struye, J. Famaey, and X. Costa Pérez, "Location Optimization and Resource Allocation of IRS in a Multi-User Indoor mmWave VR Network", in *2024 IEEE Wireless Communications and Networking Conference (WCNC)*, Dubai, United Arab Emirates, Apr. 2024, pp. 1–7. (**ComSoc Conference Travel Grant**) (*Chapter 7*)
8. **J. Jalali**, A. Khalili, H. Tabassum, R. Berkvens, J. Famaey and W. Saad, "Energy-Efficient THz NOMA for SWIPT-aided Miniature UAV Networks", *IEEE Communications Letters*, vol. 28, no. 5, pp. 1107–1111, May 2024. (**FWO Research Grant V478223N**) (*Chapter 8*)

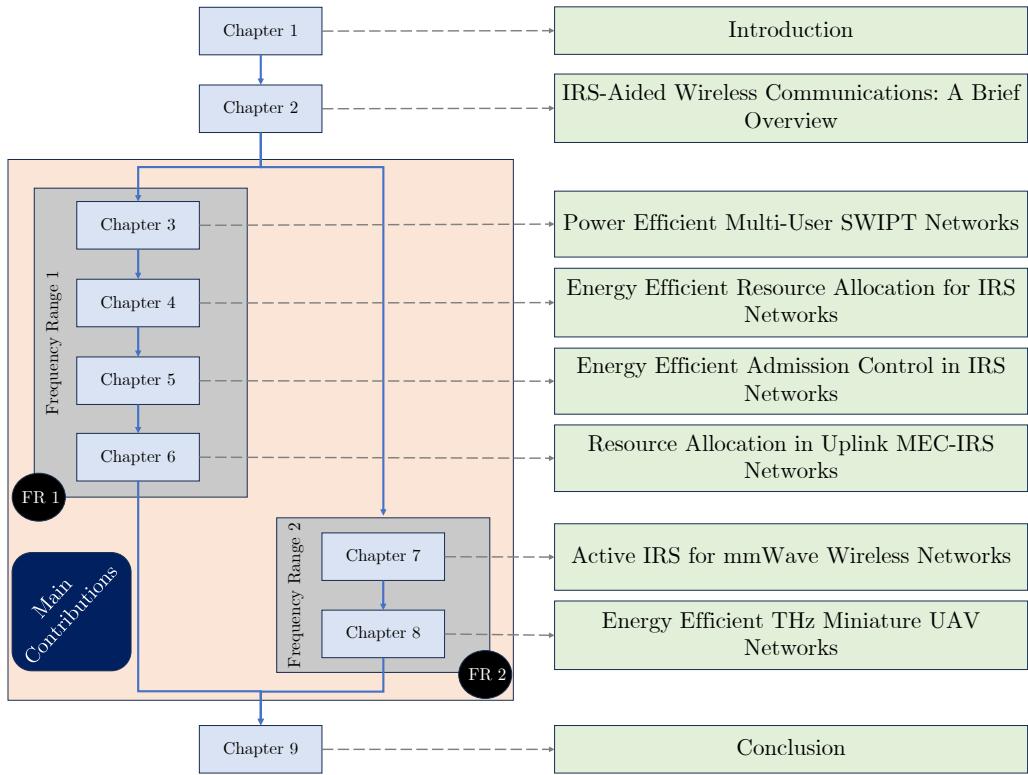


Figure 1.1: Dissertation outline.

1.3 Outline

Fig. 1.1 provides an overview of how the chapters relate to the contributions listed in the previous section. Each chapter is based on the publications listed in Section 1.2.

The dissertation proceeds with a detailed examination of IRS-enhanced wireless communication technologies in Chapter 2, providing readers with essential concepts pivotal throughout the study. Chapter 3 sets the groundwork by discussing the power efficiency metric in multi-user SWIPT networks, serving as the foundational chapter for subsequent analysis. In Chapter 4, we embark on an in-depth look at the application of IRS technology in enhancing URLLC services, followed by an exploration of IRS integration within MTC services in Chapter 5. Notably, Chapters 3, 4, and 5 focus on downlink communications, whereas Chapter 6 shifts the perspective to investigate an IRS-assisted MEC network in the uplink scenario.

Chapter 7 delves into the exploration of higher frequency bands of mmWave. Chapter 8 introduces a pioneering study on a miniature UAV network operating within the even higher frequency spectrum of THz. The dissertation culminates in Chapter 9, where we conclude and summarize our primary findings.

This dissertation comprehensively examines the transformative role of IRS across various technological domains. From boosting the efficiency of SWIPT networks and supporting URLLC and MTC, to facilitating MEC, and enhancing VR, mmWave, THz applications, IRS emerges as a key driver of innovation in wireless communication. This study illustrates the wide-ranging applications of IRS technology and establishes a foundation for future inquiries aimed at leveraging IRS to navigate the evolving challenges of contemporary wireless networks. It envisions a trajectory towards crafting more resilient, effective, and immersive communication environments.

Chapter 2

Intelligent Reflecting Surface-Aided Wireless Communications: A Brief Overview

THE future of mobile communications is set to undergo a significant transformation with the advent of 6th generation (6G) and beyond wireless networks, introducing a plethora of new applications and stringent technical demands. At the forefront of this shift is the integration of Intelligent Reflecting Surfaces (IRSs)¹, which revolutionizes the traditional view of the propagation channel. Historically seen as an unpredictable medium that adversely affects signal quality, IRS technology empowers network operators to precisely control the interaction of radio waves with their environment. This is achieved by manipulating the scattering, reflection, and refraction properties of radio waves, thus mitigating the inherent challenges of wireless propagation. IRS is particularly notable for its capability to modify the wavefront characteristics of signals, including phase, amplitude, frequency, and polarization, without necessitating complex signal processing operations. This chapter aims to provide a brief overview of IRS technology, tracing its development, distinguishing it from previous technologies, identifying key research questions, and highlighting the need for new communication-theoretical models in light of IRS.

This chapter does not aim to provide an exhaustive review of the latest advancements and the state of the art in IRS technology. Instead, it focuses on areas relevant to this dissertation. The content herein draws from a literature review conducted over the past four years in preparation for my publications. The purpose is to introduce the reader to key concepts that will be revisited throughout this dissertation.

¹Several terminologies are used to describe Reflective Surfaces, including software-controlled metasurfaces [59], Reconfigurable Intelligent Surfaces (RIS) [60], Intelligent Reflecting Surfaces (IRSs) [61], and reconfigurable intelligent metasurfaces [62]. For the sake of consistency and clarity in this dissertation, the term “IRS” will be used uniformly to refer to these technologies.

2.1 Intelligent Reflecting Surfaces

IRSs have been introduced as a novel paradigm to reconfigure the wireless propagation environment through software-controlled reflections. Comprising a planar array of numerous low-cost passive reflectors, each IRS unit is capable of independently adjusting the amplitude and/or phase of the incident signals. This capability enables precise three-Dimensional (3D) beamforming, setting IRS apart from traditional wireless link adaptation methods that modify signals at the transmitter or receiver. Instead, IRS proactively alters the wireless channel itself with highly controlled and intelligent reflections, offering a new Degree of Freedom (DoF) to boost communication performance in a smart, programmable wireless environment. With the advantage of not requiring transmit Radio Frequency (RF) chains and supporting short-range operation, IRS can be deployed densely to enhance network coverage cost-effectively and with minimal energy use, sidestepping the complex interference management typically required. Additionally, IRSs are versatile, designed to fit on various surfaces, thus accommodating a wide range of application scenarios, although this innovative approach necessitates further research in communication modeling and problem-solving [62].

2.2 Architecture of an IRS: From Concept to Implementation

IRSs are grounded in the technology of metasurfaces, which are essentially two-Dimensional (2D) arrays composed of metamaterials. These metasurfaces are made up of numerous meta-elements, each with distinct ElectroMagnetic (EM) characteristics determined by their design parameters. As shown in Fig. 2.1, an IRS incorporates these elements placed at intervals shorter than the wavelength of the signals they interact with. Upon encountering an EM signal, each IRS element generates a current, subsequently re-emitting an EM wave whose amplitude and/or phase may differ from that of the incoming signal. By strategically configuring these elements, an IRS can precisely modulate the phase and amplitude of the reflected EM waves to create a 3D beam. This coordination among IRS elements facilitates the manipulation of wireless signals, achieving passive beamforming towards specific directions without relying on external power sources. Such passive modulation of signals enables the IRS to intelligently direct reflections, significantly boosting the efficacy of wireless communication networks [54].

Diving deeper, Fig. 2.1 also illustrates the structured composition of the architecture of an IRS, which typically consists of three distinct layers alongside a sophisticated controller. The surface layer consists of numerous metallic patches arrayed on a dielectric substrate, serving the primary role of interacting with incoming signals. Positioned beneath this, a copper plate ensures no signal energy leakage. The foundational layer encompasses the control circuit board tasked with modulating each metallic element's reflection characteristics — amplitude and phase shift — under an attached smart controller. In practical setups, a Field-Programmable Gate Array (FPGA) is often deployed as this controller, doubling as a communication nexus with the broader network infrastructure (such as Base Stations (BSs), Access Points (APs), and User Equipments (UEs)) through dedicated low-rate wireless connections for information exchange.

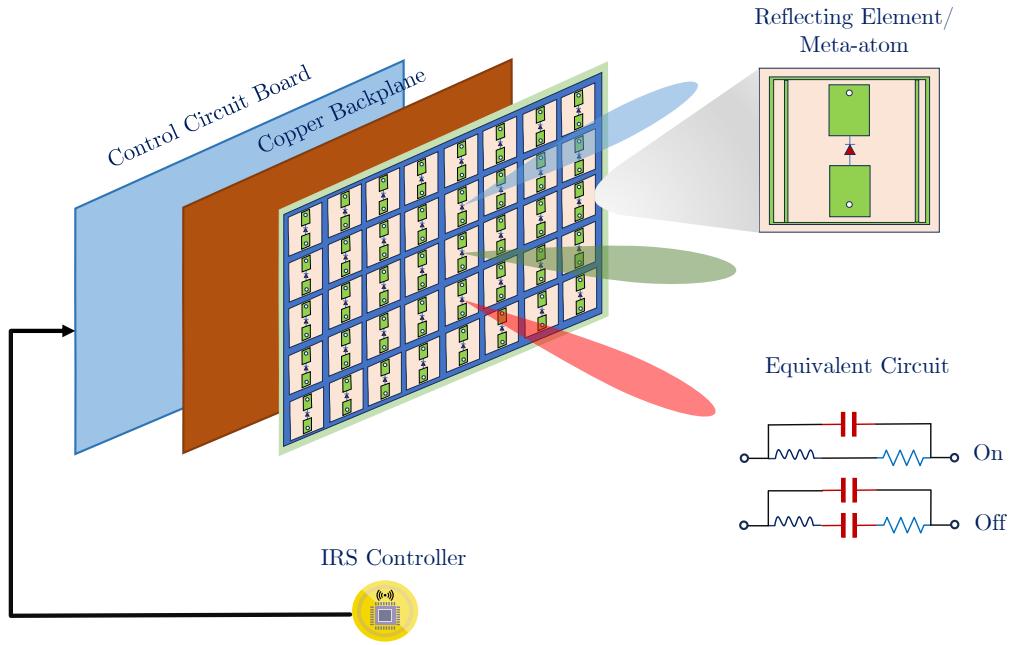


Figure 2.1: Architecture of an IRS.

A closer look at an individual IRS element's design, as depicted in Fig. 2.1, reveals the inclusion of a Positive-Intrinsic-Negative (PIN) diode². By varying the bias voltage through a DC feeding line, the PIN diode toggles between active ("On") and inactive ("Off") states, producing a phase shift of π radians. This mechanism allows for the individual calibration of each element's phase shift by adjusting their bias voltages via the smart controller. Furthermore, the design incorporates variable resistor loads to modulate the reflected signal's amplitude. Changing resistor load values adjust the dissipation of the incident signal's energy, facilitating tenable reflection amplitudes between 0 and 1. Achieving independent control over both amplitude and phase shift at each element necessitates the seamless integration of these components, highlighting the complexity and precision required in IRS design.

The utilization of IRS in wireless networks brings several benefits, including [64, 65, 66]:

- **Enhanced Signal Strength:** IRSs have the capability to intelligently redirect and fine-tune wireless signals, boosting signal strength in targeted areas. This is achieved by precisely altering the phase shifts of its elements, directing focused signal beams to

²Alternative mechanisms, such as Field-Effect Transistors (FETs) or MicroElectroMechanical System (MEMS) switches, are also viable options for implementing IRS technology [59]. For those interested in the practical implementation of IRS, a specific design for the 5.8 Gigahertz (GHz) frequency band is provided in [63]. This particular IRS features a total of 1100 metaatoms configured in a Uniform Planar Array (UPA) with 20 rows and 55 metaatoms per row. The impedance of each metaatom is controlled by two varactor diodes, facilitating the adjustment of the phase of reflected signals across a span of 240 degrees.

specific regions, thus mitigating path loss and elevating signal clarity.

- **Low Power Consumption:** Composed of numerous reflective elements, often implemented with diodes, IRSs operate without the need for active RF chains, leading to significantly reduced power consumption compared to traditional active antenna systems.
- **Programmability:** The elements within an IRS can be programmatically adjusted to alter the direction and properties of incoming EM waves, channeling them towards intended destinations. This adaptability allows for the achievement of various objectives such as signal boosting, latency minimization, interference mitigation, and coverage expansion.
- **Simplified Design:** IRSs are designed to be modular and lightweight, facilitating straightforward installation on surfaces like interior walls or ceilings. Their adaptable nature ensures easy incorporation into current wireless frameworks, enhancing system flexibility and adaptability.
- **Cost-Effectiveness:** Implementing IRS technology can be a more economical solution than the expansion of infrastructure or increasing the transmit power of base stations. By integrating with existing network architectures, IRS deployment avoids the necessity for substantial infrastructural enhancements, offering an efficient way to improve network quality and reach with minimal additional cost.

2.3 IRS vs. Other Technologies: Key Advantages and Differences

Through several key differences and advantages, IRSs distinguish themselves from other technologies, such as active relays, backscatter communication, and massive MIMO (m-MIMOs) with active surfaces [67]. Unlike active wireless relays, which amplify and retransmit signals using active components like power amplifiers, IRS simply reflects the received signals passively, without the need for active transmission modules. This not only makes the IRS more energy-efficient but also allows it to operate in a full-duplex (FD) mode, unlike the typically half-duplex active relays, thus offering greater spectrum efficiency. Although FD relays are possible, they necessitate complex and expensive self-interference cancellation solutions [68].

Contrary to traditional backscatter communication systems, such as Radio Frequency Identification (RFID) tags that modulate and reflect signals from a reader to communicate, IRS does not transmit its own information [69, 70]. Instead, it enhances an existing communication link by coherently combining direct and reflected paths at the receiver, thereby strengthening the signal for improved decoding without the need for self-interference cancellation techniques required by backscatter systems. Besides, the IRS differs from m-MIMOs systems with active surfaces in its foundational architecture and functionality. IRS employs a passive array architecture and operates by reflecting signals, as opposed to the active transmission approach of m-MIMOs. This distinction in array architecture (passive vs. active)

and operational mechanism (reflect vs. transmit) underlines the unique position of IRS in enhancing wireless communications with efficiency and cost-effectiveness [71].

The integration of IRS into wireless networks introduces a novel combination of active elements, such as BSs, APs, and user terminals, with (somehow) passive components, namely the IRS itself. This hybrid approach marks a departure from conventional networks that rely solely on active components. Consequently, this dissertation aims to explore the intricacies of IRS-aided networks, covering aspects such as signal and system modeling, the principles of passive and active beamforming, channel estimation, and deployment strategies. It particularly focuses on the primary challenges and proposes potential solutions for the development and application of IRS-aided networks, serving as a catalyst for further research in this emerging field.

2.4 Holographic Massive MIMO Surfaces

IRSs are not necessarily the same as Holographic Massive MIMO Surfaces (HMIMOS). The distinction between the two lies primarily in their functionality and design, which can be better clarified. HMIMOS are versatile components that are critical for achieving adaptable wireless environments. They can act as transmitters, receivers, or reflectors, making them highly flexible in their applications.

The term “active HMIMOS” applies when these surfaces incorporate energy-demanding RF circuits and signal processing units for transceiving purposes. Active HMIMOS represent an advanced iteration of traditional m-MIMO systems, characterized by an increased density of software-managed antenna elements on a compact 2D surface. As the quantity of these elements grows, leading to closer spacing between them, such configurations of HMIMOS are also known as Large Intelligent Surfaces (LIS). Active HMIMOS may feature a dense assembly of small antenna elements linked with reconfigurable networks, forming a unified antenna array that utilizes the hologram principle for signal dissemination and reception. Alternatively, they can employ discrete photonic antenna arrays with integrated active components for handling optical or RF signal transformations [72, 73, 74].

Conversely, passive HMIMOS, *identified as IRS*, function similarly to passive metallic mirrors or “wave collectors,” programmable to change incident EM fields. Unlike their active counterparts, passive HMIMOS consist of cost-effective elements that operate without external power, relying on energy harvesting for autonomy, aiming for energy neutrality. This technology stands out for its efficiency in molding and directing radio waves without needing power amplifiers or RF chains, and it operates without intricate signal processing. Capable of full-duplex communication, passive HMIMOS mitigate self-interference and maintain low noise levels with minimal control link requirements. Their low energy use and inexpensive hardware make them ideal for integration into various settings, from architectural facades to wearable technology, enhancing wireless networks’ flexibility and efficiency [75]. This is the power of IRS as passive HMIMOS.

2.5 Optimization of IRSs

Consider an IRS composed of Q unit cells, each acting as a diffusive scatterer capable of changing the phase of an impinging EM wave. This behavior is mathematically represented as $E_{r,q} = E_{t,q}\beta_q e^{j\phi_q}$ ³, where $E_{t,q}$ and $E_{r,q}$ denote the incident and reflected electric fields of the q -th unit cell, respectively, and $\phi_q \in \Phi \in [0, 2\pi)$ is the phase shift induced by the q -th unit cell, chosen from a set Φ , while $\beta_q \geq 1$ is reflection amplitude of the q -th IRS element [79, 80]. The task of configuring the IRS entails optimizing these phase shifts ϕ_q (and amplitudes β_q in case of an active IRS rather than passive IRS), leading to a non-convex optimization problem due to the unit-modulus constraint $|e^{j\phi_q}| = 1$. Although various strategies have been proposed in the literature to tackle this non-convex constraint, they generally do not scale well for large IRS configurations as the number of optimization variables becomes excessively large [81, 82, 60, 83, 84]. For example, an IRS of size 1 meter by 1 meter at a carrier frequency of 5 GHz, assuming a unit-cell spacing of half a wavelength, would encompass $Q = 1100$ unit cells. Consequently, direct optimization of ϕ_q for each cell in such large IRS arrays may not be feasible for on-the-fly design considerations.

In an IRS-enhanced communication framework that includes several multi-antenna BSs/APs, multiple IRS units, and a large user base equipped with either single or multiple antennas, managing IRS configurations centrally becomes a formidable challenge. This setup would necessitate the transmission of extensive control data to a central controller, leading to substantial computational and energy costs. To address these issues, there is a critical need to develop algorithms that can efficiently handle resource allocation, beamforming, IRS settings, and user scheduling. Furthermore, the network optimization process is expected to be further complicated by factors such as power distribution, spectrum utilization, and the assignment of users to specific BSs/APs and IRS units. As the network integrates more IRS units, the complexity of designing effective algorithms is expected to increase correspondingly.

In this dissertation, we address these challenges by devising resource allocation policies characterized by low computational complexity and optimal convergence behavior. Our approach encompasses a variety of use cases, exploring a broad spectrum of communication network metrics (such as power/spectral/energy efficiency, throughput, and user quality of service) and optimization strategies specifically tailored for IRS-aided networks. By adopting this methodology, we aim to enhance the efficiency and effectiveness of deploying IRS technology across diverse scenarios. Our focus on various optimization approaches enables us to tailor solutions that meet the unique requirements of each use case and mitigate the computational and energy constraints associated with centralized IRS management. This comprehensive examination of IRS-aided networks is crucial for implementing distributed algorithms that optimize network performance while effectively balancing computational load and energy usage.

³More complex models (that depend on the angle of arrival/departure, unit cell radiation pattern, and other relevant physical parameters) exist in the literature [29, 30, 31, 32, 34, 76, 77, 78].

2.6 Applications of IRSs

The introduction of IRSs is transforming wireless networks, enabling a host of promising applications. Fig 2.2 presents a vision for a future IRS-aided wireless network. IRS technology proves invaluable in extending coverage for millimeter-wave (mmWave) and Terahertz (THz) communications, which typically suffer from signal obstructions. By strategically placing IRS at cell edges, not only is the signal strength for users at these locations improved, but interference from adjacent cells is effectively reduced. This is particularly advantageous in settings such as smart offices or homes, where IRS can significantly counteract power loss over distances for systems like simultaneous wireless information and power transfer (SWIPT), thanks to its capability for directed beamforming towards proximate devices [85, 86, 87].

Further, IRS can be seamlessly integrated into the indoor infrastructure, attached to ceilings, walls, or even discreetly behind decorations, to foster enhanced network coverage and establish high-capacity connectivity zones. Such advancements are critical for enhanced Mobile BroadBand (eMBB) and massive Machine Type Communications (mMTC) services within diverse environments, including smart factories and commercial venues, offering improved device activity detection and efficiency by exploiting IRS' extra controllable paths in varying propagation conditions [86, 88]. Outdoors, the IRS finds utility in various structures, from building exteriors to public fixtures, supporting a range of applications, including URLLC for remote operation and intelligent transportation systems. By mitigating Doppler and delay spread effects, IRS contributes to the stabilization of wireless channels, thereby enhancing communication reliability essential for minimizing packet retransmissions and reducing latency (which is the key for URLLC applications) [89, 90].

As a pioneering technology, IRS is set to revolutionize the “dumb” infrastructure of today into an intelligent, interactive topography, promising substantial advantages across multiple sectors in the evolving 6G ecosystem. Its potential has ignited industry interest in developing and commercializing IRS-like technologies to forge new value chains. Concurrently, numerous pilot projects are underway, advancing research in this innovative field. Despite the diverse nomenclature, ranging from intelligent walls and smart reflect-arrays to large intelligent surfaces, all these innovations fundamentally rely on the principle of passive and adjustable surfaces for manipulating signal reflection or refraction [91, 92, 93].

2.6.1 Practical Scenarios: From Blockage Mitigation to Advance Usage

Fig. 2.3 depicts various practical scenarios where IRS-aided wireless networks can significantly enhance connectivity and security. Signal blockages and obstructions pose significant challenges to signal coverage, often compromising wireless communication systems' connectivity. In the first scenario, an IRS overcomes the challenge of signal blockage between a user and a BSs/APs, which is common in environments prone to obstruction, like indoor spaces affecting mmWave communications. Through the strategic placement of IRS elements, it is possible to reroute incident signals around obstacles, effectively creating a virtual Line-of-Sight (LoS) connection and extending coverage to areas otherwise shadowed by an APs [94, 95]. The second scenario addresses enhancing physical layer security. When an

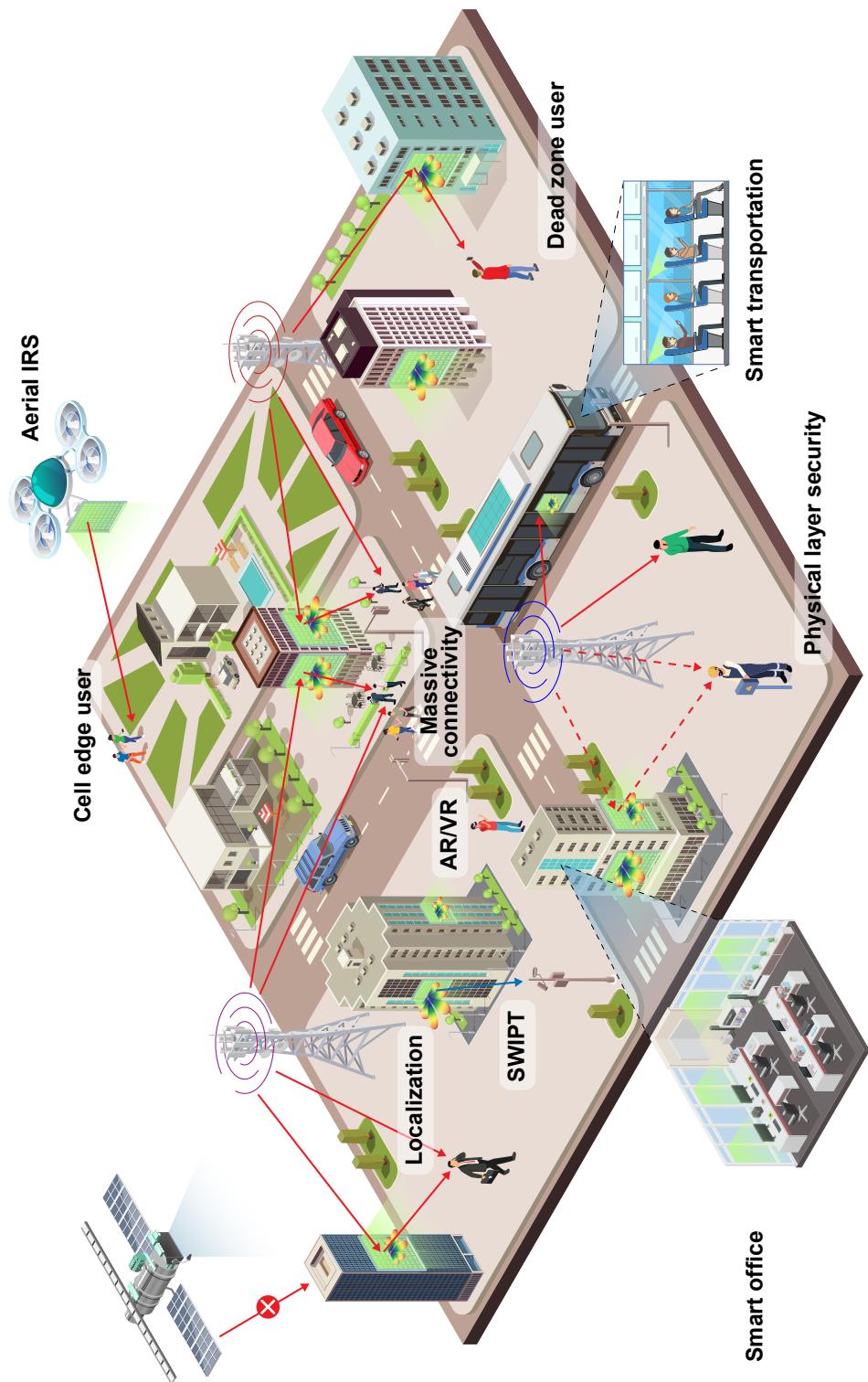


Figure 2.2: Illustration of IRS applications in future wireless network.

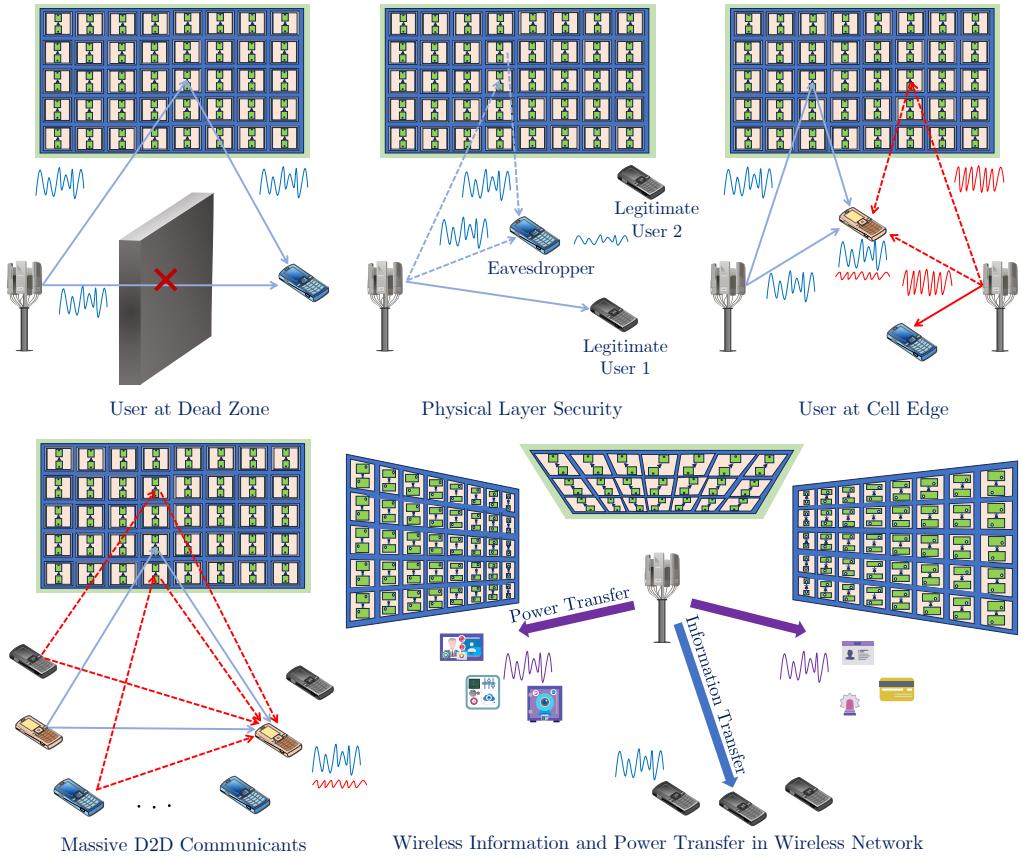


Figure 2.3: Typical IRS applications in wireless network.

eavesdropper is closer to the BSs than a legitimate user or in the same direction, traditional transmit beamforming struggles to prevent information leakage. Here, an IRS near the eavesdropper can adjust its signal reflection to negate the signal reaching the eavesdropper, thereby securing the communication [96, 97, 98]. In the third use case, an IRS aids a cell-edge user experiencing both weak signal strength from its BSs and interference from a neighboring BSs. By deploying an IRS at the cell edge and fine-tuning its beamforming, the desired signal is amplified while interference is minimized, creating both a “signal hotspot” and an “interference-free zone” [81, 99, 100, 101]. The fourth application demonstrates the IRS’s role in facilitating massive D2D communications. Here, the IRS serves as a central hub for signal reflection, enabling multiple low-power D2D links by mitigating interference among them. Lastly, the potential of the IRS in SWIPT scenarios is explored, particularly for Internet-of-Things (IoT) networks. Utilizing the large surface area of an IRS, it is possible to significantly improve the efficiency of wireless power transfer to IoT devices by compensating for power losses over distances with passive beamforming, showcasing the diverse



Figure 2.4: Application scenario for an IRS to mitigate blockage in an indoor factory.

capabilities of IRS in modern wireless communication systems [102, 103, 104, 105].

2.6.2 Indoor Environments: A Factory Setting

In industrial settings, as shown in Fig 2.4, the complexity of the environment often results in signal blockage due to numerous obstructions such as machinery, tools, and the constant movement of personnel, robots, and vehicles. The IRS offers a solution to this problem by facilitating a reliable connection between an APs and mobile users within the premises. By strategically positioning the IRS, it can dynamically redirect beams to navigate around obstructions, ensuring a robust and continuous link with users and the machines, even amidst changes or reconfigurations in the factory layout. The ability of the IRS to redirect its reflecting beams to accommodate the movement of users and bypass physical barriers guarantees uninterrupted wireless communication. For enhanced coverage and flexibility, mounting the IRS on a mobile platform allows for the adjustment of its location to suit the factory's current configuration and the users' positions, optimizing the signal quality and addressing blockage challenges effectively. Furthermore, the deployment of a considerable number of IRS elements allows the system to adapt flexibly to the wireless

environment, navigating around blockages. These elements collaboratively build a robust reflective Non-Line-of-Sight (NLoS) link, thereby amplifying signal propagation and broadening the coverage area [106, 107].

The implementation of an IRS in such environments circumvents the need for major infrastructural modifications or the extensive repositioning of network components [108]. With its ability to be reconfigured, the IRS enhances connectivity, significantly improving both operational and communication efficiency in smart manufacturing environments [109]. Its capacity to maintain a dependable connection with moving users boosts productivity and enhances the factory's overall performance [110]. Further insights into the application of IRS in industrial settings, particularly for systems accommodating Machine-Type Communication (MTC)⁴ and URLLC services [111, 112], are detailed in Chapters 4 and 5 of this dissertation.

2.7 IRS-aided System Design: Dissertation Objectives

Research on optimizing IRS configurations has proliferated due to their ability to reshape wireless propagation channels [76, 113, 114, 115, 116, 117, 118, 119, 120, 121]. Various studies demonstrate the technology's versatility, such as those proposing robust resource allocation algorithms for IRS-enhanced systems and exploring resource distribution in IRS-supported OFDMA frameworks [122, 123]. While enhancements in multi-cell [81] and full-duplex [124] systems have been explored, along with IRS's integration into non-orthogonal multiple access (NOMA) setups [125], the emphasis has largely been on eMBB traffic and predominantly single-carrier systems [66, 88, 126]. The incorporation of IRS into both URLLC and MTC services introduces a transformative shift, notably reducing latency and improving system reliability. In this dissertation, our investigation extends IRS into such services, further expanding the scope of the research.

The strategic manipulation of EM waves enables IRS to amplify power efficiency within multi-user SWIPT systems [50]. This dual functionality underscores the flexibility of IRS in promoting energy harvesting and reliable data transmission. IRS proves invaluable for Virtual Reality (VR) networks that encounter significant path loss or obstructions, ensuring uninterrupted, high-quality experiences. IRS also enhances Mobile Edge Computing (MEC) by optimizing signal delivery for efficient edge data processing [101]. This improvement is critical for applications demanding immediate feedback and strict data integrity, such as autonomous vehicles and industrial automation, highlighting the IRS's crucial role in addressing next-generation communication requirements. Subsequent to our examination, additional studies have delved into the use of miniature unmanned aerial vehicle (UAV) networks. The exploration in this dissertation further spans the strategic application of IRS across a wide frequency spectrum, from Frequency Range 1 (FR1) to Frequency Range 2 (FR2), and into the advanced territories of mmWave and THz frequencies, illustrating its profound impact on the future of telecommunications.

⁴MTC refers to automated communication between devices or machines, characterized by minimal or no human intervention, typically utilizing protocols and technologies designed for efficient and scalable data exchange in IoT and industrial automation environments.

To assess the efficacy of IRS-enhanced networks, this dissertation identifies a suite of KPIs, including data rate, power and energy efficiency, Signal-to-Interference-plus-Noise (SINR), transmit signal power budget, and received power strength. These KPIs are crucial for evaluating and refining the network's performance through the development of an effective resource allocation strategy. Non-linear, non-convex, and MINLP problems arise when addressing the resource allocation optimization problem. These problems are NP-hard due to the complex relationship between variables and the system's constraints. Given the complexity of these optimization problems, different strategies are used to simplify and approach their solution. By relaxing the objective function and constraints that are non-convex to a more tractable format, the problems became more manageable. This relaxation approach often involved transforming the optimization problem into its convex equivalent or utilizing approximation techniques to linearize or convexify non-convex terms.

Algorithms are developed that are capable of solving the main problem either globally or sub-optimally but sufficiently close to the global optimum. These solutions employ optimization solvers and computer simulations, exploiting advanced mathematical tools and techniques such as the big-M method for linearizing product terms involving binary variables and Successive Convex Approximation (SCA) to obtain convex approximations of non-convex terms. The iterative nature of these solutions allowed for step-by-step refinement, gradually moving towards an optimal configuration of a resource allocation design despite the initial problem's complexity.

Through exhaustive simulations, this dissertation unveils the diverse performance improvements achievable through resource allocation in IRS-assisted networks, providing rich insights into how IRS technology can improve wireless communication systems. These simulations serve as a critical bridge, connecting theoretical predictions with empirical evidence and validating the practical feasibility of the proposed IRS-enhanced network. By exploring various IRS configurations — examining both passive and active types and varying the number of reflective elements — and their implementation in different environments and settings, this study not only confirms the theoretical models' accuracy but also explains the conditions under which IRS deployments yield maximal performance gains, manifesting the IRS versatility in adapting new technologies.

Collectively, this dissertation studies the impact of IRS across a broad range of technologies. By enhancing the performance of SWIPT networks, facilitating URLLC and MTC, enabling VR, and revolutionizing VR, mmWave, and THz applications, IRS stands at the forefront of wireless communication innovation. This work demonstrates the diverse applications of IRS technology and lays the foundation for future research aimed at utilizing IRS to tackle the dynamic challenges of modern wireless networks. It charts a path toward the development of more robust, efficient, and engaging communication ecosystems.

In this dissertation, we contribute to the IRS-aided system design for different service type users with three principal innovations:

- **Extending IRS to URLLC, MTC and other Use-Cases:** We explore the use of IRS in reducing latency and enhancing reliability for URLLC and MTC services, broadening the scope of IRS applications beyond traditional eMBB traffic. We further investigate IRS functionality in VR, MEC, and UAV networks.

- **Developing Optimization Algorithms:** We introduce algorithms for solving complex optimization problems in IRS-assisted networks, utilizing advanced mathematical techniques to approach global or near-global optima efficiently.
- **Performance Validation through Simulations:** Through simulations, we demonstrate the significant improvements IRS can bring to wireless networks, validating theoretical predictions and showcasing the technology's adaptability across various applications.

Chapter 3

Power Efficient Multi-User SWIPT Networks

BEFORE diving into Intelligent Reflecting Surface (IRS) technology, we need to develop a baseline. This chapter lays the groundwork for a comprehensive understanding of power efficiency in a downlink multi-antenna, multi-user single-cell network, particularly in the context of Simultaneous Wireless Information and Power Transfer (SWIPT). The proposed power efficiency problem aims to maximize the harvested energy and minimize transmission power consumption simultaneously. The emphasis on maximizing the harvested energy while minimizing transmission power consumption addresses a crucial aspect of wireless communication networks. The approach includes optimizing beamforming and antenna selection procedures at the receivers, considering minimum data rate requirements, which introduces a complex optimization landscape. The problem is identified as a non-linear programming problem, highlighting the complex nature of achieving power efficiency in such networks.

As a result, a joint optimization of beamforming and antenna selection design is performed based on the scheduling chosen for information decoding and energy harvesting. This method involves breaking down the main problem into two manageable subproblems: *antenna selection optimization* based on maximum channel gain across all antennas and *transmit beamforming optimization* through a unique two-layer iterative process leveraging the sum-of-ratio programming. This dual approach enables a low-complexity, locally optimal solution, significantly improving power and energy efficiency.

Furthermore, the chapter presents simulation results to validate the effectiveness of the proposed scheme, demonstrating notable improvements in power and energy efficiency. These results also uncover an intriguing balance between power and energy efficiency, suggesting an inherent tradeoff that must be navigated in the design of SWIPT-enabled networks.

This first technical chapter aims to develop a baseline fundamental backbone, building the foundation for the subsequent chapters. The concepts, methodologies, and solutions introduced here are essential for understanding and advancing the state of power efficiency in multi-antenna, multi-user single-cell networks. As such, this chapter serves as the cornerstone upon which the rest of the discussions and explorations in this book will be built, ensuring a coherent and progressive development of ideas and technologies in the realm of IRS-aided wireless communications.

This chapter is based on:

J. Jalali, A. Khalili, A. Rezaei, J. Famaey and W. Saad, "Power-efficient Antenna Switching and Beamforming Design for Multi-User SWIPT with Non-Linear Energy Harvesting", in *2023 IEEE 20th Consumer Communications & Networking Conference (CCNC)*. Las Vegas, NV, USA, Jan. 2023, pp. 746–751. <https://doi.org/10.1109/CCNC51644.2023.10059879>

3.1 Introduction

Recently, the technology of Simultaneous Wireless Information and Power Transfer (SWIPT) has been recognized as a promising approach for simultaneously enhancing the Energy Efficiency (EE) and extending the battery lifespan of wireless devices [127]. Research in this field has primarily concentrated on either increasing the amount of energy harvested or improving throughput levels [128, 129]. However, focusing exclusively on enhancing throughput may lead to elevated power consumption within the network, while prioritizing the maximization of harvested energy through SWIPT could degrade Quality of Service (QoS). In response to this dilemma, EE has been proposed as a pivotal metric designed to handle the delicate balance between minimizing power consumption and maximizing throughput.

This metric introduces a different understanding of network performance, emphasizing the importance of achieving a harmonious equilibrium between the dual objectives of energy conservation and information transmission efficiency. By evaluating both energy harvest and throughput within a unified framework, EE offers a comprehensive metric that addresses the complexities and inherent trade-offs involved in SWIPT technology. This approach not only underscores the multi-layered challenges of optimizing network operations but also highlights the potential for innovative solutions that can reconcile competing objectives in the pursuit of sustainable and efficient wireless communication systems.

The quest for energy-efficient resource allocation design in SWIPT networks is a focal point of numerous scholarly studies, as evidenced by the array of research tackling this issue [130, 131, 132, 133, 134, 135, 136]. These investigations have unfolded a variety of innovative strategies designed to elevate the EE across diverse network architectures. Among these strategies, beamforming techniques within complex multi-cell, multi-user environments have been emphasized for their potential to significantly enhance network performance [130]. Furthermore, the optimization of EE in Multiple-Input Single-Output (MISO) Orthogonal Frequency Division Multiple Access (OFDMA) frameworks through advanced beamforming methodologies, such as zero-forcing, has been thoroughly examined [137].

The body of work extends beyond these approaches, with additional research focusing on the enhancement of EE through beamforming strategies specifically designed for OFDMA networks [131]. This is complemented by in-depth investigations into Non-Orthogonal Multiple Access (NOMA) based SWIPT networks. Such studies focus on achieving maximal EE by finely tuning the balance between power allocation and Time-Switching (TS) control in TS-based SWIPT architectures [132]. This collective research effort represents a concerted move towards optimizing the dual functionality of SWIPT networks, ensuring not only the efficient transmission of information but also the effective harvesting of energy.

Building upon the same research dedicated to enhancing EE in SWIPT networks, the study presented in [133] explores a heterogeneous NOMA SWIPT network, introducing a novel solution for EE maximization. This solution considers the matching theory concept coupled with the application of Lagrangian duality, presenting a sophisticated mathematical framework for addressing EE optimization problems. Furthermore, [134] advances the discourse by proposing an EE optimization strategy through a meticulous subcarrier allocation design policy. This strategy is aimed at fostering green communication in wireless sensor networks by harnessing the capabilities of SWIPT technology, marking a significant step

towards sustainable network operations.

Additionally, the research in [135] delves into a Multiple-Input Single-Output (MISO) SWIPT network, distinguished by a non-linear energy harvesting model. Here, a comprehensive approach to EE maximization is undertaken, involving the joint optimization of Power-Splitting (PS) ratios and beamforming design. This approach features the intricate balance required to maximize EE without compromising the network's operational capabilities. Meanwhile, the work in [136] introduces an Energy Efficiency Indicator (EEI), specifically designed to mediate the relationship between data rate and energy harvesting. This indicator serves as a tool for calibrating network performance with an eye towards achieving an optimal balance between these two critical dimensions.

Despite these considerable advancements, it is striking to note that previous studies have largely overlooked the potential of Antenna Switching (AS) techniques at the receiver level to balance the trade-off between Information Decoding (ID) and Energy Harvesting (EH) [130, 131, 132, 133, 134, 135, 136]. This oversight suggests an opportunity for further exploration into how AS techniques could enhance the dynamic interplay between ID and EH, potentially unlocking new pathways to superior energy efficiency. The ability of multiple antennas to switch between decoding and harvesting modes, based on real-time assessments, could dramatically elevate the efficiency and adaptability of SWIPT networks, heralding a new era of smart, energy-efficient wireless communication.

Intuitively, the concept of utilizing multiple receive antennas presents a promising avenue for enhancing both energy harvesting capabilities and the efficiency of information transfer in SWIPT networks. This approach benefits from the potential for increased energy collection and improved signal reception, thus facilitating a more robust and efficient communication system. Furthermore, AS provides a trade-off between operational costs, system complexity, and overall network performance. By selectively activating certain antennas based on prevailing conditions, networks can achieve optimal functionality with minimized resource expenditure.

The process of receiver antenna selection extends the AS scheme within the specialized environment of co-located SWIPT networks [138]. This advanced methodology allows for each antenna at the user's device to be dynamically designated for either ID or EH according to real-time Channel State Information (CSI). Such flexibility not only enhances the adaptability of the network but also optimizes the dual objectives of SWIPT technology. We refer to this methodology as "*generalized AS*" technique in SWIPT-based networks, signifying a leap forward in the strategic deployment of network resources. Here, the AS function acts as a decisive "switch," orchestrating the operational mode of antennas to ensure they fulfill their dual potential in EH and ID as necessitated by the network demands, as illustrated in Fig 3.2.

The exploration of AS techniques in SWIPT systems represents a niche yet significant area of study, with only a handful of research works delving into the intricacies of how antennas can dynamically alternate between ID and EH functionalities [139, 140, 141, 142]. Among these, the study conducted in [141] is notable for its innovative approach, proposing an antenna-clustering methodology that leverages hybrid Deep Reinforcement Learning (DRL) to optimize the average data rate for systems equipped with multiple antennas at both the receiver and transmitter ends. Similarly, the work presented in [142] explores an AS

strategy specifically designed for multi-antenna secondary receivers within cognitive-based networks, employing a thresholding method to facilitate the switching process. Despite these advancements, it is observed that the existing literature, including the studies [129, 139, 141], and [142], primarily overlooks the EE aspect of networks employing AS. In contrast, [140] takes a step forward by assessing EE within a Point-to-Point (P2P) Multiple-Input and Multiple-Output (MIMO) SWIPT system, underlining the potential for improving network performance through efficient energy use.

Motivated by the practical scenarios, the critical gap identified in the literature is the absence of any study focusing on the optimization of “power efficiency” in multi-user SWIPT networks through a combined approach of generalized AS and beamforming techniques. This approach not only aims to enhance the power harvested and data rate but also seeks to minimize power consumption, thereby ensuring a sustainable balance that meets the network’s quality conditions. Given the evolving needs of modern wireless networks, such as increased data demand and the push for energy sustainability, the integration of generalized AS with beamforming strategies presents a promising avenue for research. This novel direction could pave the way for significantly more energy- and power-efficient SWIPT networks, tailored to meet the challenges of contemporary and future wireless communication systems.

This chapter delineates a pioneering approach through the implementation of generalized AS alongside a beamforming design tailored for multi-user SWIPT systems, aiming primarily to elevate the network’s power efficiency. The essence of our contribution is encapsulated in the strategic deployment of these technologies to refine the balance between energy harvesting and power consumption, thereby optimizing the overall network functionality. The core aspects of our contributions are detailed as follows:

- We first present the difference between *generalized* AS and other SWIPT architecture. Our investigation then begins with a critical evaluation of the system’s power efficiency, which we articulate as the ratio of the aggregate harvested energy to the total power expenditure [143]. This foundational concept serves as the basis for our subsequent optimizations and analyses in the generalized AS SWIPT network.
- Building upon this, our objective expands to encompass the holistic optimization of the network’s effective power. By designing beamforming for both information and energy signals within a MIMO framework that incorporates generalized AS, we undertake a comprehensive approach. Essential to our methodology is the assurance of maintaining each user’s minimum Quality of Service (QoS), in order to balance the tradeoff between ID and EH. We maximize the effective power throughput subject to minimum data rate and maximum power transfer constraints. Achieving these ambitious goals necessitates a deliberate selection of receiver antennas and an optimization of transmit beamforming, tailored distinctly to the network’s unique attributes. The complexity and non-linear nature of this problem are acknowledged, setting the stage for innovative solutions.
- In response to the challenging nature of the original optimization problem, we adopt a strategic approach by decomposing it into two manageable subproblems, thereby paving the way for a locally optimal solution. In particular, the first subproblem is solved via searching for the best channel gain across all antennas. The objective function in the second subproblem follows the sum of objective ratio functions that will

be transformed into an equivalent subtractive form. To navigate these challenges, we leverage the Semi-Definite Programming (SDP) relaxation technique coupled with a one-dimensional search methodology, facilitating an iterative progression toward obtaining an optimal solution.

- The practical validation of our theoretical constructs is achieved through rigorous simulation exercises. These simulations not only affirm the robustness of the proposed algorithm in enhancing power efficiency and energy efficiency across a spectrum of antenna and sensor user configurations but also illuminate the inherent trade-offs between these two critical efficiencies. The insights learned from these simulations are instrumental in demonstrating the tangible impacts of our contributions.

In essence, this chapter not only provides the innovative application of generalized AS and beamforming within the realm of SWIPT networks but also illustrates the profound implications of these strategies on the operational efficiency of such networks. Through a methodical dissection of the problem and the strategic employment of advanced mathematical techniques, we offer a blueprint for achieving an optimal balance between energy harvesting and information transmission, heralding a new era in the design and optimization of SWIPT systems.

This chapter aims to develop a baseline fundamental backbone, building the foundation for the subsequent chapters. The concepts, methodologies, and solutions introduced here are essential for understanding and advancing the state of power efficiency in multi-antenna, multi-user single-cell networks. As such, this chapter serves as the cornerstone upon which the rest of the discussions and explorations in this dissertation will be built, ensuring a coherent and progressive development of ideas and technologies in the realm of wireless communications.

The rest of the chapter is organized as follows. This chapter unfolds with a comparative analysis of *generalized* AS against traditional SWIPT architectures in Section 3.2. We then detail the system model and problem formulation in Section 3.3. Section 3.4 introduces a solution to the proposed scheme. This is followed by Section 3.5, where we validate our theoretical models with extensive simulations, demonstrating the practical viability and benefits of our proposed method. The chapter concludes in Section 3.6, where we summarize our findings and reflect on their implications for the advancement of SWIPT systems.

Notations: We denote the matrices and column vectors by boldfaced lowercase and uppercase letters, e.g., \mathbf{A} and \mathbf{a} , respectively. $\|\mathbf{a}\|$ indicates the Euclidean norm of vector \mathbf{a} and $|a|$ describes the magnitude of a complex number a . The transpose and Hermitian of a matrix are expressed as $(\cdot)^T$ and $(\cdot)^H$, respectively. Moreover, $\text{Tr}(\mathbf{A})$ and $\text{rank}(\mathbf{A})$ define the trace and the rank of matrix \mathbf{A} , respectively. $\text{diag}(\mathbf{A})$ specifies a vector whose elements are taken from the main diagonal elements of the matrix \mathbf{A} . $\mathbf{A} \succeq 0$ indicates \mathbf{A} is a positive semidefinite matrix. $\mathbb{C}^{N \times M}$ is the space of an $N \times M$ with complex element entries. The set containing the elements a and b is denoted as $\{a, b\}$. Matrix \mathbf{I}_M expresses an $M \times M$ identity matrix. $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ gives the distribution of a Circularly Symmetric Complex Gaussian (CSCG) random vector with the mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, where \sim means “with the distribution of.”

3.2 Generalized AS versus other SWIPT Architectures

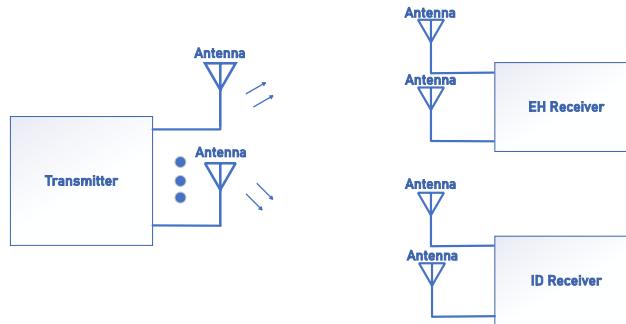
In our discussion, we delve into the concept of generalized AS within the context of SWIPT networks, a burgeoning area of research that intersects with multiple Wireless Power Transfer (WPT) technologies. The core idea behind SWIPT is its ability to utilize radio frequency (RF) signals for dual purposes — both to convey information and to transfer energy to energy-constrained wireless User Equipments (UEs). This innovative approach allows UEs to simultaneously harvest energy and process information from RF signals emanating from a Base Station (BS), a mobile AP (e.g., a drone), or an Access Point (AP).

SWIPT systems are designed to facilitate the concurrent transfer of energy and information signals in the DL direction from one or multiple BSs or APs to one or several receivers. This setup is optimized for simultaneous ID and EH, fundamentally altering the dynamics of wireless communication systems by enhancing their energy efficiency and operational capabilities. Ideally, a receiver equipped for SWIPT would possess integrated circuitry capable of performing both ID and EH concurrently. This integration represents a departure from traditional designs, where separate circuits were utilized for EH and ID, marking a significant evolution in the design and functionality of receivers within SWIPT networks [144, 145, 146]¹.

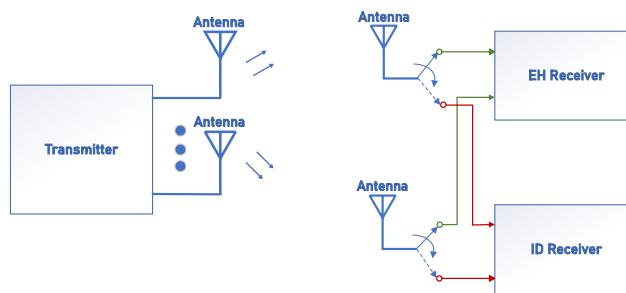
While simultaneous EH and ID operations are embodied in the SWIPT technology, it is crucial to understand that these processes do not necessarily occur on the same segment of the received signal. Practically, attempting to harvest energy directly from the signal carrying information would compromise the integrity of the data within the RF domain, rendering the information content unusable. Moreover, relying on a singular antenna for both EH and ID tasks might not provide a consistent energy supply, given the inherent limitations in energy collection capabilities of a single antenna setup. To circumvent these challenges and effectively enable SWIPT, distinct strategies are employed, such as dedicating separate antennas for EH and ID operations or dividing the incoming RF signal into two distinct paths — one for EH and another for ID — through the use of a *splitter*. This differentiation is essential for the practical implementation of SWIPT systems, ensuring both energy harvesting and information decoding can occur efficiently without interference between the two processes.

The architectural design of EH and ID receivers within SWIPT systems can be broadly categorized into two types: separated and co-located architectures (cf. Fig. 3.1). In a separated architecture, EH and ID functions are performed by two distinct devices, each equipped with its own antenna and experiencing different channel conditions from the transmitter. As shown in Fig. (3.1a), the EH receiver, designed for low-power operation, focuses solely on energy collection, whereas the ID receiver processes the data. Given the efficiency of energy harvesting diminishes with distance, EH receivers are typically positioned closer to the BS or AP compared to ID receivers, necessitating spatial separation and often defined by an inner and outer radius to demarcate EH and ID zones. Conversely, a co-located SWIPT architecture integrates both EH and ID capabilities within a single device, receiving identical channel conditions from the transmitter. This setup enables the device to simultaneously perform EH and ID without the need for spatial separation, offering a more elegant approach to SWIPT but requiring sophisticated internal mechanisms to manage the different func-

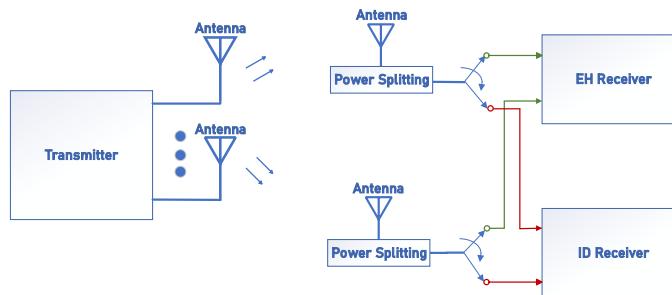
¹The concept of generalized AS in this framework further enriches this evolving landscape, promising new avenues for maximizing the efficiency and effectiveness of SWIPT systems.



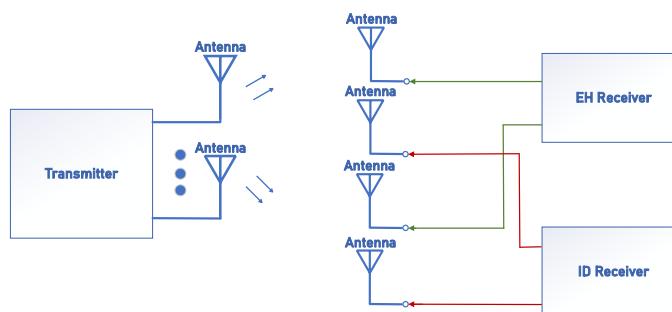
(a) Separated receiver architecture.



(b) Time-Switching (TS) approach to realize co-located SWIPT architecture.



(c) Power-Splitting (PS) approach to realize co-located SWIPT architecture.



(d) Antenna-Switching (AS) approach to realize co-located SWIPT architecture.

Figure 3.1: Integrated receiver architecture designs for SWIPT.

tionalities efficiently. Each architecture offers distinct advantages and challenges, shaping the deployment and effectiveness of SWIPT systems in various operational environments.

Within the realm of co-located receiver architectures for SWIPT, three pragmatic approaches — Time Switching-(TS), Power-Splitting (PS), and Antenna-Switching (AS) — stand out for their innovative means of enabling simultaneous EH and ID operations. These methods ensure that EH and ID receivers, despite sharing the same physical space, can efficiently manage the tasks without compromising on performance.

In the TS approach, as illustrated in Fig. (3.1b), the architecture includes an EH module, an ID module, and a switch. This setup allows the receiving antenna to alternate between EH and ID modes according to a predefined, yet adjustable, TS sequence. This method requires precise scheduling of information and energy reception, alongside accurate timing mechanisms, to ensure the seamless transition between modes. Conversely, the PS approach, depicted in Fig. (3.1c), employs a strategy where the incoming RF signal is divided into two distinct streams at varying power levels dedicated to EH and ID tasks, respectively. This division is governed by an optimizable PS ratio, enabling a balanced distribution of power for both functions within the same time frame. The AS methodology, as in Fig. (3.1d), introduces a paradigm where independent antennas are designated for EH and ID activities, facilitated by a straightforward AS algorithm. This setup typically involves an antenna array at the receiver, employing spatial multiplexing to segregate the antenna into subsets dedicated to either EH or ID functions. This configuration allows for a dynamic allocation of resources, with one subset of antennas focusing on energy harvesting while the remaining antennas are tasked with decoding information.

The AS approach is notably recognized for its simplicity and practicality, making it a particularly appealing option for SWIPT architecture designs. It offers a straightforward solution without the complex scheduling required by TS or the precise power allocation demands of PS [147]. Moreover, its flexibility and lower complexity render it an optimal strategy for not only co-located receiver architectures but also for enhancing separated receiver designs, as suggested by Fig. (3.1a). This adaptability and ease of implementation highlight the AS method as a superior choice for real-world applications of SWIPT systems. In addition, the AS approach can be similarly adopted to optimize the separated receiver architecture as shown in Fig. (3.1a) [148].

The AS approach, particularly in its advanced form of *generalized AS*, emerges as a notably strategic and flexible methodology in co-located SWIPT networks. By extending the basic premise of AS to allow for the dynamic selection of receiver antennas, based on real-time CSI for either ID or ED, this methodology significantly broadens the operational versatility of SWIPT systems. Such an advanced approach introduces a paradigm shift in how antennas are utilized, seamlessly alternating between EH and ID to maximize network efficiency and resource utilization. In the traditional AS framework, antennas are separated into distinct groups, each dedicated to either EH or ID. Conversely, generalized AS supplies each antenna with both EH and ID functionality. This allows antennas in the generalized setup to 'switch' between two specific roles, unlike in conventional AS, where a physical switch is used to toggle between different antennas for EH and ID tasks. Our introduction of the *generalized AS* concept represents a first in the field, offering a novel and sophisticated solution that optimally aligns with the objectives of SWIPT technology. This innovative leap enhances the adaptability and efficiency of SWIPT networks, ensuring that each antenna within a user's

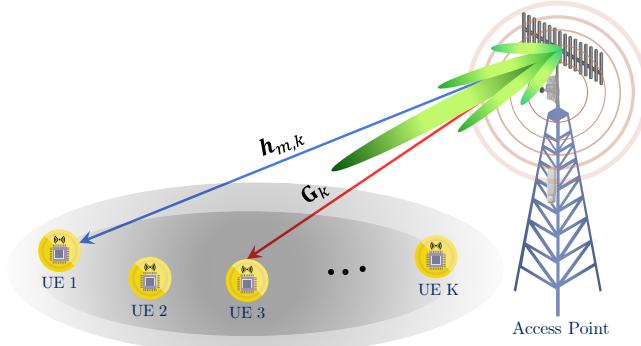


Figure 3.2: Generalized AS approach to realize SWIPT architecture in a single-cell multi-user MIMO network.

device is employed to its fullest potential in accordance with the instantaneous demands of the network environment.

3.3 System Model and Problem Formulation

In our study, we focus on a DL Orthogonal Frequency Division Multiplexing (OFDM) network architecture, where an AP provides coverage to multiple sensor User Equipments (UEs) within its range. The AP is equipped with N_T antennas, while each sensor user is equipped with M antennas, allowing for enhanced communication capabilities and network efficiency. We define the set of k sensor users under the AP's coverage as $\mathcal{K} = \{1, 2, \dots, K\}$, representing a diverse array of devices with varying communication needs.

A key assumption in our model is the availability of perfect CSI at the central resource allocator². This assumption enables the design of an optimized resource allocation policy that can dynamically adjust to various network conditions. The perfect CSI ensures that the resource allocation mechanism can make informed decisions, maximizing the network's overall performance by efficiently utilizing the available spectral resources. This approach emphasizes the importance of accurate information in the strategic planning and execution of network resource distribution, aiming to enhance the communication experience of all sensor users within the DL OFDM network.

In our system model, where each user is equipped with multiple antennas, the framework allows for the strategic selection of the optimal antenna for specific tasks — either ID or energy EH, in accordance with the predefined resource allocation policy. This configuration enables a user to engage in both ID and EH activities simultaneously, though each activity is allocated to different antennas to avoid interference and maximize efficiency. Specifically, the finest antenna, chosen from the antenna set \mathcal{M} , is dedicated to ID, while the other available antennas are utilized for EH. This methodical selection ensures that each antenna

²It is assumed that the AP has perfect CSI through a feedback channel. In particular, the AP sends some orthogonal preambles in the downlink to the sensor users and obtains the CSI by listening to the sounding reference signals transmitted by the sensor users [149].

Table 3.1: Summary of Our Main Notations in Power-efficiency Optimization of a Multi-user SWIPT Network.

Symbol	Definition
$\mathbf{h}_{m,k} \in \mathbb{C}^{N_t \times 1}$	The DL channel gain vector for the information transfer from the AP to the m^{th} antenna of user k .
$\mathbf{G}_k \in \mathbb{C}^{N_t \times M}$	The DL channel matrix for the wireless power transfer from the AP to the user k .
$b_{m,k} \in \{0, 1\}$	Binary indicator that selects the m^{th} antenna from the AP to the k^{th} user for data transmission.
$\mathbf{w}_k \in \mathbb{C}^{N_t \times 1}$	The transmit information beamforming of the AP for the k^{th} user.
$\mathbf{w}_e \in \mathbb{C}^{N_t \times 1}$	The transmit energy signal of the AP broadcasted to all sensor users.

is used to its fullest potential, aligning with the system's operational objectives. To aid in the comprehension of our system model and the underlying principles of antenna selection for ID and EH, we have delineated key variables and their definitions in *Table 3.1*. For readability, we summarized some of the essential variables used to describe the system model in *Table 3.1*. We further assume that the AP transmits both the information and energy signals simultaneously. Thus, the discrete-time signal transmitted by the AP can be mathematically represented as follows:

$$\mathbf{x} = \sum_{k \in \mathcal{K}} \mathbf{w}_k s_k + \mathbf{w}_e, \quad (3.1)$$

where $s_k \in \mathbb{C}$ is a unit-energy information carrying symbol intended for the k -th user. This formulation, (3.1), encapsulates the essence of SWIPT by integrating the transmission of information-bearing signals, identified by the weighting vectors \mathbf{w}_k for each user k , with an energy signal, denoted by \mathbf{w}_e . It is important to highlight that the energy component of the transmission, while devoid of information, plays a crucial role in the energy harvesting process at the user end. This energy signal is not arbitrary but is instead crafted at the AP using a deterministic pseudo-random sequence. The characteristics of this sequence are well-defined, possessing a zero mean and a specific covariance matrix \mathbf{W}_e , which is mathematically expressed as $\mathbf{w}_e \sim \mathcal{CN}(\mathbf{0}, \mathbf{W}_e)$. The deterministic nature of this sequence, generated with a predefined seed, ensures that it is known to all users, facilitating efficient energy harvesting without compromising the integrity of the transmitted information. This approach underscores the intricate balance between information dissemination and energy transmission inherent in SWIPT systems.

For simplicity of the analysis, we adopt the assumption of a narrow-band block-fading propagation channel [146, 150]. This approach simplifies the channel's representation and is particularly suitable for scenarios where the channel's properties do not change significantly over the block of transmission. The signals received for ID and EH in such a channel

environment can be given as follows:

$$y_k^{\text{ID}} = \sum_{j \in \mathcal{K}} \mathbf{h}_{m,k}^H (b_{m,k} \mathbf{w}_j s_j + \mathbf{w}_e) + n_k^{\text{ID}}, \quad \forall k \in \mathcal{K}, \quad (3.2)$$

$$\mathbf{y}_k^{\text{EH}} = (\mathbf{I}_M - \text{diag}(\mathbf{b}_k)) \sum_{j \in \mathcal{K}} \mathbf{G}_k^H (\mathbf{w}_j s_j + \mathbf{w}_e) + \mathbf{n}_k^{\text{EH}}, \quad \forall k \in \mathcal{K}, \quad (3.3)$$

where the noise components n_k^{ID} and \mathbf{n}_k^{EH} are considered to be Additive White Gaussian Noise (AWGN) with properties reflecting circularly symmetric Gaussian distributions. Specifically, the noise term for the ID process is distributed as $n_k^{\text{ID}} \sim \mathcal{CN}(0, \sigma_k^{\text{ID}^2})$, indicating a zero-mean complex Gaussian distribution with variance $\sigma_k^{\text{ID}^2}$. Similarly, the noise affecting the EH process follows $\mathbf{n}_k^{\text{EH}} \sim \mathcal{CN}(\mathbf{0}, \sigma_k^{\text{EH}^2} \mathbf{I}_M)$, where \mathbf{I}_M is the identity matrix of size M , reflecting the dimensionality of the EH noise vector.

Within the framework of our model, we employ the generalized AS technique, which can distinguish between signals intended for information transfer and those allocated for power transfer. Through the application of generalized AS, we strategically partition the antennas at the receiver into two distinct groups. One set is dedicated exclusively to EH, absorbing the power transmitted by the AP to replenish the device's battery or to power other operations. Concurrently, the remaining antennas are tasked with wireless information processing, capturing and decoding the data transmitted from the AP. This dual-path strategy ensures that the system maximizes the utility of every antenna, optimizing both energy intake and information throughput [151]. In this optimized configuration, the achievable data rate for user k through a selected received antenna m is described by the following relationship:

$$R_{m,k}(b_{m,k}, \mathbf{w}_k) = \log_2 \left(1 + \frac{b_{m,k} |\mathbf{h}_{m,k}^H \mathbf{w}_k|^2}{\sigma_k^{\text{ID}^2} + I_{m,k}} \right), \quad \forall m \in \mathcal{M}, \forall k \in \mathcal{K}, \quad (3.4)$$

where the AWGN is considered at the k^{th} user with zero mean and variance $\sigma_k^{\text{ID}^2}$. The term $I_{m,k}$ in (3.4) indicates the multi-user interference and is given by:

$$I_{m,k} = \sum_{k' \neq k, k' \in \mathcal{K}} b_{m,k} |\mathbf{h}_{m,k}^H \mathbf{w}_{k'}|^2, \quad \forall m \in \mathcal{M}, \forall k \in \mathcal{K}. \quad (3.5)$$

We should note that the EH beams may cause interference in the data rate function in Equation (3.4). It is crucial to acknowledge that while EH beams are primarily aimed at powering devices, they may inadvertently interfere with the data rate calculations as specified in Equation (3.4). Nevertheless, the anticipated challenge posed by this interference is mitigated by the fact that the energy signals, being deterministic and known to the sensor users, allow for a proactive approach to interference management. Users are equipped to eliminate these energy signals as a form of interference, even prior to the decoding of the information-bearing signals, through the application of Successive Interference Cancellation (SIC) techniques [152]. This process indicates the adaptability and resilience of our system design, ensuring that data transmission efficiency is maintained despite the concurrent energy transfer.

For facilitating the presentation, we define $\mathbf{b}_k = [b_{1,k}, \dots, b_{M,k}]^T \in \mathbb{Z}^{1 \times M}$ as the vector representing the antenna selection within the optimization problem framework. Consequently, the achievable data rate of user k considering the antenna selection and the effects of potential

interference from EH beams, is expressed as:

$$R_k(\mathbf{b}_k, \mathbf{w}_k) = \sum_{m \in \mathcal{M}} R_{m,k}(b_{m,k}, \mathbf{w}_k), \quad \forall k \in \mathcal{K}. \quad (3.6)$$

Further enhancing our analytical framework, we introduce a novel performance metric dedicated to evaluating the efficiency of wireless power transfer. This new metric serves as a tool for assessing the effectiveness of our proposed system in optimizing the simultaneous delivery of information and power, thereby summarizing the core objectives of SWIPT technology in enhancing network performance and user experience. The new performance metric, $\mathcal{P}^{\text{eff}}(\mathbf{b}_k, \mathbf{w}_k, \mathbf{W}_e)$, for the wireless power transfer efficiency which is given by [136, 143]:

$$\mathcal{P}^{\text{eff}}(\mathbf{b}_k, \mathbf{w}_k, \mathbf{W}_e) = \frac{\sum_{k \in \mathcal{K}} P_{\text{NL}_k}^{\text{EH}}(\mathbf{b}_k, \mathbf{w}_k, \mathbf{W}_e)}{P_{\text{T}}(\mathbf{w}_k, \mathbf{W}_e)}. \quad (3.7)$$

The denominator of (3.7), $P_{\text{T}}(\mathbf{w}_k, \mathbf{W}_e)$, is the total power dissipated in the system in [Joule/Second] given by:

$$P_{\text{T}}(\mathbf{w}_k, \mathbf{W}_e) = \frac{\sum_{k \in \mathcal{K}} \|\mathbf{w}_k\|^2 + \text{Tr}(\mathbf{W}_e)}{\beta} + N_T P_{\text{ant}} + P_c, \quad (3.8)$$

where P_{ant} and P_c are the circuit power in each transmit antenna and fixed consumed power for baseband signal processing, respectively [153]. We note the first term in (3.8) is the so-called RF's transmit power consumption that is divided by $0 < \beta \leq 1$, the constant AP power amplifier efficiency. The numerator in (3.7), $P_{\text{NL}_k}^{\text{EH}}(\mathbf{b}_k, \mathbf{w}_k, \mathbf{W}_e)$ is the total harvested energy in the network topology. The harvesting is realized using the active EH antennas for each user. The total harvested energy is then given by [152, 143]:

$$P_{\text{NL}_k}^{\text{EH}}(\mathbf{b}_k, \mathbf{w}_k, \mathbf{W}_e) = \frac{[\Theta_k - \Omega_k \Delta_k]}{1 - \Delta_k}, \quad \forall k \in \mathcal{K}, \quad (3.9)$$

$$\Delta_k = \frac{1}{1 + \exp(\alpha_k \zeta_k)}, \quad \forall k \in \mathcal{K}, \quad (3.10)$$

$$\Theta_k = \frac{\Omega_k}{1 + \exp(-\alpha_k (P_{\text{L}_k}^{\text{EH}}(\mathbf{b}_k, \mathbf{w}_k, \mathbf{W}_e) - \zeta_k))}, \quad \forall k \in \mathcal{K}. \quad (3.11)$$

The constant Δ_k is introduced to guarantee a zero-input/zero-output response for EH [152]. In the traditional logistic function (3.11), the linear factor is given by

$$P_{\text{L}_k}^{\text{EH}}(\mathbf{b}_k, \mathbf{w}_k, \mathbf{W}_e) = \epsilon_k \text{Tr} \left(\sum_{j \in \mathcal{K}} \tilde{\mathbf{G}}_k^{\text{H}} (\mathbf{w}_j \mathbf{w}_j^H + \mathbf{W}_e) \tilde{\mathbf{G}}_k \right), \quad \forall k \in \mathcal{K}, \quad (3.12)$$

where

$$\tilde{\mathbf{G}}_k = (\mathbf{I} - \text{diag}(\mathbf{b}_k)) \mathbf{G}_k, \quad \forall k \in \mathcal{K}. \quad (3.13)$$

In the total linear received RF power formula (3.12), $0 < \epsilon_k < 1$ is introduced as the power conversion efficiency for the m^{th} active EH antenna of the k^{th} receiver. This parameter quantifies the efficiency with which the EH antenna converts the received RF power into usable electrical energy. The parameter Ω_k is defined as the maximum power that can be harvested by user k when the EH circuit becomes saturated. This saturation threshold prevents the EH circuit from being overloaded by excessive RF power, which could potentially

damage the circuit or reduce its efficiency. Furthermore, α_k and ζ_k are constant parameters that characterize the non-linear behavior of the EH process. These parameters can be precisely determined through the use of a curve-fitting tool, allowing for a more accurate modeling of the EH efficiency across different power levels. We should note that the contribution of the noise power to the $P_{NL_k}^{EH}(\mathbf{b}_k, \mathbf{w}_k, \mathbf{W}_e)$ formula can be neglected, as it is significantly lower than the main signal power, thereby having a minimal impact on the total energy harvested.

With these parameters and considerations in place, we proceed to formulate the main optimization problem, which focuses on beamforming design coupled with antenna selection within a generalized AS-based SWIPT framework, targeting a single-cell multi-user network configuration. The objective is to enhance the overall network performance by optimizing the allocation of antennas for EH and ID, in conjunction with the beamforming vectors. This optimization problem aims to balance the maximizing data transmission efficiency while ensuring optimal energy harvesting by the network's users. The formulation of this problem captures the intricacies of the generalized AS technique and its impact on the SWIPT system's performance, which is essential for advancing our understanding and capabilities in managing the complex interplay between information and power transfer in modern wireless networks. The optimization problem can be written as follows:

$$P_1 : \max_{\mathbf{b}_k, \mathbf{w}_k, \mathbf{W}_e} \mathcal{P}^{eff}(\mathbf{b}_k, \mathbf{w}_k, \mathbf{W}_e) \quad (3.14a)$$

$$\text{s.t.} : \sum_{k \in \mathcal{K}} \|\mathbf{w}_k\|^2 + \text{Tr}(\mathbf{W}_e) \leq p_{\max}, \quad (3.14b)$$

$$R_k(\mathbf{b}_k, \mathbf{w}_k) \geq R_{\min}, \quad \forall k \in \mathcal{K}, \quad (3.14c)$$

$$\sum_{m \in \mathcal{M}} b_{m,k} = 1, \quad \forall k \in \mathcal{K}, \quad (3.14d)$$

$$b_{m,k} \in \{0, 1\}, \quad \forall k \in \mathcal{K}, \forall m \in \mathcal{M}. \quad (3.14e)$$

In the optimization framework of our study, we formulate the optimization problem P_1 which is central to the deployment of a generalized AS-based SWIPT network in a single-cell multi-user setting. This optimization problem is defined by several key constraints that ensure the operational feasibility and efficiency of the system. The constraint (3.14b) limits the total transmit power of the AP that should not exceed its maximum threshold (p_{\max}). This limitation is vital for maintaining energy efficiency and adhering to regulatory power emission standards. Constraint (3.14c) guarantees a minimum data rate requirement, R_{\min} , for each user k . This guarantee ensures that all users receive a baseline QoS, critical for user satisfaction and system reliability. Constraint (3.14d) determines that each user utilizes only one antenna for ID, a rule that underpins the operational logic of the generalized AS technique by designating clear roles for each antenna at the user end. Finally, (3.14e) specifies the antenna selection variable takes only binary values. This binary nature reflects the decision-making process in antenna selection — whether an antenna is allocated for EH or ID.

Given these constraints, the optimization problem P_1 emerges as a Mixed-Integer Non-Linear

Programming (MINLP) challenge. The MINLP nature of the problem stems from the binary decision variables involved in antenna selection and the non-linear relationships encapsulated within the system's power and data rate equations. This complexity renders the problem generally intractable, necessitating innovative solution designs that can efficiently navigate the problem space [154]. Our objective is to develop a solution approach that not only adheres to the stipulated constraints but also optimizes the system's overall performance in terms of power efficiency.

3.4 A Two-layer Optimal Solution Design

Addressing the complex challenge presented by our optimization problem P_1 , we approach the solution through a strategic decomposition into two distinct but interrelated subproblems: antenna selection and beamforming. This decomposition is crucial for simplifying the problem's structure and focusing on specific aspects of the system's optimization individually, thus allowing for a more granular and effective solution strategy.

The antenna selection subproblem is tackled by prioritizing the identification of the antenna that offers the maximum channel gain. This step involves evaluating each antenna's performance to determine the most effective configuration for ID. The criterion for selection is straightforward; maximize the channel gain to ensure the highest QoS in terms of data rate for each user. Once the optimal antenna for ID is identified, the remaining antennas are allocated for EH tasks. Following the resolution of the antenna selection subproblem, attention shifts to the beamforming subproblem. This phase of the solution design is crucial for shaping the transmission strategy in a way that respects the predefined objective function and aligns with the system's goals of efficiency and reliability. The second subproblem is optimally solved via a two-layer iterative structure based on the sum-of-ratios programming.

The complexity of each subproblem necessitates a modular approach, especially given the limitations of traditional solution methodologies like the Dinkelbach method or the Charnes-Cooper transformation in handling sum-of-ratios objective functions [155, 156]. To circumvent these challenges, we employ a two-layer iterative structure specifically devised for sum-of-ratio programming (in the second subproblem). This two-phased approach, encompassing both antenna selection and beamforming optimization, underscores our comprehensive strategy for tackling the SWIPT system's optimization problem. By methodically addressing each component, we ensure a cohesive solution that not only meets the individual requirements of each user but also enhances the system's overall performance and efficiency. In what follows, we explain each step in detail.

3.4.1 Antenna Selection

Our approach initially focuses on the antenna selection process, assuming a scenario with predetermined transmit beamforming configurations. The main optimization problem, P_1 , clearly indicates that each user is to allocate precisely one antenna for ID purposes, with the remaining antennas dedicated to EH within this framework. This allocation strategy

is critical for optimizing the system's dual functionality—enhancing both data transmission quality and energy collection efficiency.

The main dilemma of the optimization problem P_1 lies in adhering to the stringent data rate QoS requirements imposed for each user. This necessitates a strategic selection of antennas, prioritizing those with the maximal channel gain for ID tasks. Such a selection criterion is pivotal, as the quality and reliability of information transmission directly hinge on the strength and clarity of the received signal. Hence, to ensure that every user's data rate demands are satisfactorily fulfilled, we employ the following principle: the antenna offering the supreme channel gain amongst all available antennas for a user is selected for ID. This principle can be encapsulated in the formula:

$$b_{m,k} = \begin{cases} 1, & \underset{m \in \mathcal{M}}{\operatorname{argmax}} \ h_{m,k}, \\ 0, & \text{otherwise,} \end{cases}, \quad \forall m \in \mathcal{M}, \forall k \in \mathcal{K}. \quad (3.15)$$

In essence, we assess the channel quality between the AP and all user's antennas via (3.15). This evaluation is crucial for identifying which antenna among the available options provides the highest channel quality, thereby determining the most suitable antenna for ID purposes. Following this determination, the remaining antennas are allocated for EH, optimizing the system's dual-function capabilities³.

By allocating the antenna with the best channel gain conditions exclusively for ID, rather than for EH, we consciously prioritize the transmission of information signals. This strategic decision underscores our commitment to ensuring the feasibility and practicality of our design policy. Such prioritization is key in SWIPT-enabled networks, where the efficient decoding of information signals is paramount to the network's functionality and user satisfaction.

It is important to highlight that the complexity associated with the antenna selection algorithm is not as daunting as it might seem. Despite the potential for a large solution space, the actual number of antennas present on a typical mobile receiver is relatively modest. Consequently, the process of selecting the optimal antenna for ID from a limited array does not introduce exponential complexity. This manageable scale ensures that the antenna selection procedure remains practical and implementable within the constraints of current technology and device capabilities, thus reinforcing the viability of our solution in real-world applications. Through this careful and strategic antenna allocation, we aim to enhance both the efficiency and effectiveness of SWIPT systems, ensuring that they can meet the demands of modern wireless communication networks.

3.4.2 Beamforming Design

Following the successful allocation of antennas for ID and EH purposes, our next step in our two-phased approach involves meticulously crafting the beamforming strategy for both information and energy signals. This step is crucial for enhancing the system's power efficiency, which is our defined novel metric for assessing the performance and sustainability of the SWIPT-aided network.

³Here, we assume that the users are sensor nodes. These nodes do not need to transmit with a high data rate and are more interested in EH.

To facilitate this process, we introduce beamforming matrices defined by $\mathbf{W}_k = \mathbf{w}_k \mathbf{w}_k^H$, where each matrix $\mathbf{W}_k \in \mathbb{H}^{N_T \times N_T}$ represents the beamforming strategy for the user k . Additionally, we define the channel matrix $\mathbf{H}_k = \mathbf{h}_{m,k} \mathbf{h}_{m,k}^H$, to encapsulate the channel characteristics between the AP and the m -th antenna of user k , which is critical for the beamforming design. For simplicity, we ignore the constant terms (P_{ant} and P_c) in the total power consumption model in (3.8). This simplification aids in focusing the optimization on the variables that we can influence directly through our design choices.

Employing Semi-Definite Programming (SDP), a powerful tool in optimization theory, we can reformulate the original problem P_1 into a more tractable format. SDP allows us to handle the quadratic nature of the beamforming matrices and the linear constraints of the system within a convex optimization framework. This approach simplifies the computational process and ensures that we can find a global optimum for the beamforming design problem, subject to the constraints and objectives defined. Thus, using SDP, the original optimization problem in P_1 can be reformulated as follows:

$$P_2 : \max_{\mathbf{W}_k, \mathbf{W}_e} \frac{\sum_{k \in \mathcal{K}} P_{\text{NL}_k}^{\text{EH}}(\mathbf{W}_k, \mathbf{W}_e)}{\sum_{k \in \mathcal{K}} \text{Tr}(\mathbf{W}_k) + \text{Tr}(\mathbf{W}_e)} \quad (3.16a)$$

$$\text{s.t.} : \sum_{k \in \mathcal{K}} \text{Tr}(\mathbf{W}_k) + \text{Tr}(\mathbf{W}_e) \leq p_{\max}, \forall k \in \mathcal{K}, \quad (3.16b)$$

$$\bar{R}_k(\mathbf{W}_k) \geq R_{\min}, \quad \forall k \in \mathcal{K}, \quad (3.16c)$$

$$\text{rank}(\mathbf{W}_k) \leq 1, \quad \forall k \in \mathcal{K}, \quad (3.16d)$$

$$\mathbf{W}_e \succeq 0. \quad (3.16e)$$

we adopt the SDP relaxation. This method involves a strategic simplification of the problem by omitting the rank-one constraint (3.16d), which traditionally ensures that the solution translates directly into a physically implementable beamforming vector. The removal of this constraint allows us to transform the beamforming design problem into a semi-definite programming problem, which is easier to solve using available optimization tools. Next, we handle the constraint (3.16c). In doing so, we restate this constraint as follows:

$$\frac{\text{Tr}(\mathbf{H}_k \mathbf{W}_k)}{\gamma_{\text{req}}} \geq \sum_{k' \neq k} \text{Tr}(\mathbf{H}_k \mathbf{W}_{k'}) + \sigma_k^2, \quad \forall k \in \mathcal{K}, \quad (3.17)$$

where

$$\gamma_{\text{req}} = 2^{R_{\min}} - 1. \quad (3.18)$$

A significant challenge in our optimization problem is the inherent non-convexity of the non-linear objective function, which complicates the process of finding a global optimum. However, by categorizing this function within the sum-of-ratio class of objective functions, we create an opportunity to go about this complexity more effectively. To accomplish this, we introduce a new slack variable, denoted as ϱ , effectively transforming the optimization problem. This introduction of ϱ allows us to reformulate the non-linear, non-convex objective

function into a format that is more amenable to optimization techniques commonly used for sum-of-ratios problems. To this end, we formulate the optimization problem as follows:

$$P_3 : \max_{\mathbf{W}_k, \mathbf{W}_e} \frac{\sum_{k \in \mathcal{K}} P_{NL_k}^{\text{EH}}(\mathbf{W}_k, \mathbf{W}_e)}{\varrho}, \quad (3.19a)$$

$$\text{s.t.} : \sum_{k \in \mathcal{K}} \text{Tr}(\mathbf{W}_k) + \text{Tr}(\mathbf{W}_e) \leq p_{\max}, \quad \forall k \in \mathcal{K}, \quad (3.19b)$$

$$\sum_{k \in \mathcal{K}} \text{Tr}(\mathbf{W}_k) + \text{Tr}(\mathbf{W}_e) = \varrho, \quad (3.19c)$$

$$\frac{\text{Tr}(\mathbf{H}_k \mathbf{W}_k)}{\gamma_{\text{req}}} \geq \sum_{k' \neq k} \text{Tr}(\mathbf{H}_k \mathbf{W}_{k'}) + \sigma_k^2, \quad \forall k \in \mathcal{K}, \quad (3.19d)$$

$$\mathbf{W}_e \succeq 0. \quad (3.19e)$$

In order to solve this optimization problem, we consider an iterative algorithm composed of two essential layers to refine the solution progressively. The initial layer is dedicated to determining the optimal configurations for the beamforming matrices, \mathbf{W}_k and \mathbf{W}_e , given a preset value of the slack variable, ϱ . The subsequent layer then undertakes the critical task of updating ϱ based on newly obtained beamforming matrices from the preceding stage. This dynamic adjustment ensures that the system continuously evolves towards achieving an optimal balance between transmission efficiency and energy utilization.

The challenge posed by the sum-of-ratio objective functions in the first layer of the optimization problem in (3.19) necessitates an innovative approach beyond traditional methods like Dinkelbach's algorithm, which proves inadequate for this context. This transformation simplifies the mathematical treatment of the problem and preserves the integrity of the optimal solution, facilitating a more straightforward path to achieving our optimization objectives. Through this layered algorithm, we systematically address the complexity of the optimization problem, ensuring a thorough and effective solution strategy that enhances the system's overall performance. Therefore, we find an equivalent subtractive form yielding the same optimal solution based on the following lemma from [154].

Lemma 1 [154]: For (3.19), there exist two vectors $\psi^* = [\psi^*, \dots, \psi_K^*]^T$ and $\beta^* = [\beta_1^*, \dots, \beta_K^*]^T$ in which \mathbf{W}_k^* and \mathbf{W}_e^* are the optimal solutions to the following optimization problem

$$\max_{\{\mathbf{W}_k^*, \mathbf{W}_e^*\} \in \mathcal{S}} \frac{1}{\varrho} \sum_{k \in \mathcal{K}} \psi_k^* \left[\Omega_k \left(1 - \Delta_k \Gamma_k \right) - \beta_k^* \left(\Gamma_k (1 - \Delta_k) \right) \right], \quad (3.20a)$$

where \mathcal{S} is the set belonging to the feasible solution of P_3 . In (3.20), the Γ_k term is:

$$\Gamma_k = 1 + \exp(-\alpha_k ((P_L^{\text{EH}}((\mathbf{W}_k, \mathbf{W}_e) - \zeta_k)). \quad \forall k \in \mathcal{K}. \quad (3.21)$$

Note that $\{\mathbf{W}_k^*, \mathbf{W}_e^*\}$ should satisfy the following equations

$$\Omega_k \left(1 - \Delta_k \Gamma_k \right) - \beta_k^* \left(\Gamma_k (1 - \Delta_k) \right) = 0, \quad \forall k \in \mathcal{K}, \quad (3.22)$$

$$\psi_k^* \left(\Gamma_k (1 - \Delta_k) \right) - 1 = 0, \quad \forall k \in \mathcal{K}. \quad (3.23)$$

Algorithm 1 Power Efficient Resource Allocation Algorithm for Beamforming in Multi-User SWIPT Networks

- 1: **Initialize**
iteration index of resource allocation policy $i = 1$,
limitation over two layer iteration of I_{\max}
define feasible set vector ϱ , and constant set $\{\alpha, \zeta, \Omega, \epsilon_k, \tau, \kappa\}$.
- 2: **repeat**
- 3: Set $\{\mathbf{W}_k^i, \mathbf{W}_e^i\} = \{\mathbf{W}_k^{s^*}, \mathbf{W}_e^{s^*}\}$.
- 4: Solve the inner-layer of (3.20) to update $\{\mathbf{W}_k^{i+1}, \mathbf{W}_e^{i+1}\}$.
- 5: Solve the outer-layer of (3.20) to update $\{\beta^{i+1}, \Psi^{i+1}\}$ regarding (3.28) and (3.29).
- 6: **until** $i = I_{\max}$
- 7: Update ϱ for the obtained $\{\mathbf{W}_k^{i+1}, \mathbf{W}_e^{i+1}\}$ via one dimensional search method.
- 8: **return** $\{\varrho, \mathbf{W}_k^{i+1}, \mathbf{W}_e^{i+1}\}$

Subproblem (3.20) can be solved with two-layer iterative structure including an inner and an outer layer. In the following, we describe these layers' functionality.

3.4.2.1 Inner Layer Solution

Delving deeper into the solution mechanism, within the inner layer of the algorithm, we define the following optimization problem denoted as P_4 . Given the specific assumptions regarding the parameters ψ and β based on *Lemma 1*, P_4 emerges as a convex problem. This convexity implies that the problem can be solved efficiently through standard optimization techniques, offering a pathway to achieving the desired optimization with relative ease and efficacy. Thus, the inner layer optimizing problem P_4 reads as:

$$P_4 : \max_{\mathbf{W}_k, \mathbf{W}_e, \lambda_k} \frac{1}{\varrho} \sum_{k \in \mathcal{K}} \psi_k \left[\Omega_k - \beta_k (1 + \exp(-\alpha_k(\lambda_k - \zeta_k))) \right], \quad (3.24a)$$

$$\text{s.t.} : \sum_{k \in \mathcal{K}} \text{Tr}(\mathbf{W}_k) + \text{Tr}(\mathbf{W}_e) \leq p_{\max}, \quad \forall k \in \mathcal{K}, \quad (3.24b)$$

$$\sum_{k \in \mathcal{K}} \text{Tr}(\mathbf{W}_k) + \text{Tr}(\mathbf{W}_e) = \varrho, \quad (3.24c)$$

$$\frac{\text{Tr}(\mathbf{H}_k \mathbf{W}_k)}{\gamma_{\text{req}}} \geq \sum_{k' \neq k} \text{Tr}(\mathbf{H}_k \mathbf{W}_{k'}) + \sigma_k^2, \quad \forall k \in \mathcal{K}, \quad (3.24d)$$

$$\lambda_k \leq \epsilon_k \text{Tr} \left(\sum_{j \in \mathcal{K}} \tilde{\mathbf{G}}_k^H (\mathbf{w}_j \mathbf{w}_j^H + \mathbf{W}_e) \tilde{\mathbf{G}}_k \right), \quad \forall k \in \mathcal{K}, \quad (3.24e)$$

$$\mathbf{W}_e \succeq 0, \quad (3.24f)$$

where λ_k is the auxiliary optimization variable.

3.4.2.2 Outer-Layer Solution

In the outer layer of our iterative optimization process, we employ a damped Newton method to obtain the values of ψ and β , which are pivotal for determining the optimal solution. To facilitate this process, we introduce specific functions for each user k , encompassing both the β and ψ parameters, denoted as $\phi_k(\beta_k)$ and $\phi_{K+k}(\psi_k)$, respectively. These functions are defined as follows, capturing the relationship between the parameters and the system's operational dynamics:

$$\phi_k(\beta_k) = \Omega_k \left(1 - \Delta_k \Gamma_k \right) - \beta_k^* \left(\Gamma_k (1 - \Delta_k) \right), \quad \forall k \in \mathcal{K}, \quad (3.25)$$

$$\phi_{K+k}(\psi_k) = \psi_k \left(\Gamma_k (1 - \Delta_k) \right) - 1, \quad \forall k \in \mathcal{K}, \quad (3.26)$$

where $k \in \{1, 2, \dots, K\}$. It has been shown in [152] that the optimal solution $\{\beta^*, \psi^*\}$ can be found if and only if:

$$\phi(\beta, \psi) = [\phi_1, \dots, \phi_{2K}]^T = 0. \quad (3.27)$$

This condition forms the basis for the iterative updates of ψ_k and β_k during each iteration of the outer layer. As a result the update rule at i -th iteration are given by:

$$\beta^{i+1} = \beta^i + \tau^i \eta_{1:K}^i, \quad (3.28)$$

$$\psi^{i+1} = \psi^i + \tau^i \eta_{K+1:2K}^i, \quad (3.29)$$

where

$$\eta = [\phi'(\beta, \psi)]^{-1} \phi(\beta, \psi), \quad (3.30)$$

in which $\phi'(\beta, \psi)$ is the Jacobian matrix of $\phi(\beta, \psi)$.

Furthermore, τ^i is the largest value of ε^i that should satisfy the following criterion

$$\|\phi(\psi^i + \varepsilon^i \eta_{K+1:2K}^i, \beta^i + \varepsilon^i \eta_{1:K}^i)\| \leq (1 - \kappa \varepsilon^i) \|\phi(\beta, \psi)\|, \quad (3.31)$$

where $i \in \{1, 2, \dots\}$, $\varepsilon^i \in (0, 1)$, and $\kappa \in (0, 1)$, moderating the adjustment's extent to ensure gradual and controlled convergence.

Since the optimization problem (3.20) is convex it lends itself to an efficient solution methodology. This approach is systematically outlined in the pseudo-code provided in **Algorithm 1**, which serves as a structured guide through the optimization process. A noteworthy aspect of this process is the implementation of a one-dimensional search over the variable ϱ , necessitating the evaluation of problem P_4 across a spectrum of ϱ values⁴.

The optimization problem in question falls within the domain of convex SDP, a category well-accommodated by established numerical algorithms for convex optimization tasks, such as the interior point method. These algorithms are known for their robustness and efficiency, offering a reliable means to navigate and solve convex SDP challenges. An optimistic development in our beamforming design procedure is the existence of a rank-one solution, a feature that significantly rationalizes the optimization process. As affirmed by existing

⁴The upper bound for ϱ is p_{\max} . This means the search area of the problem is inherently restricted by the system's maximum transmit power, p_{\max} , which is the feasible domain for ϱ .

Table 3.2: Overview of Simulation Parameters for a Multi-Antenna, Multi-User SWIPT Network.

Parameter	Value
Number of sensor users (K)	8
Maximum cell coverage (d_{\max})	20 meters
Number of AP antennas (N_T)	4
Number of user antennas (M)	3
AP antenna power consumption (P_{ant})	30 dBm
Static circuit power consumption (P_c)	40 dBm
Central carrier frequency	3 GHz
Number of subcarriers (N)	16
Bandwidth of each subcarrier	180 kiloHertz (kHz)
Background noise (σ^2)	-120 dBm
Rician factor (ρ)	3 deciBel (dB)
Path loss exponent (α)	2.8
Standard deviation of log-normal shadowing	8 dB
Power conversion efficiency (ϵ)	0.3
Power amplifier efficiency (β)	0.2
Target transmission rate (γ_{req})	10 dB

research [152], the presence of a rank-one solution validates the feasibility of achieving an optimal beamforming configuration, thereby enhancing the efficacy of the proposed SDP relaxation approach. This facet of the solution simplifies the practical application of beamforming strategies and accentuates the efficiency of utilizing SDP relaxation techniques to optimize system configurations, ultimately contributing to the goal of improving power efficiency in SWIPT-aided system.

3.5 Simulation Results

This section presents the system performance through simulation results, focusing on the power efficiency of antenna switching and beamforming design in a multi-antenna, multi-user SWIPT system (cf., Table 3.2). In evaluating the achievable power efficiency of the proposed scheme, eight sensor users, $K = 8$, are uniformly distributed within a single cell, where the maximum coverage of the cell is $d_{\max} = 20$ meters. The AP and the sensor users are equipped with four ($N_T = 4$) and three ($M = 3$) antennas, respectively, facilitating robust communication and energy harvesting capabilities. Key parameters include the AP antenna power consumption ($P_{\text{ant}} = 30$ dBm) and static circuit power consumption ($P_c = 40$ dBm).

The numerical simulations are conducted under the assumption of a flat fading channel with a central carrier frequency set at 3 GHz, reflecting a realistic communication environment. The system utilizes $N = 16$ subcarriers, each with a bandwidth of 180 kHz, to efficiently manage the available spectral resources. The receivers' background noise across all antennas is uniformly modeled as $|\sigma_k^{\text{ID}}|^2 = \sigma_k^2 = \sigma^2 = -120$ dBm, indicative of the operational challenges in low-signal environments. Given the proximity of users to the transmitter, a Line-of-Sight (LoS) communication channel is presumed, and a small-scale fading channel is modeled as

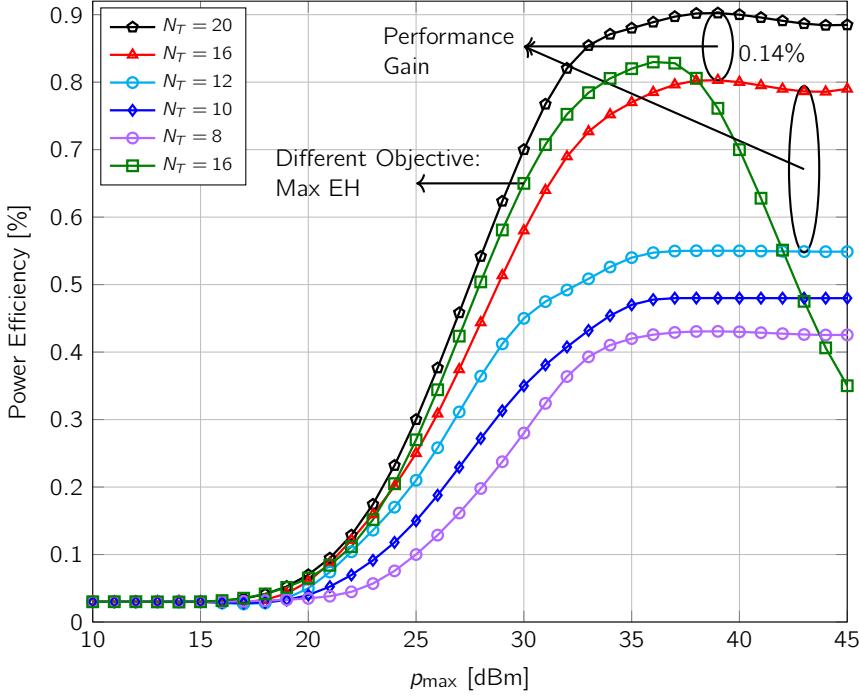


Figure 3.3: Power efficiency versus maximum allowed transmit power in the downlink of SWIPT-aided multi-user single-cell network.

Rician fading with Rician factor $\rho = 3$ dB. This small-scale Rician flat fading channel gains incorporates both a distance-dependent path loss component and a log-normal shadowing component with a standard deviation of 8 dB, where the path loss exponent is equal to $\alpha = 2.8$ [151]. The simulations account for the power conversion efficiency, $\epsilon_k = \epsilon = 0.3$, of all active EH antennas, alongside the power amplifier efficiency, $\beta = 0.2$, of the AP, which are critical parameters influencing the system's energy efficiency and power efficiency metrics and sustainability. The target transmission rate for ensuring satisfactory QoS for each user is set at $\gamma_{\text{req}} = 10$ dB, a benchmark for evaluating the efficacy of the proposed scheme.

To ensure the reliability and robustness of our simulation results, we employ Monte Carlo simulations, generating numerous random realizations of the channel gains. This approach allows us to compute the average EE across various scenarios, offering a comprehensive assessment of the proposed scheme's performance under diverse conditions [151]. Through these simulations, we aim to demonstrate the viability and advantages of the proposed power-efficient AS and beamforming design in enhancing the operational efficiency of multi-antenna, multi-user SWIPT systems.

Figure 3.3 illustrates the impact of enhancing the maximum allowable power budget on the network's power efficiency, revealing a non-linear relationship. Particularly notable is the pronounced effect observed at higher p_{max} values, where the power efficiency dynamics exhibit distinct characteristics across different power budget ranges. In the lower spectrum, between 5 dBm and 20 dBm of p_{max} , there is a negligible alteration in power efficiency,

indicating a phase of relative stability. The scenario shifts as the power budget spans from 20 dBm to 35 dBm, during which the power efficiency experiences a modest increase for systems with a fewer number of antennas, and a more pronounced surge for configurations with a larger antenna array. This distinction underscores the role of antenna count in influencing efficiency gains. A further escalation of p_{\max} beyond 35 dBm up to 45 dBm ushers in a saturation phase for power efficiency, signifying a plateau where additional increases in the power budget cease to yield significant efficiency improvements. This plateau effect is attributed to the dominance of fixed circuit power consumption in the lower transmit power regions, which initially allows for a gradual uplift in efficiency rates. As the AP's transmit power budget expands, surpassing a certain threshold, the relative contribution of RF's transmit power consumption begins to overshadow that of the fixed circuit power, rendering the system increasingly sensitive to increments in the power budget.

Moreover, this figure, Fig. 3.3, also elucidates the beneficial impact of increasing the number of transmit antennas (N_T) on augmenting effective power efficiency. This enhancement is logically anticipated, considering that power efficiency exhibits a quasi-linear dependency on the transmission parameters, including both information and energy beamforming vectors. For a comparative analysis, we compare our proposed optimization algorithm against a baseline scheme focused on EH maximization (Max EH), wherein the objective is to maximize the numerator of the power efficiency metric (cf., (3.7)) through a similar optimization approach as in **Algorithm 1**. The comparative results distinctly demonstrate the superiority of our proposed methodology, which, unlike the baseline, also integrates the minimization of total power consumption into the efficiency maximization process. Additionally, it is observed that while exclusively targeting EH maximization contributes to a rise in power efficiency for lower to moderate p_{\max} values, an inverse trend is noted at higher power budget levels. At this juncture, achieving maximum EH leads to an increase in the total transmitted power, which, in turn, elevates the denominator of the power efficiency equation, culminating in a marked efficiency downturn. This phenomenon highlights the critical balance between transmitted power and energy harvesting in optimizing power efficiency, underscoring the nuanced interplay of system parameters in achieving optimal network performance.

In the domain of wireless communication systems, particularly within the context of SWIPT systems, there is a fundamental tradeoff between power efficiency and EE. This behavior arises because the objectives of maximizing power efficiency and EE inherently conflict, especially when considering the balance between data rate optimization and the minimization of power consumption. While power efficiency focuses on optimizing the utility of power in transmitting data, EE aims to achieve the highest data rate relative to the total power expended by the network. The crux of this tradeoff lies in the fact that strategies aimed at enhancing power efficiency often do so at the expense of overall energy consumption, and vice versa. Maximizing power efficiency typically involves optimizing the transmission power to enhance the signal's clarity and reach, which can lead to increased power consumption. On the other hand, maximizing EE necessitates minimizing power consumption while still achieving satisfactory data transmission rates, which may not always align with the strategies that prioritize power efficiency. To explore and substantiate this tradeoff further, we embark on a detailed analysis by defining EE mathematically as the ratio of the achievable data rate (3.6) to the total network's power consumption (3.8) as follows:

$$\mathcal{E}^{\text{eff}}(\mathbf{b}_k, \mathbf{w}_k, \mathbf{W}_e) = \frac{\sum_{k \in \mathcal{K}} R_k(\mathbf{b}_k, \mathbf{w}_k)}{P_T(\mathbf{w}_k, \mathbf{W}_e)}. \quad (3.32)$$

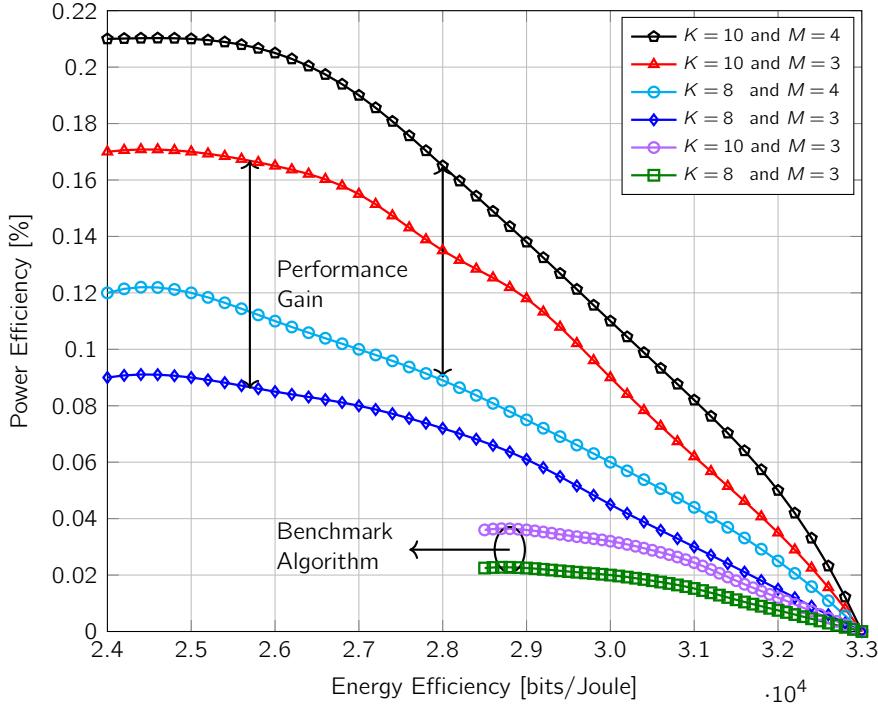


Figure 3.4: System performance tradeoff between power efficiency and EE for $p_{\max} = 40$ dBm in the downlink of SWIPT-aided multi-user single-cell network.

This ratio encloses the essence of EE by quantifying the data rate that can be achieved per unit of power consumed by the network, offering a clear metric for evaluating the efficiency of resource allocation policies in the system. Through this definition and subsequent analysis, we aim to discuss the trade-off between optimizing power efficiency and maximizing EE, highlighting the considerations that support resource allocation decisions in SWIPT systems.

As shown in Fig. 3.4, a tradeoff between power and energy efficiencies is depicted, following the maximization of EE as defined in (3.32). This graphical representation demonstrates that power efficiency (EE) exhibits a monotonically decreasing trend as EE (power efficiency) increases, highlighting the inherent balance that must be struck between these two efficiencies. A notable observation from the figure is the positive impact of augmenting the number of receiver antennas on power efficiency. This improvement is attributable to the enhanced capability of the network to harvest energy, thereby strengthening its power efficiency. Moreover, the dynamics introduced by varying the number of sensor users within the network are evident. An increase in sensor users does not merely expand the performance tradeoff gap between power and energy efficiencies; it also significantly boosts power efficiency for a given EE level. This enhancement stems from the increased collective energy-harvesting capacity brought about by a larger pool of sensor users, effectively enabling more efficient utilization of the power emitted by the AP⁵. For comparison, we also plot the tradeoff region for the case the energy signal (\mathbf{W}_E) is set to zero, and maximum ratio transmission is adopted to

⁵This is because more sensor users mean more harvesting antennas, i.e., more of the emitted power from the AP can be harvested when more receivers (more EH antennas) partake in the energy harvesting.

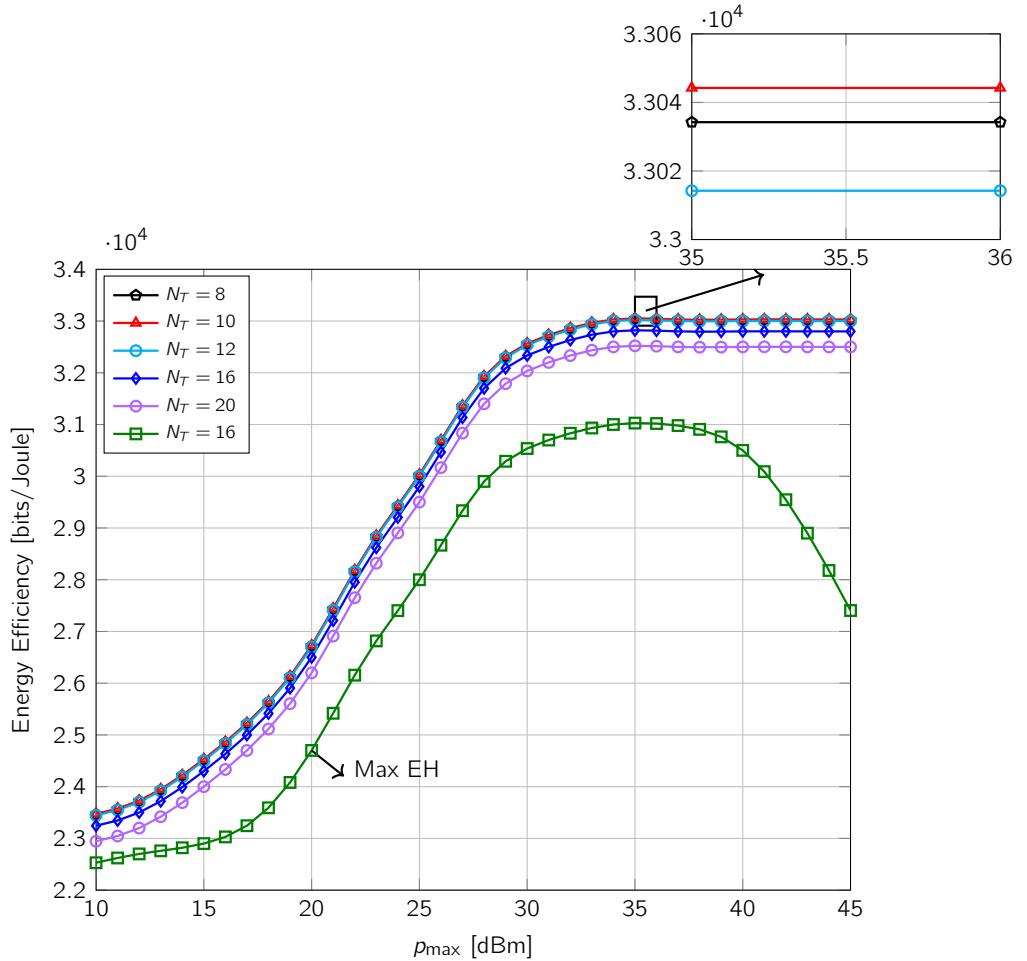


Figure 3.5: *EE versus maximum allowed transmit power in the downlink of SWIPT-aided multi-user single-cell network.*

optimize the information beamforming as a benchmark algorithm. This benchmark algorithm serves as a reference point, highlighting the benefits of the proposed approach in harnessing energy signals for improved network performance.

Shifting the focus to the last figure, Fig. 3.5, an exploration into the effects of escalating the power budget on EE reveals a monotonically increasing trend up to a saturation point around $p_{\max} \approx 30$ dBm. Beyond this threshold, EE plateaus, indicating that further amplification of p_{\max} does not translate into proportional EE gains. This saturation is primarily due to the escalating interference power, which adversely affects the quality of the received ID signal, thereby capping the achievable data rate. Additionally, while an increase in the number of transmit antennas (N_T) markedly enhances power efficiency, its influence on EE is relatively subdued. The logarithmic relationship between the data rate function and N_T implies that the EE gains attainable through additional transmit antennas are marginal compared to the

linear escalation of circuit power costs associated with higher N_T . Consequently, a significant expansion in N_T may not yield commensurate benefits for ID and is less effective for EH enhancement. Furthermore, Fig. 3.5 underscores the superiority of the proposed algorithm over the baseline Maximum EH (Max EH) scheme. This comparison reveals the strategic advantage of minimizing total power consumption alongside EE maximization, revealing the proposed algorithm's efficacy in achieving optimal system performance.

3.6 Conclusion

To obtain a feasible solution, an optimization problem with a transformed objective function was designed based on an iterative algorithm which yields a locally optimal solution. In particular, the antenna selection problem was solved based on maximum channel gain across all antennas. The second subproblem was solved based on a two-layer method. Simulation results revealed the superiority of the generalized AS scheme by demonstrating a good balance of improvement in terms of power and EE.

In this chapter, we introduced an innovative optimization framework tailored for a MIMO-OFDM network that employs generalized AS-based receivers, integrating the principles of SWIPT. This new framework takes into account a realistic, non-linear power model for EH, setting its sights on enhancing a new metric in wireless communications: power efficiency. The optimization problem we presented is characterized by the simultaneous consideration of antenna selection and beamforming strategies, a task rendered complex due to its non-convex, non-linear nature, and the inclusion of binary variables. These attributes collectively contribute to the problem's inherent difficulty.

To navigate through the complexities of this optimization problem non-convexity and arrive at a practically viable solution, we crafted an approach that modifies the original objective function. This approach hinges on an iterative algorithm meticulously designed to converge towards a locally optimal solution. The resolution of the antenna selection subproblem, prioritizing maximum channel gain across available antennas, marked the initial phase of our solution strategy. Subsequently, a two-layer method was applied to address the beamforming subproblem, further refining the solution.

Our simulations underscore the efficacy of the generalized AS scheme, shedding light on its capability to achieve a commendable synergy between power efficiency and EE. The results unequivocally illustrate the advantages of this scheme, demonstrating notable improvements in both power and EE metrics. Through this comprehensive investigation, the proposed optimization framework not only addresses the technical challenges associated with SWIPT-enabled MIMO-OFDM networks but also paves the way for significant advancements in the domain of wireless communication, particularly in optimizing the dual objectives of efficient power usage and effective energy harvesting.

In this chapter, we laid the groundwork for future discussions by establishing a fundamental framework that underpins the rest of this work. The principles, approaches, and solutions presented form a crucial basis for exploring advanced topics, notably Intelligent Reflecting Surfaces (IRS), which will be the focal point of the forthcoming chapters. Our exploration of

IRS in the next chapter aims to delve deeper into its applications and implications, building upon the foundational knowledge and insights gathered here.

Chapter 4

Energy Efficient Resource Allocation in IRS Networks

THIS chapter formally introduces the concept of Intelligent Reflecting Surfaces (IRS) and examines their advanced technical capabilities within the context of a multi-user Multiple-Input Single-Output (MISO) system. Central to this investigation is a system specifically designed to enhance Ultra-reliable Low-Latency Communications (URLLCs). It incorporates a multi-antenna Access Point (AP) that effectively transmits information symbols to a group of URLLC users, carefully integrating short packet transmission techniques to address the critical need for reduced latency in wireless communications. A key goal of this study is to minimize the total system's transmission power by simultaneously optimizing both active and passive beamformers at the AP and the IRS, respectively. An efficient algorithm based on Alternating Optimization (AO) principles is designed to tackle the main optimization problem through a step-by-step iterative approach.

The development of the algorithm begins with the application of the Difference of Convex (DC) functions combined with Successive Convex Approximation (SCA) techniques to find a near-optimal solution for the AP's active beamformer. This step is followed by the adoption of a penalty-based strategy, complemented by SCA, to effectively manage the unit-modulus constraints at the IRS. This two-pronged approach to optimization not only highlights the delicate interplay between active and passive beamforming but also proposes a specific objective aimed at improving the convergence rate of our algorithm. To validate the effectiveness and efficiency of the proposed solution, the chapter includes a series of simulation studies. These simulations benchmark the performance of our algorithm against several baseline models, providing a solid empirical foundation for its superiority in optimizing URLLC-enabled IRS systems. This thorough analysis sheds light on the potential of IRS to transform the landscape of wireless communications, paving the way for future research in this emerging domain.

This chapter is based on:

J. Jalali, A. Rezaei, A. Khalili and J. Famaey, "Power-efficient Joint Resource Allocation and Decoding Error Probability for Multiuser Downlink MISO with Finite Block Length Codes", in *2022 25th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, Herning, Denmark, Oct. 2022, pp. 232–237. <https://doi.org/10.1109/WPMC55625.2022.10014778>

4.1 Introduction

The exploration of Intelligent Reflecting Surfaces (IRS) has emerged as a groundbreaking area of research, aiming to significantly enhance the spectral and energy efficiencies of future communication networks through its straightforward deployment strategies [157]. The IRS, characterized as a planar meta-surface equipped with a large number of passive reflecting elements¹, possesses the unique ability to manipulate the radio propagation environment. This is achieved by dynamically adjusting the amplitudes and phases of incoming signals, thereby facilitating a more controlled and efficient communication pathway. Notably, the IRS operates in a Full-Duplex (FD) mode, achieving this without the need for active Radio Frequency (RF) chains for signal transmission and reception, nor for mechanisms to cancel self-interference. Traditional FD systems require complex and costly hardware to manage self-interference, but the IRS leverages passive elements to reflect and modulate incident electromagnetic waves, enabling simultaneous transmission and reception without generating self-interference. This intelligent manipulation of the phase and amplitude of incoming signals reduces complexity, power consumption, and system costs, making IRS a highly cost-effective solution for Beyond fifth-Generation (B5G) communication systems. By supporting efficient FD communication, IRS enhances data rates, addressing the demands of B5G and future wireless networks. This capability positions IRS technology as a pivotal component in the evolution of next-generation communication infrastructures, promoting versatile, scalable, and energy efficient wireless networks [158].

The synergy between active beamforming techniques at the Base Station (BS) and passive beamforming at the IRS opens new avenues for optimizing the Spectral Efficiency (SE) and enhancing the network's overall data throughput, as demonstrated in previous studies [159]. Moreover, the integration of IRS with technologies like wireless power transfer and simultaneous information and power transfer (SWIPT) stands out as a pivotal strategy for fostering green communication. Such collaborations have been shown to significantly improve the network's Energy Efficiency (EE) by judiciously optimizing both the phase shifts at the IRS and the active beamforming strategies at the transmitter [136]. Research into IRS-aided Multi-Input Single-Output (MISO) systems has further highlighted the potential of IRS technology in maximizing SE through the use of sophisticated algorithms, such as the branch-and-bound method to achieve globally optimal solutions for phase shifts and active beamforming at the IRS and AP, respectively [160]. Moreover, the targeted optimization of the weighted sum-rate maximization problem underscores the benefits of designing coordinated active and passive beamforming strategies at both the BS and IRS [83], showcasing the IRS's critical role in shaping the future of wireless communication networks.

On the other hand, Ultra-Reliable Low-Latency Communication (URLLC) represents a cornerstone of B5G wireless systems, aimed at addressing the needs for rapid data transmission and minimal delay in critical applications such as healthcare, autonomous driving, and the tactical Internet [161, 162]. The strict requirements of URLLC, including short packet transmission and ultra-low latency, necessitate a reevaluation of traditional communication theories, notably the conventional Shannon capacity formula, which falls short under the URLLC regime due to its incompatibility with the short packet paradigm [163]. Innovative

¹IRS can also have active elements. This feature, known as active IRS in the literature, will be the subject of Chapter 7.

approaches to resource allocation within URLLC systems have been explored, with [164] achieving a global optimum in bandwidth, power allocation, and antenna configuration to significantly reduce the total power consumption across DownLink (DL) and UpLink (UL) channels. Besides, the authors in [165] designed the active BS beamforming vectors to maximize the sum data rate performance of a Multiple-Input Single-Output (MISO) orthogonal frequency division multiple access (OFDMA)-URLLC system.

Further advancing the performance of URLLC systems, the integration of IRS offers a promising avenue to tackle the challenges of latency. Studies such as [166] and [167] have highlighted the potential of IRS in enhancing the data rate and reducing latency through joint optimization strategies involving active beamforming at the BS and adaptive phase shifts at the IRS. Specifically, [166] focused on optimizing the weighted sum data rate in an IRS-assisted OFDMA-URLLC system, while [167] targeted latency minimization in an IRS-supported mobile edge computing framework by coordinating edge computing resources, computation offloading, and beamforming techniques. The exploration of IRS in URLLC contexts extends to evaluating the average Decoding Error Probability (DEP) and achievable data rates, as conducted in [168], indicating a substantial improvement in system performance. Furthermore, the concept of user grouping has been leveraged to optimize latency across IRS-enhanced networks catering to URLLC demands, as demonstrated in [90], thereby underscoring the transformative impact of IRS technology in fulfilling the rigorous requirements of URLLC systems.

While significant progress has been made in integrating IRS within the domain of URLLC systems, the existing literature [166, 167, 168, 90] still does not fully capture the profound impact of IRS technology, especially in terms of DEP and the management of average traffic loads. The IRS's capability to significantly enhance the network's Quality of Service (QoS) invites further exploration into its deployment in high-stakes scenarios, which require not just enhanced data rates but also an elevated Signal-to-Interference-Plus-Noise Ratio (SINR). This underlines the necessity for our research, which aims to develop a novel resource allocation algorithm tailored for a DL MISO URLLC framework that leverages IRS technology. Our approach involves employing a multi-antenna AP that communicates with multiple single-antenna URLLC receivers, facilitated by a dynamically reconfigurable IRS. This setup is poised to bridge the identified research gaps by focusing on minimizing the overall transmission power, thus paving the way for a deeper understanding of the systemic design intricacies. Our study contributes valuable insights into the optimization of URLLC-enabled IRS systems and underscores the transformative potential of the IRS in enhancing URLLC services within increasingly complex and demanding operational contexts. Through this investigation, we aspire to unlock new possibilities for IRS technology, thereby contributing to the evolution of next-generation wireless communication systems.

In this chapter, we delineate our primary contributions toward enhancing the efficiency and reliability of URLLC systems through the strategic deployment of the IRS. The highlights of our work are summarized as follows:

- Central to our study is the objective to 'reduce the system's overall transmission power.' This is achieved by designing a comprehensive joint optimization problem of both the active and passive beamforming strategies employed at AP and IRS, respectively. Furthermore, we aim to optimize the DEP while ensuring adherence to the

minimum data rate requirements of each URLLC user. A novel aspect of our approach involves modeling the URLLC user traffic, especially those with finite blocklength data rates, through a chance constraint formulation. This methodology facilitates the network's ability to properly allocate resources, thereby efficiently managing the collective traffic load.

- To tackle the complexities of the optimization problem formulated, we exploit an Alternating Optimization (AO) resource allocation algorithm. This algorithm allows for an iterative solution process wherein we first establish a lower bound for the SINR for the active beamformers at the AP. Subsequently, we employ the Difference of Convex (DC) functions and Successive Convex Approximation (SCA) technique to derive a near-optimal solution. A penalty-based methodology is then utilized in conjunction with the SCA technique for the passive beamformers at the IRS, effectively addressing the unit-modulus constraints. Additionally, we introduce a specific objective aimed at devising more effective phase shifts, thus ensuring improved convergence of the optimization process.
- Through simulation studies, our findings underline the substantial benefits of incorporating an IRS alongside a multi-antenna AP within URLLC systems. Notably, the implementation of IRS technology contributes to notable gains in system performance, particularly in achieving lower latency and higher reliability. Moreover, our results indicate the energy efficiency of utilizing IRS technology in comparison to the alternative of equipping the AP with multiple antennas. This supports the potential of using the IRS to facilitate more sustainable and efficient URLLC communication infrastructures.

These contributions collectively demonstrate the innovative strides our work makes in harnessing IRS technology to elevate the operational efficacy of URLLC systems, setting a new benchmark for future research in the field.

The rest of the chapter is structured as follows. Section 4.2 begins with a system model analysis of energy-efficient resource allocation in an IRS-assisted URLLC network. We proceed to formulate the problem in Section 4.3. In Section 4.4, we present a solution for the proposed scheme. Section 4.5 validates our theoretical models through extensive simulations, showcasing the practical feasibility and advantages of our proposed method. Finally, Section 4.6 concludes the chapter, summarizing our findings and considering their implications for the advancement of IRS systems.

Notations: Matrices and vectors are denoted by boldface capital letters \mathbf{A} and lower case letters \mathbf{a} , respectively. For a square matrix \mathbf{A} , \mathbf{A}^T , \mathbf{A}^H , $\text{rank}(\mathbf{A})$, $\text{Tr}(\mathbf{A})$, and $\|\mathbf{A}\|_*$ are transpose, Hermitian conjugate transpose, rank of a matrix, trace, norm of a matrix, respectively. $\mathbf{A} \succeq \mathbf{0}$ shows a positive semidefinite matrix. \mathbf{I}_N denotes the N -by- N identity matrix. $\text{diag}(\cdot)$ is the diagonalization operation. $\text{diag}(\mathbf{A})$ indicates a vector whose elements are extracted from the main diagonal elements of matrix \mathbf{A} . The absolute value of a complex scalar, and the Euclidean norm of a complex vector are expressed by $|\cdot|$ and $\|\cdot\|$, respectively. $\sim \mathcal{CN}(\boldsymbol{\mu}, \mathbf{C})$ denotes the distribution of a Circularly Symmetric Complex Gaussian (CSCG) random vector with mean $\boldsymbol{\mu}$ and covariance matrix \mathbf{C} . The largest eigenvalue of matrix \mathbf{X} is denoted by $\lambda_{\max}(\mathbf{X})$. $Q^{-1}(\cdot)$ stands for the inverse of the Gaussian Q-function. $\Pr(X > a)$ denotes the probability that the random variable X assumes a particular value strictly greater

than a . Finally, $\mathbb{C}^{M \times N}$ represents an $M \times N$ dimensional complex matrix and $\nabla_{\mathbf{x}}$ expresses the gradient vector with respect to \mathbf{x} .

4.2 System Model of an IRS-assisted URLLC Network

In this study, we consider the architecture of a DL MISO communication system, incorporating an IRS comprised of N elements, an AP equipped with M antennas, and K users each possessing a single antenna. The set of users is denoted by $\mathcal{K} = \{1, \dots, K\}$, as illustrated in Fig. 4.1. For each user k within this system, a designated number of L_k information bits are allocated. These bits are subsequently encoded by the AP into a codeword consisting of m_d symbols, represented as $z_k[l]$, where l spans the set $\mathcal{L} = \{1, 2, \dots, m_d\}$. Following the encoding process at the AP, the formulated transmit signal intended for transmission can be mathematically represented by the equation:

$$\mathbf{s}[l] = \sum_{k \in \mathcal{K}} \mathbf{w}_k z_k[l], \quad l \in \mathcal{L}, \quad (4.1)$$

wherein $\mathbf{w}_k \in \mathbb{C}^{M \times 1}$ is the dedicated beamforming vector allocated for user k . This beamforming strategy is pivotal, as it directly influences the signal's directionality and strength towards each user, thereby optimizing the communication efficacy within the system. In the context of this DL MISO system, it is assumed that the channel links exhibit time-invariance and also belong to the category of slow fading channel models [169]. This assumption is critical for the stability and predictability of the communication channel, ensuring consistent performance over the communication duration. The system explicitly models the baseband equivalent channel responses to include the transmission pathways effectively: from the AP to the IRS, denoted by $\mathbf{H} \in \mathbb{C}^{N \times M}$, from the IRS to user k , represented by $\mathbf{h}_k^{\text{IU}} \in \mathbb{C}^{N \times 1}$, and directly from the AP to user k , indicated by $\mathbf{h}_k^{\text{AU}} \in \mathbb{C}^{M \times 1}$. These channel models are foundational to understanding and optimizing the interaction between the transmitted signals, the reflective IRS, and the receiving users, enabling a comprehensive analysis of the system's overall performance and efficiency. Moreover, it is assumed that the Channel State Information (CSI) and the delay requirements of all users are perfectly known at the AP (see [136, 158, 159, 165])².

Following the establishment of the transmit signal at the AP, it is important to understand the mechanics of signal reflection and reception within this system. Let's define the reflection-coefficients matrix at the IRS as:

$$\Theta = \text{diag}(\beta_1 e^{j\alpha_1}, \beta_2 e^{j\alpha_2}, \dots, \beta_N e^{j\alpha_N}), \quad (4.2)$$

where $\beta_n \in [0, 1]$ represents the reflection amplitude, and $\alpha_n \in (0, 2\pi]$, $\forall n \in \mathcal{N} \in \{1, \dots, N\}$, indicates the phase shift of the n -th reflection coefficient at the IRS³. The overall equivalent channel link, considering both the direct path and the reflected path via the IRS, for user k is expressed as:

$$\mathbf{h}_k^H \triangleq (\mathbf{h}_k^{\text{IU}})^H \Theta \mathbf{H} + (\mathbf{h}_k^{\text{AU}})^H, \quad \forall k \in \mathcal{K}. \quad (4.3)$$

²The results in this chapter serve as theoretical performance upper bounds for the URLLC-enable IRS systems with imperfect CSI in practice [170, 171].

³For reflection efficiency maximization, the amplitudes of all IRS elements are assumed to be one [158] i.e., $\beta_n = 1$, $\forall n \in \mathcal{N}$.

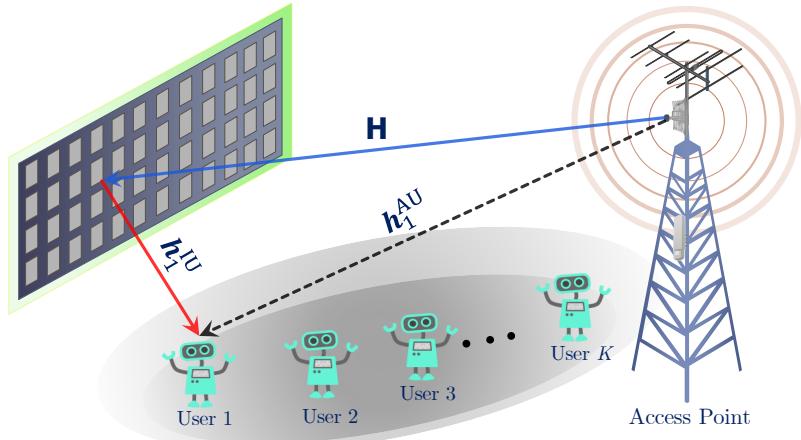


Figure 4.1: Energy-Efficient Resource Allocation for a multi-user DL MISO URLLC-enabled IRS system.

Defining the equivalent channel link brings us to the formulation of the received signal at each user k :

$$y_k[l] = \mathbf{h}_k^H \mathbf{s}[l] + n_k[l] \stackrel{\Delta}{=} \sum_{k \in \mathcal{K}} \mathbf{h}_k^H \mathbf{w}_k z_k[l] + n_k[l], \quad \forall k \in \mathcal{K}, \forall k \in \mathcal{L}, \quad (4.4)$$

where $n_k[l] \sim \mathcal{CN}(0, \sigma_k^2)$ symbolizes the Additive White Gaussian Noise (AWGN) at user k , characterized by a mean of zero and variance σ_k^2 . Consequently, the SINR experienced by user k can be expressed as:

$$\gamma_k = \frac{|\mathbf{h}_k^H \mathbf{w}_k|^2}{\sum_{i \neq k, i \in \mathcal{K}} |\mathbf{h}_k^H \mathbf{w}_i|^2 + \sigma_k^2}, \quad \forall k \in \mathcal{K}. \quad (4.5)$$

It should be noted that in URLLC systems, the data blocks must be finite and have a short length to guarantee low-latency and high-reliability wireless communication. This requirement stems from the essential goal of URLLC to minimize latency while maximizing the reliability of transmissions, a balance critical for applications where even minimal delays or errors could lead to significant consequences. Reflecting this necessity, a precise approximation of the achievable data rate for each user within such systems is paramount. The formula for this approximation, as detailed in [163], serves as a cornerstone for evaluating the efficiency and effectiveness of URLLC systems. This approximation takes into account the finite blocklength regime, where traditional approaches to calculating channel capacity, assuming infinite blocklengths, fall short of providing accurate or useful predictions for URLLC scenarios. By incorporating these considerations, URLLC systems are better positioned to meet the requirements of high reliability and low latency, ensuring that wireless communication remains both robust and agile in environments where performance and speed are non-negotiable. The precise approximation for the achievable data rate of each user is

given by:

$$R_k(\epsilon_k, \mathbf{w}_k, \Theta) = F_k(\mathbf{w}_k, \Theta) - G_k(\epsilon_k, \mathbf{w}_k, \Theta), \quad \forall k \in \mathcal{K}, \quad (4.6)$$

where

$$F_k(\mathbf{w}_k, \Theta) = \log_2(1 + \gamma_k), \quad \forall k \in \mathcal{K}, \quad (4.7)$$

$$G_k(\epsilon_k, \mathbf{w}_k, \Theta) = Q^{-1}(\epsilon_k) \sqrt{\frac{1}{m_d} V_k}, \quad \forall k \in \mathcal{K}. \quad (4.8)$$

Furthermore, within the context of URLLC systems, ϵ_k characterizes the decoding error probability for user k , providing a quantifiable measure of the likelihood that a transmitted message is incorrectly decoded. The term m_d specifies the blocklength, indicating the number of symbols contained in each transmitted block of data. This parameter is crucial in balancing the trade-off between latency and reliability, as shorter blocklengths contribute to reduced transmission times at the cost of potentially increased error rates. Additionally, V_k represents the channel dispersion for user k , a metric that reflects the variability in the channel's capacity to convey information reliably. It is mathematically defined as:

$$V_k = a^2 \left(1 - (1 + \gamma_k)^{-2} \right), \quad \forall k \in \mathcal{K}, \quad (4.9)$$

where $a = \log_2(e)$. The Equation (4.8) captures the channel dispersion and the effect of the finite blocklength on the achievable data rate, with γ_k denoting the SINR for user k .

In the analysis of URLLC traffic, the load associated with user k is generally modeled as a random variable, expressed as $L_k = \nu_k \Omega_k$. Here, ν_k is the size of individual data packets, whereas Ω_k denotes the rate at which these packets arrive. This model reflects the stochastic nature of network traffic, accommodating fluctuations in data demand and arrival patterns. To ensure that the QoS requirements for each user are met, it is essential that the probability of the traffic load surpassing the user's allocated total data rate remains below a predefined threshold, ζ . This threshold represents the maximum tolerable probability of failure in supporting the traffic load, a critical consideration in the design and optimization of URLLC systems to uphold stringent performance standards [172]. By adhering to these parameters and models, URLLC systems can achieve the delicate balance between low latency, high reliability, and efficient use of network resources, ensuring robust and responsive communication for critical applications. A probabilistic constraint could be established that reflects the critical balance between the system's ability to support user traffic and the intrinsic randomness of that traffic, ensuring that the QoS for each user is upheld within acceptable limits of reliability. Such a constraint is formally written as:

$$\Pr(L_k > R_k(\epsilon_k, \mathbf{w}_k, \Theta)) \leq \zeta, \quad 0 < \zeta < 1, \quad \forall k \in \mathcal{K}. \quad (4.10)$$

This inequality essentially stipulates that the probability of the traffic load L_k exceeding the achievable data rate $R_k(\epsilon_k, \mathbf{w}_k, \Theta)$ for each user k should not surpass a predefined threshold ζ , which lies between 0 and 1. This threshold shows the system's maximum acceptable risk level of failing to meet the data rate requirements due to variability in traffic load. In particular, in our system model, each user k 's packet size ν_k is assumed to be constant, while the packet arrival rate Ω_k is modeled to follow a Poisson distribution⁴ with a mean

⁴The variability introduced by the Poisson distribution captures the stochastic nature of traffic arrival, which is a critical aspect of accurately modeling and managing network resources in URLLC systems.

rate parameter Λ_k . The Cumulative Distribution Function (CDF) of the packet arrival rate for user k , denoted by $F_{\Omega_k}(\cdot)$, translates the probabilistic constraint into a tangible condition for system design and optimization. Through algebraic manipulation, the aforementioned probabilistic constraint can be reformulated into a more direct relationship between the system parameters and the QoS requirements, as follows:

$$R_k(\epsilon_k, \mathbf{w}_k, \Theta) \geq \nu_k F_{\Omega_k}^{-1}(1 - \zeta), \quad \forall k \in \mathcal{K}, \quad (4.11)$$

where $F_{\Omega_k}^{-1}$ represents the inverse CDF (quantile function) of Ω_k . This expression sets a minimum required data rate based on the packet size and the statistical behavior of packet arrivals, adjusted by the tolerable failure probability ζ .

This formulation allows network designers to specify and enforce QoS guarantees directly, considering both the inherent uncertainty of wireless channels and the stochastic nature of network traffic, thereby ensuring that URLLC systems can deliver the requisite performance levels for critical applications.

4.3 Resource Allocation Problem Formulation

In this segment of our investigation, the primary objective is to reduce the overall transmission power within the system under consideration. This goal is to be achieved through the simultaneous optimization of several key variables: the active beamforming vectors employed by the AP, the phase shifts at the IRS, and the DEP associated with each communication link. The optimization problem, thus, seeks to find an efficient balance between minimizing power consumption and fulfilling the system's operational constraints, ensuring that both performance and efficiency targets are met. The formal mathematical representation of this optimization challenge can be given as follows:

$$P_1 : \min_{\mathbf{w}_k, \Theta, \epsilon_k} \sum_{k \in \mathcal{K}} \|\mathbf{w}_k\|^2 \quad (4.12a)$$

$$\text{s.t. : } R_k(\epsilon_k, \mathbf{w}_k, \Theta) \geq \nu_k F_{\Omega_k}^{-1}(1 - \zeta), \quad \forall k \in \mathcal{K}, \quad (4.12b)$$

$$|\Theta_{nn}| = 1, \quad \forall n \in \mathcal{N}, \quad (4.12c)$$

$$\epsilon_k \leq \epsilon_{k,\max}, \quad \forall k \in \mathcal{K}. \quad (4.12d)$$

The constraint labeled as (4.12b) specifies the minimum data rate that must be sustained for user k , a critical requirement that ensures each user receives sufficient bandwidth to meet their QoS needs. Meanwhile, the constraint described in (4.12c) mandates that the elements along the main diagonal of the diagonal phase shift matrix — corresponding to the IRS's phase shifters — must all possess unit modulus. This requirement is the key for maintaining the integrity of the signal reflection process, ensuring that the IRS can effectively manipulate the incident signals to enhance communication links. Additionally, the constraint presented in (4.12d) is designed to uphold the reliability standards for each URLLC user within the network. Here, ϵ_{\max} represents the upper bound on the allowable

error rate for data transmission, highlighting the importance of maintaining high reliability in URLLC contexts. This constraint is crucial for applications demanding ultra-reliable communications, where even minimal losses or errors can have significant implications.

By addressing the optimization problem P_1 , this chapter elucidates strategies for reducing power consumption in IRS-aided URLLC wireless networks while simultaneously enhancing their capacity to support high-performance, reliable communications. Through the management of beamforming techniques, phase shift adjustments, and error probabilities, the study aims to contribute valuable insights into the design and operation of energy-efficient, high-capacity IRS-aided URLLC systems.

4.4 Proposed Solution

The optimization problem P_1 presents a significant challenge due to its non-convex nature, primarily arising from the complex interdependencies among the optimization variables. Such non-convexity typically renders direct, efficient solutions elusive, as there lacks a straightforward method to navigate the complex landscape of potential solutions. To tackle this issue, we propose employing an AO strategy, known for its lower computational complexity, as a pragmatic approach to approximating a sub-optimal solution.

The AO method we suggest decomposes the main problem, P_1 , into more manageable sub-problems, each focusing on a subset of the optimization variables. This decomposition allows for iterative refinement of the variables in a manner that gradually converges towards an improved solution. Specifically, the process begins by addressing the first sub-problem, which involves the design of the active beamforming vectors at the AP. Here, we leverage the SCA technique alongside the DC approach. The SCA technique handles a non-convex optimization by iteratively solving convex approximations of the original problem, thereby easing the computational burden. Simultaneously, the DC approach facilitates the breakdown of non-convex components into convex and concave parts, further simplifying the optimization process.

Following the resolution of the beamforming design, the focus shifts to the second sub-problem, which targets the optimization of the phase shifts at the IRS. This phase of optimization utilizes a penalty approach in conjunction with the SCA technique. The penalty method introduces auxiliary constraints to transform the original problem into a penalized version, where constraints are incorporated into the objective function as penalty terms. This transformation often simplifies constraint handling, making the optimization problem more tractable. The application of the SCA technique within this context ensures that each iteration moves closer to fulfilling the original constraints while progressively optimizing the phase shifts.

Additionally, to circumvent potential issues related to the feasibility of solutions — common in complex optimization scenarios — a new objective function is proposed. This objective aims to guide the optimization process more effectively, ensuring that the search for sub-optimal solutions remains within the bounds of practical and achievable solution space. Through this structured two-layered approach, our method promises to yield efficient and effective solutions to the challenging non-convex problem P_1 , facilitating advancements in

the design and optimization of communication systems where active beamforming and IRS phase shifts are crucial elements.

4.4.1 First-stage: Optimizing \mathbf{w}_k and ϵ_k with Fixed Θ

At this stage of our optimization process, we direct our focus towards the active beamformers at the AP and the DEP for each user, while treating the phase shifts implemented by the IRS, denoted as Θ , as fixed parameters. This assumption simplifies the optimization problem by temporarily isolating a subset of variables, thereby allowing for a more targeted approach in optimizing the active beamformers, \mathbf{w}_k , and DEP, ϵ_k .

To facilitate this optimization, we employ the technique of Semi-Definite Programming (SDP), a powerful mathematical framework well-suited for handling optimization problems involving linear matrix inequalities. By adopting SDP, we introduce matrices $\mathbf{W}_k = \mathbf{w}_k \mathbf{w}_k^H$ and $\mathbf{H}_k = \mathbf{h}_k \mathbf{h}_k^H$ for all users $k \in \mathcal{K}$. This transformation converts the original beamforming vectors and channel coefficients into their respective matrix forms. Under the SDP framework, the original problem P_1 undergoes a reformulation, denoted here as⁵:

$$P_2 : \min_{\mathbf{W}_k, \epsilon_k} \sum_{k \in \mathcal{K}} \text{Tr}(\mathbf{W}_k) \quad (4.13a)$$

$$\text{s.t.: } F_k(\mathbf{W}_k) - G_k(\mathbf{W}_k, \epsilon_k) \geq \nu_k F_{\Omega_k}^{-1}(1 - \zeta), \forall k \in \mathcal{K}, \quad (4.13b)$$

$$\text{rank}(\mathbf{W}_k) \leq 1, \quad \forall k \in \mathcal{K}, \quad (4.13c)$$

$$\mathbf{W}_k \succeq \mathbf{0}, \quad \forall k \in \mathcal{K}, \quad (4.13d)$$

$$\epsilon_k \leq \epsilon_{k,\max}, \quad \forall k \in \mathcal{K}, \quad (4.13e)$$

where γ_k in $F_k(\mathbf{W}_k)$ and $G_k(\mathbf{W}_k)$ can be expressed as:

$$\gamma_k = \frac{\text{Tr}(\mathbf{h}_k \mathbf{w}_k)}{\sum_{i \in \mathcal{K}, i \neq k} \text{Tr}(\mathbf{h}_k \mathbf{w}_i) + \sigma_k^2}, \forall k \in \mathcal{K}. \quad (4.14)$$

This stage of the optimization process is crucial for iteratively refining the system's performance, setting the stage for subsequent optimization of the IRS phase shifts. By effectively decoupling the problem into manageable subproblems and employing sophisticated mathematical techniques like SDP, we inch closer to achieving our goal of minimizing the total transmit power while adhering to the system's operational constraints and quality of service requirements.

Addressing the challenge posed by the non-concavity of constraint (4.13b) in the optimization problem P_2 requires a strategic approach to ensure the tractability of the optimization

⁵This reformulated version harnesses the power of SDP to navigate the complexities of optimizing beamformers and decoding error probabilities within the constraints of fixed IRS phase shifts. By representing the beamforming vectors and channel links as semidefinite matrices, the problem becomes more tractable, allowing for the (potential of) utilization of SDP solvers to find optimal or near-optimal solutions.

process. To tackle this issue, we introduce a novel approach that involves the deployment of auxiliary variables μ_k , $\forall k \in \mathcal{K}$. These auxiliary variables are designed to establish a lower bound for the SINR, facilitating a more manageable form of the SINR expression that can be optimized effectively. With this approach, the SINR constraint specified in (4.14) can be reformulated as follows:

$$0 \leq \mu_k \leq \gamma_k = \frac{f_k(\mathbf{W}_k)}{g_k(\mathbf{W}_k)}, \forall k \in \mathcal{K}. \quad (4.15)$$

This reformulation allows us to decompose the SINR into components that are more conducive to optimization. Specifically, the numerator and denominator of the reformulated SINR expression, as indicated in (4.15), can be detailed as:

$$f_k(\mathbf{W}_k) = \text{Tr}(\mathbf{h}_k \mathbf{w}_k), \quad \forall k \in \mathcal{K}, \quad (4.16)$$

$$g_k(\mathbf{W}_k) = \sum_{i \in \mathcal{K}, i \neq k} \text{Tr}(\mathbf{h}_k \mathbf{w}_i) + \sigma_k^2, \forall k \in \mathcal{K}, \quad (4.17)$$

respectively. Based on this decomposition, both the signal power and the interference plus noise power components are articulated in terms that allow for the application of optimization techniques. Leveraging the lower bound provided by the auxiliary variables μ_k , the first stage of the optimization problem can then be reformulated. This rephrased optimization problem, retaining the essence of minimizing the total transmit power while satisfying the system's constraints, can be restated as:

$$\mathcal{P}_3 : \min_{\mathbf{W}_k, \mu_k, \epsilon_k} \sum_{k \in \mathcal{K}} \text{Tr}(\mathbf{W}_k) \quad (4.18a)$$

$$\text{s.t.: } \mu_k \geq 0, \quad \forall k \in \mathcal{K}, \quad (4.18b)$$

$$\mu_k \leq \frac{f_k(\mathbf{W}_k)}{g_k(\mathbf{W}_k)}, \quad \forall k \in \mathcal{K}, \quad (4.18c)$$

$$R_k(\epsilon_k, \mu_k) \geq \nu_k F_{\Omega_k}^{-1}(1 - \zeta), \quad \forall k \in \mathcal{K}, \quad (4.18d)$$

$$\text{rank}(\mathbf{W}_k) \leq 1, \quad \forall k \in \mathcal{K}, \quad (4.18e)$$

$$\mathbf{W}_k \succeq \mathbf{0}, \quad \forall k \in \mathcal{K}, \quad (4.18f)$$

$$\epsilon_k \leq \epsilon_{k,\max}, \quad \forall k \in \mathcal{K}, \quad (4.18g)$$

where

$$R_k(\epsilon_k, \mu_k) = F_k(\mu_k) - G_k(\epsilon_k, \mu_k), \quad \forall k \in \mathcal{K}, \quad (4.19)$$

In constraint (4.18d), and the terms $F_k(\mu_k)$ and $G_k(\epsilon_k, \mu_k)$ are given by:

$$F_k(\mu_k) = \log(1 + \mu_k), \quad \forall k \in \mathcal{K}, \quad (4.20)$$

$$G_k(\epsilon_k, \mu_k) = Q^{-1}(\epsilon_k) \sqrt{\frac{a^2}{m_d} \left(1 - (1 + \mu_k)^{-2}\right)}, \forall k \in \mathcal{K}. \quad (4.21)$$

To address the non-convex nature of \mathcal{P}_3 , the optimization problem is initially transformed into a canonical form that aligns with the requirements for DC programming techniques.

This transformation is crucial as it prepares the ground for subsequent approximation and manipulation techniques aimed at dealing with non-convex terms. Once the problem is cast into the canonical form required for DC programming, the next phase involves employing the first-order Taylor expansion. This mathematical tool allows us to approximate the non-convex components of the problem with convex ones, thus rendering the problem more tractable. Specifically, this approximation strategy is applied to constraint (4.18c), which can be represented as follows:

$$\begin{aligned} \mu_k g_k(\mathbf{W}_k) &\leq f_k(\mathbf{W}_k) \\ \Rightarrow \mu_k \mathcal{A}(\mathbf{W}_k) &\leq f_k(\mathbf{W}_k) - \mu_k \sigma_k^2, \forall k \in \mathcal{K}, \end{aligned} \quad (4.22)$$

where

$$\mathcal{A}(\mathbf{W}_k) = \sum_{i \in \mathcal{K}, i \neq k} \text{Tr}(\mathbf{h}_k \mathbf{w}_i), \quad \forall k \in \mathcal{K}. \quad (4.23)$$

This transformation is key to progressing the optimization process, as it breaks down complex, non-linear relationships into forms that are more amenable to analytical and numerical optimization techniques.

Nevertheless, a notable challenge arises with constraint (4.22), which inherently embodies non-convexity due to the multiplication of two optimization variables: the beamforming matrix \mathbf{W}_i and the auxiliary variable μ_k , applicable for all $i, k \in \mathcal{K}$. This type of constraint typically complicates the optimization process, as direct optimization methods struggle to handle such non-linear interdependencies effectively. To mitigate this challenge, a decoupling strategy is proposed. By adopting a specific form for the problematic constraint, as detailed in the literature [165, 2], the issue can be tackled. The proposed form, indicated as:

$$\mu_k \mathcal{A}(\mathbf{W}_k) = \mathcal{P}_k(\mu_k, \mathbf{W}_k) - \mathcal{Q}_k(\mu_k, \mathbf{W}_k), \forall k \in \mathcal{K}, \quad (4.24)$$

where

$$\mathcal{P}_k(\mu_k, \mathbf{W}_k) = \frac{1}{2} (\mu_k + \mathcal{A}(\mathbf{W}_k))^2, \quad \forall k \in \mathcal{K}, \quad (4.25)$$

$$\mathcal{Q}_k(\mu_k, \mathbf{W}_k) = \frac{1}{2} (\mu_k)^2 + (\mathcal{A}(\mathbf{W}_k))^2, \forall k \in \mathcal{K}, \quad (4.26)$$

allows for the separation of the intertwined variables, thereby simplifying the constraint into components that can be more readily optimized. This decoupling is key to overcoming the inherent non-convexity, enabling the identification of sub-optimal solutions within a complex problem space.

Through these strategic steps — casting the problem into a suitable DC-compatible form, applying convex approximations to non-convex terms, and decoupling intertwined variables — the once daunting task of optimizing P_3 becomes more approachable. This approach enables the pursuit of sub-optimal solutions to complex optimization problems within the domain of communications system design. By denoting $\Upsilon_k = \{\epsilon_k, \mu_k, \mathbf{W}_k\}$ as a set of optimization variables, P_3 can be recast as follows:

$$P_4 : \min_{\Upsilon_k} \sum_{k \in \mathcal{K}} \text{Tr}(\mathbf{W}_k) \quad (4.27a)$$

$$\text{s.t. : } F_k(\mu_k) - G_k(\mu_k) \geq \nu_k F_{\Omega_k}^{-1}(1 - \zeta), \quad \forall k \in \mathcal{K}, \quad (4.27b)$$

$$\mathcal{P}_k(\Upsilon_k) - \mathcal{Q}_k(\Upsilon_k) \leq f_k(\mathbf{W}_k) - \mu_k \sigma_k^2, \quad \forall k \in \mathcal{K}, \quad (4.27c)$$

$$\mu_k \geq 0, \quad \forall k \in \mathcal{K}, \quad (4.27d)$$

$$\epsilon_k \leq \epsilon_{k,\max}, \quad \forall k \in \mathcal{K}, \quad (4.27e)$$

$$\text{rank}(\mathbf{W}_k) \leq 1, \quad \forall k \in \mathcal{K}, \quad (4.27f)$$

$$\mathbf{W}_k \succeq \mathbf{0}, \quad \forall k \in \mathcal{K}. \quad (4.27g)$$

Addressing the optimization problem outlined in P_4 presents yet another obstacle: the incorporation of the $Q^{-1}(\cdot)$ function. This function, representing the inverse of the Gaussian Q-function, is notoriously difficult to handle directly within optimization frameworks due to its non-linear and non-convex nature. To effectively deal with this issue and advance towards a solution, we introduce a *Lemma* that facilitates an approximate representation of the $Q^{-1}(\cdot)$ function, thereby simplifying the optimization process.

Lemma 1 For $0 < \epsilon_k < 1$, an approximation of $Q^{-1}(\epsilon_k)$ is given by:

$$Q^{-1}(\epsilon_k) \approx \sqrt{\frac{\pi}{2}} (B - C\epsilon_k), \quad \forall k \in \mathcal{K}, \quad (4.28)$$

where B and C are defined as:

$$B = \left(1 + \frac{\pi}{12} + \frac{7\pi^2}{480} + \frac{127\pi^3}{40320} + \dots \right), \quad (4.29)$$

$$C = \left(1 + \frac{\pi}{2} + \frac{7\pi^2}{48} + \frac{127\pi^3}{2880} + \dots \right). \quad (4.30)$$

$$(4.31)$$

By adopting the **Lemma 1**, the transformation of the data rate constraint function as specified in (4.27b) becomes feasible, allowing us to recast it in a more tractable form. By adopting the lemma's approximation for the $Q^{-1}(\cdot)$ function, we arrive at a new representation of the constraint as follows:

$$F_k(\mu_k) - \sqrt{\frac{\pi}{2 m_d}} \left(1 - \frac{1}{(1 + \gamma_k)^2} \right) (B - C\epsilon_k) \geq R_{\min}, \quad \forall k \in \mathcal{K}. \quad (4.32)$$

This reformulation significantly simplifies the original problem by providing an explicit relation that incorporates the effects of decoding error probability, ϵ_k , and the SINR, γ_k , into the data rate constraint. However, the constraint as expressed in (4.32) remains non-convex, largely due to the presence of the channel dispersion term, which complicates the direct application of convex optimization techniques. To circumvent this issue, we introduce an assumption applicable to the high SINR regime, where the channel dispersion, V_k , can be

closely approximated by:

$$V_k = \left(1 - \frac{1}{(1 + \gamma_k)^2}\right) \approx 1, \quad \forall k \in \mathcal{K}. \quad (4.33)$$

This approximation assumes that as SINR, γ_k , increases, the impact of the channel's variance on the system's performance diminishes, allowing us to treat V_k as approximately constant. Under this assumption, the complex non-convex equation (4.32) can be simplified further, resulting in:

$$F_k(\mu_k) - \underbrace{\sqrt{\frac{\pi}{2 m_k}}}_{= \tilde{G}_k(\epsilon_k, \mu_k)} (B - C\epsilon_k) \geq R_{\min}, \quad \forall k \in \mathcal{K}. \quad (4.34)$$

This revised formulation, (4.34), represents a significant step towards enabling the application of convex optimization strategies by describing the constraint in terms that are inherently more helpful to such approaches. Here, $\tilde{G}_k(\epsilon_k, \mu_k)$ is a modified term that contains the interplay between the decoding error probability and the auxiliary variables, μ_k . Through this strategic approximation and reformulation, the pathway to addressing the optimization challenge in P_4 becomes clearer, facilitating the advancement toward finding viable, efficient solutions within the defined constraints.

It should be noted that the constraint (4.27c) also belongs to the class of DC problems. Thus, the SCA technique can be directly applied to approximate the non-convex problem in each iteration. Based on this similar recognition, the first-order Taylor expansion becomes the key in crafting a globally lower-bound approximation for the function $\mathcal{Q}_k(\Upsilon_k)$ for each user $k \in \mathcal{K}$. At a given iteration t , the lower-bound approximations of these functions are given by:

$$\begin{aligned} \mathcal{Q}_k(\Upsilon_k) \geq \tilde{\mathcal{Q}}_k(\Upsilon_k) \triangleq & \mathcal{Q}_k(\Upsilon_k^t) + \partial_{\mu_k}^T \mathcal{Q}_k(\Upsilon_k^t) (\mu_k - \mu_k^t) \\ & + \text{Tr} \left(\nabla_{\mathbf{W}_k}^H \mathcal{Q}_k(\Upsilon_k^t) (\mathbf{W}_k - \mathbf{W}_k^t) \right), \quad \forall k \in \mathcal{K}, \end{aligned} \quad (4.35)$$

facilitating a piecewise convex approximation of the original, more complex, problem. Following this strategic maneuver and by dropping the inherently non-convex rank-one constraint, P_4 with any given local point at iteration t can be approximated as:

$$P_5 : \min_{\Upsilon_k} \sum_{k \in \mathcal{K}} \text{Tr}(\mathbf{W}_k) \quad (4.36a)$$

$$\text{s.t.} : F_k(\mu_k) - \tilde{G}_k(\epsilon_k, \mu_k) \geq R_{\min}, \quad \forall k \in \mathcal{K}, \quad (4.36b)$$

$$\mathcal{P}_k(\Upsilon_k) - \tilde{\mathcal{Q}}_k(\Upsilon_k) \leq f_k(\mathbf{W}_k) - \mu_k \sigma_k^2, \quad \forall k \in \mathcal{K}, \quad (4.36c)$$

$$\mathbf{W}_k \succeq \mathbf{0}, \quad \forall k \in \mathcal{K}, \quad (4.36d)$$

$$\mu_k \geq 0, \quad \forall k \in \mathcal{K}, \quad (4.36e)$$

$$\epsilon_k \leq \epsilon_{k,\max}, \quad \forall k \in \mathcal{K}. \quad (4.36f)$$

Algorithm 2 Iterative SCA Algorithm for Energy-efficient Resource Allocation Policy in IRS-aided URLLC Networks

Input: Set iteration number $t = 0$, maximum number of iterations T_{\max} , and initialize the decoding error $\epsilon_k = \epsilon_k^0$, the auxiliary variable $\mu_k = \mu_k^0$, and the active beamformers as $\mathbf{W}_k = \mathbf{W}_k^{(0)}$.

- 1: **repeat**
- 2: Calculate $\tilde{G}_k(\mu_k)$ and $\tilde{Q}_k(\Upsilon_k)$ as stated in (4.34) and (4.35), respectively.
- 3: Solve P_5 to obtain $\{\epsilon_k^t, \mu_k^t, \mathbf{W}_k^t\}$.
- 4: Set $t = t + 1$.
- 5: **until** $t = T_{\max}$
- 6: **Return** $\Upsilon_k^* = \{\epsilon_k^t, \mu_k^t, \mathbf{W}_k^t\} = \{\epsilon_k^*, \mu_k^*, \mathbf{W}_k^*\}$.

The optimization problem P_5 is now a convex optimization problem that can be efficiently solved by standard convex optimization solvers such as CVX. P_5 , now framed as a convex optimization challenge, paves the way for iterative, efficient resolution through SCA. The iterative SCA algorithm for P_5 is given in **Algorithm 2**.

4.4.2 Second-stage: Optimizing Θ with Fixed \mathbf{w}_k and ϵ_k

In the second stage of our optimization process, we focus on optimizing the phase shifts at the IRS, denoted as Θ , while keeping the active beamforming vectors \mathbf{w}_k and the DEPs ϵ_k fixed. This step presents a unique set of challenges, primarily due to the constraints associated with the phase shifts.

The main difficulty in this context is constraint (4.12c), which imposes a unit-modulus requirement on the phase shifts. This unit-modulus constraint is inherently non-convex, making the direct optimization of the phase shifts intractable with standard convex optimization tools. To overcome this issue, we introduce a new variable representation for the phase shifts. We define the vector $\boldsymbol{\theta} = (e^{j\alpha_1}, \dots, e^{j\alpha_N})^H \in \mathbb{C}^{N \times 1}$, containing the phase shifts of the IRS elements, and augment it with a dummy variable $\tau \in \mathbb{C}$, such that $|\tau| = 1$, to form the extended vector $\tilde{\boldsymbol{\theta}} = [\boldsymbol{\theta}^T \ \tau]^T \in \mathbb{C}^{(N+1) \times 1}$.

To further facilitate the solution, we introduce the matrix $\mathbf{V} = \tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^H \in \mathbb{C}^{(N+1) \times (N+1)}$. This representation ensures that \mathbf{V} is semi-definite and satisfies the condition $\text{rank}(\mathbf{V}) \leq 1$. Leveraging this formulation allows us to circumvent the direct handling of the unit-modulus constraint by focusing on the properties of \mathbf{V} . Thus, we obtain:

$$\begin{aligned} \left| \left((\mathbf{h}_k^{\text{IU}})^H \boldsymbol{\Theta} \mathbf{H} + (\mathbf{h}_k^{\text{AU}})^H \right) \mathbf{w}_k \right|^2 &\triangleq \text{Tr}(\mathbf{V} \mathbf{X}_k \mathbf{W}_k \mathbf{X}_k^H) \\ &= \text{Tr}(\mathbf{W}_k \mathbf{Y}_k), \forall k \in \mathcal{K}, \end{aligned} \quad (4.37)$$

where

$$\mathbf{X}_k = \begin{bmatrix} \left(\text{diag}((\mathbf{h}_k^{\text{IU}})^H) \mathbf{H} \right)^T & (\mathbf{h}_k^{\text{AU}})^* \end{bmatrix}^T, \forall k \in \mathcal{K}, \quad (4.38)$$

$$\mathbf{Y}_k = \mathbf{X}_k^H \mathbf{V} \mathbf{X}_k, \quad \forall k \in \mathcal{K}. \quad (4.39)$$

Given that the objective function in P_1 does not depend on \mathbf{V} , the problem effectively becomes a question of feasibility. To resolve this and effectively derive the phase shift matrix Θ , we explore an alternative optimization strategy that bypasses the direct feasibility challenge. This approach involves studying the subsequent optimization problem:

$$P_6 : \underset{\mathbf{V}, \alpha_k}{\text{maximize}} \quad \sum_{k=1}^K \alpha_k \quad (4.40a)$$

$$\text{s.t.: } \text{Tr}(\mathbf{W}_k \mathbf{Y}_k) - \sum_{i \in \mathcal{K}, i \neq k} \text{Tr}(\mu_k^* \mathbf{W}_i \mathbf{Y}_k) \geq \mu_k^* \sigma_k^2 + \alpha_k, \forall k \in \mathcal{K}, \quad (4.40b)$$

$$\text{diag}(\mathbf{V}) = \mathbf{1}_{N+1}, \quad (4.40c)$$

$$\mathbf{V} \succeq \mathbf{0}, \quad (4.40d)$$

$$\text{rank}(\mathbf{V}) \leq 1. \quad (4.40e)$$

With the objective of enhancing the SINR margin beyond the minimum requirements outlined in P_1 , the optimization process seeks to precisely determine the configuration of Θ , the phase shift matrix at the IRS. This step enables the maximum utilization of the communication system's performance by fine-tuning the phase shifts to align and strengthen the signal at the intended receivers, thereby increasing the SINR margin. An essential aspect of this optimization is adhering to constraint (4.40c), which enforces the unit-modulus nature of the IRS's reflective elements. This constraint maintains the physical feasibility of the phase shifts, ensuring that each element of Θ reflects signals without amplifying or attenuating their power.

In practice, solving P_6 , which incorporates the unit-modulus constraint, often results in a solution matrix with a rank greater than one. This poses a challenge since the ideal solution would have a rank of one to correspond with the physical implementation of a single phase shift per IRS element. Therefore, it is not justifiable to neglect and drop the constraint (4.40c) as we did so in P_4 . Exploiting the DC programming method we explored earlier could be beneficial in tackling this issue. By applying the DC method, the unit-modulus constraint can be recast into an equivalent form that is mathematically tractable and more suitable for optimization processes. Thus, the equivalent form of constraint (4.40c) can be represented as⁶:

$$\|\mathbf{V}\|_* - \|\mathbf{V}\|_2 \leq 0. \quad (4.41)$$

⁶By strategically addressing the rank and unit-modulus constraints, the optimization process can converge towards a solution that not only satisfies the mathematical model but also aligns with the physical capabilities and limitations of the IRS technology. This approach underscores the intricate balance between theoretical optimization strategies and practical implementation considerations, ultimately enabling the realization of IRS-assisted communication systems that leverage optimized phase shifts to achieve enhanced performance metrics.

Note that:

$$\|\mathbf{V}\|_* = \sum_i \sigma_i \geq \|\mathbf{V}\|_2 = \max_i \{\sigma_i\}. \quad (4.42)$$

This holds for any given $\mathbf{V} \in \mathbb{H}^{N \times N}$, where σ_i is the i -th singular value of \mathbf{V} . The equality holds if and only if \mathbf{V} achieves rank one i.e., $\text{rank}(\mathbf{V}) = 1$ [173, 174]. Now, we take the first-order Taylor approximation of $\|\mathbf{V}\|_2$ as:

$$\|\mathbf{V}\|_2 \geq \|\mathbf{V}^{(t)}\|_2 + \text{Tr} \left(\lambda_{\max} \left(\mathbf{V}^{(t)} \right) \lambda_{\max}^H \left(\mathbf{V}^{(t)} \right) (\mathbf{V} - \mathbf{V}^t) \right). \quad (4.43)$$

By resorting to (4.43), a convex approximation can be obtained for (4.41) which is given by:

$$\|\mathbf{V}\|_* - \|\mathbf{V}^{(t)}\|_2 - \text{Tr} \left(\lambda_{\max} \left(\mathbf{V}^{(t)} \right) \lambda_{\max}^H \left(\mathbf{V}^{(t)} \right) (\mathbf{V} - \mathbf{V}^t) \right) \leq 0. \quad (4.44)$$

Finally, with the convex constraint (4.44) at hand, the optimization problem in the $(t+1)$ -iteration can be written as follows:

$$P_7 : \underset{\mathbf{V}, \alpha_k}{\text{maximize}} \quad \sum_{k=1}^K \alpha_k \quad (4.45a)$$

$$\text{s.t. : } \text{Tr}(\mathbf{W}_k \mathbf{Y}_k) - \sum_{i \in \mathcal{K}, i \neq k} \text{Tr}(\mu_k^* \mathbf{W}_i \mathbf{Y}_k) \geq \mu_k^* \sigma_k^2 + \alpha_k, \quad \forall k \in \mathcal{K}, \quad (4.45b)$$

$$\text{diag}(\mathbf{V}) = \mathbf{1}_{N+1}, \quad (4.45c)$$

$$\mathbf{V} \succeq \mathbf{0}, \quad (4.45d)$$

$$\|\mathbf{V}\|_* - \|\mathbf{V}^{(t)}\|_2 - \text{Tr} \left(\lambda_{\max} \left(\mathbf{V}^{(t)} \right) \lambda_{\max}^H \left(\mathbf{V}^{(t)} \right) (\mathbf{V} - \mathbf{V}^t) \right) \leq 0. \quad (4.45e)$$

Following the reformulation and addressing of the unit-modulus and rank constraints via the DC method, the optimization problem, now referred to as P_7 , achieves a convex structure. This transformation is significant because it transitions the problem into a domain where established convex optimization techniques, such as CVX, can be applied effectively [175].

The convexity of P_7 ensures that the optimization can be carried out with guarantees of reaching a global maximum within the defined solution space. The culmination of our efforts is encapsulated in the final iterative-based AO algorithm, which is detailed in **Algorithm 3**. This algorithm iteratively applies the AO method, alternating between optimizing different sets of variables while keeping others fixed, gradually converging to a solution that optimizes the system's total transmit power, phase shifts at the IRS, and the DEP for each user. By systematically addressing the various aspects of the optimization problem in stages, the AO algorithm navigates the complexities of the design space, leveraging the strengths of convex optimization to ensure efficient and effective solution convergence.

The iterative nature of the algorithm allows for continuous refinement of the solution, with each iteration bringing the system configuration closer to the optimal setting. This process demonstrates the power of combining theoretical optimization frameworks with practical

Algorithm 3 Iterative AO algorithm for Energy-efficient Resource Allocation Policy in IRS-aided URLLC Networks

Input: Set $i = 0$, l_{\max} , and initialize the phase shifts as $\Theta = \Theta^0$.

1: **Repeat**

2: Solve problem P_5 for given Θ^i , and obtain the

optimal solutions $\{\epsilon_k^i, \mu_k^i, \mathbf{W}_k^i\}$.

3: Solve problem P_7 for given $\{\epsilon_k^i, \mu_k^i, \mathbf{W}_k^i\}$.

4: $i = i + 1$.

5: **until** $i = l_{\max}$

6: **Return** $\{\epsilon_k^*, \mu_k^*, \mathbf{w}_k^*, \Theta^*\} = \{\epsilon_k^i, \mu_k^i, \mathbf{W}_k^i, \Theta^i\}$.

solution techniques like CVX and highlights the potential of IRS-assisted communication systems to achieve enhanced performance through careful system design and optimization.

Proposition 1 P_7 is non-increasing as the objective function value increases over each iteration in **Algorithm 3**. In particular, after each iteration, the iterative **Algorithm 3** improves the objective function value of P_7 and converges to a locally optimal solution

Proof 1 Please see [12]. ■

For a detailed exposition of the proof, including mathematical demonstrations and the algorithmic steps that ensure convergence to a locally optimal solution, refer to the supplementary materials and discussions provided in the associated literature [151].

This proposition and its supporting proof underscore the efficacy of the optimized framework employed to tackle the complexities of the problem space, iteratively steering the system towards enhanced performance metrics. The proof further solidifies the theoretical foundation of the algorithm's design, ensuring its applicability and reliability in practical optimization scenarios within IRS-assisted communication systems.

4.5 Numerical Results

This section demonstrates the efficiency and effectiveness of our newly developed algorithm, **Algorithm 3**, for MISO URLLC-supported IRS systems, particularly under scenarios utilizing finite blocklength codes. We establish our simulation environment within a defined square space, measuring (100, 100) meters. Here, the AP is strategically positioned at the coordinates (0, 0) meters, with the IRS located at (50, 0) meters. We numerically simulate a dynamic environment where all users are randomly distributed across this rectangular space, adding a layer of realism to our evaluation. The signal path loss is modeled using the equation $35.3 + 37.6 \log_{10}(d_k)$ dB, where d_k represents the distance in meters from the AP to user k , providing a 3gpp-compliant assessment of signal attenuation over distance [55]. For the purposes of our simulations, we have set the convergence tolerance at 10^{-2} and assumed a thermal noise density of -174 dBm/Hertz (Hz), which aligns with standard wireless communication scenarios. Furthermore, we impose a strict requirement on the maximum DEP

Table 4.1: Simulation Parameters for Multi-user MISO URLLC-enabled IRS System.

Parameter	Value
Area dimensions	(100, 100) meters
AP location	(0, 0) meters
IRS location	(50, 0) meters
User distribution	Random within area
Path loss model	$35.3 + 37.6 \log_{10}(d_k)$ dB
Convergence tolerance	10^{-2}
Thermal noise density	-174 dBm/Hz
Max DEP ($\epsilon_{k,\max}$)	10^{-7}
Average traffic load	0.1 Mbps
Central carrier frequency	3.5 GHz
Bandwidth	350 kHz
Number of URLLC users (K)	4
Number of IRS elements (N)	50
Block code length	200 symbols
Static power consumption (P_s)	100 milli Watt (mW)
Dynamic power per IRS element (P_d)	0.33 mW
Circuit power at AP (P_c)	1 Watt (W)
Dynamic power of AP per antenna (P_{Dyn})	100 mW

for any user k , which is defined as $\epsilon_{k,\max} = 10^{-7}$, ensuring ultra-reliability in the URLLC context. Additionally, we model the network to accommodate an average traffic load of 0.1 Megabits per second (Mbps) per URLLC user, with a total system bandwidth of 350 kHz and a set number of URLLC users ($K = 4$) and IRS elements ($N = 50$).

To ensure the reliability and robustness of our simulation results, we employ Monte Carlo simulations, generating numerous random realizations of the channel gains. This approach allows us to compute the average EE across various scenarios, offering a comprehensive assessment of the proposed scheme's performance under diverse conditions. Through these simulations, we aim to demonstrate the viability and advantages of the proposed algorithm in enhancing the operational efficiency of IRS-aided URLLC networks.

Figure 4.2 illustrates the relationship between the DEP, denoted as $\epsilon_{k,\max}$, and the average transmit power required for block codes of length 200 symbols, symbolized by $m_d = 250$. An intriguing observation from this figure is the inverse relationship between the decoding error and the required transmit power; as the decoding error increases, the necessity for transmit power diminishes. This trend underscores a fundamental principle in network reliability: enhancing a network's reliability necessitates higher transmit power due to the decreasing nature of $Q^{-1}(\epsilon_k)$ with respect to ϵ_k , leading to a reduction in $G_k(\epsilon_k, \mathbf{w}_k, \Theta)$. Consequently, achieving the minimum data rate requirements becomes feasible with lower transmit power, ultimately reducing the AP's overall transmit power. Furthermore, the figure investigates the effect of increasing the number of reflecting elements at the IRS, which leads to a reduction in transmit power at the AP. In other words, the AP transmit power scales down with an increasing number of reflecting elements. A comparative analysis with two baseline schemes is also presented within this context. The first baseline scheme assumes a scenario with fixed beamforming at the IRS, whereas the second scenario operates without an IRS.

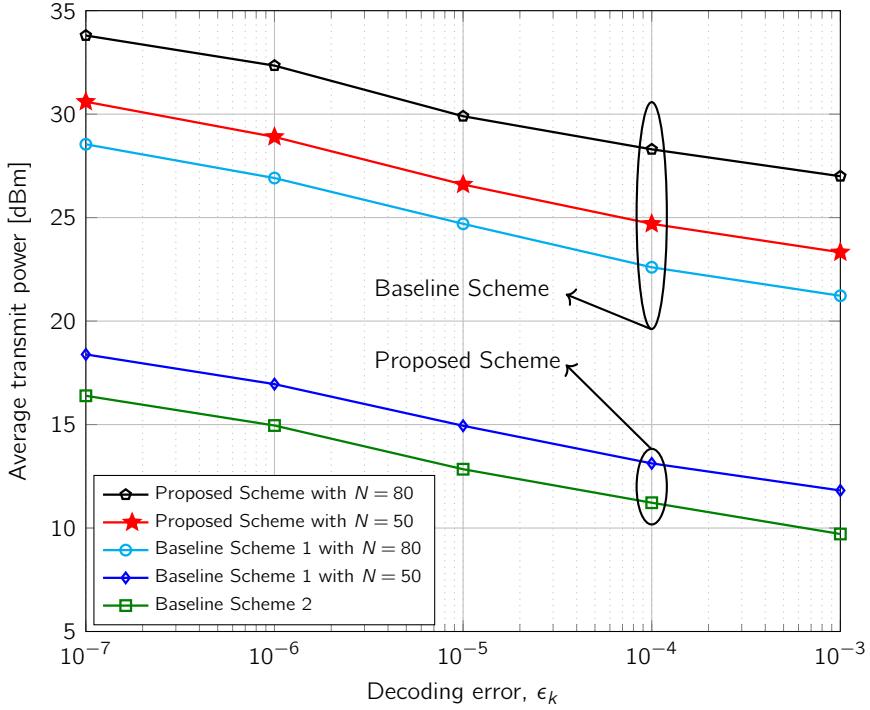


Figure 4.2: Impact of decoding error, ϵ_k , on the average transmit power for a downlink of multi-user IRS-aided URLLC network.

The comparison highlights the superior performance of our proposed algorithm, attributing its enhanced efficiency to the strategic deployment of IRS and the joint optimization of beamforming matrices at both the AP and IRS, thereby outperforming the first baseline scenario and significantly surpassing the second scenario that lacks IRS integration. This analysis demonstrates the efficacy of incorporating IRS into network design and emphasizes the key role of adaptive resource allocation in optimizing network performance.

The concept of IRS has been recognized as a revolutionary approach towards achieving environmentally friendly wireless communication systems. To quantitatively assess this, we introduce the metric of energy efficiency (EE), defined as the total system data rate divided by the overall network power consumption, measured in bits per joule. This relationship can be mathematically expressed as follows:

$$\mathcal{E}_{\text{eff}}(\epsilon_k, \mathbf{w}_k, \Theta) = \frac{\sum_{k \in \mathcal{K}} R_k(\epsilon_k, \mathbf{w}_k, \Theta)}{\sum_{k \in \mathcal{K}} \|\mathbf{w}_k\|^2 + P_s + N_T P_d + P_c + M P_{\text{Dyn}}}, \quad (4.46)$$

where $P_s = 100$ mW indicates the static power consumption as required to maintain the basic circuit operations of the IRSs, $P_d = 0.33$ mW is the dynamic power dissipation per reflecting component, $P_c = 1$ W is the circuit power at the AP, and $P_{\text{Dyn}} = 100$ mW is the dynamic power consumption of the AP per antenna.

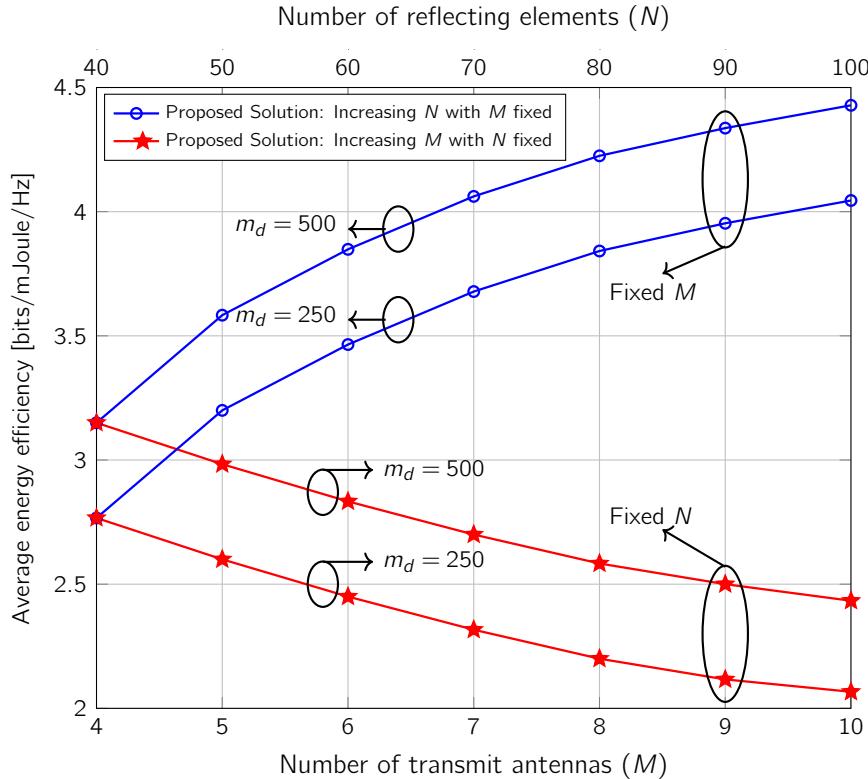


Figure 4.3: Average EE vs. the number of transmit antennas and reflecting elements for a downlink of multi-user IRS-aided URLLC network.

It is observed that the EE increases with increasing the number of reflecting elements. On the other hand, the EE of the system decreases as the number of antennas increases. This is because with increasing the number of antennas at the AP, the power consumption at the AP increases due to increasing the number of RF chains, which degrades the performance of the system in terms of EE. One can conclude that the IRS are more efficient for green wireless communication as it do not consume more transmit power as they are passive device. Besides, increasing the number of reflecting elements provides more degrees of freedom for the network to increase the network's data rate while reducing the system's transmit power.

Fig. 4.3 illustrates the relationship between the IRS's reflective elements and the AP's transmit antennas versus EE. Notably, the figure demonstrates an increase in EE with the augmentation of reflective elements on the IRS. This increment can be attributed to the reflective elements' role in enhancing the signal's directionality and strength without necessitating additional power. Conversely, an increase in the number of transmit antennas at the AP correlates with a decrease in EE. This decline is primarily due to the heightened power requirements associated with the additional RF chains needed for more antennas, adversely affecting the system's overall EE.

The insights gained from this analysis underline the IRS's capability to provide green wireless communication efforts. Being inherently passive, IRS units do not contribute to increased

transmit power, thereby presenting a sustainable alternative to conventional methods that rely on augmenting the number of active antennas at the AP. Furthermore, by expanding the number of reflecting elements, the IRS provides additional degrees of freedom. This expansion facilitates a more efficient network operation by enabling higher data rates and reduced transmit power, showcasing the IRS's substantial impact on enhancing wireless network efficiency and sustainability.

4.6 Conclusion

This chapter delved into the intricacies of resource allocation strategies for a DownLink (DL) multi-user Multiple-Input Single-Output (MISO) system, augmented by Ultra-Reliable low-Latency Communication (URLLC) capabilities through the integration of an Intelligent Reflecting Surface (IRS). The focus was on designing a resource allocation framework that optimizes both active and passive beamforming. This approach aimed to significantly reduce the total transmission power, all while considering the unique traffic loads of each URLLC user and adhering to strict Quality of Service (QoS) requirements based on the implementation of short packet transmission. The challenge presented by the non-convex nature of the problem was addressed by employing the Alternating Optimization (AO) method. This method strategically decomposed the main problem into manageable sub-problems, specifically focusing on the optimization of active and passive beamforming matrices one at a time. These sub-problems were then tackled using the Successive Convex Approximation (SCA) approach and a penalty-based method, respectively.

The effectiveness of the proposed scheme was rigorously evaluated through simulations, which highlighted the IRS's critical role in fulfilling the URLLC system's QoS demands and achieving substantial reductions in transmission power, demonstrating a marked improvement over traditional methodologies. Moreover, the simulations support the IRS's contribution to enhancing EE, indicating its potential to revolutionize power-efficient green communications. These findings illustrate the transformative impact of IRS technology in enhancing the performance and sustainability of future wireless communication systems, offering promising pathways for further research and development in the field of URLLC-aided wireless communications.

Chapter 5

Energy Efficiency and Admission Control in IRS Networks

In the previous chapter, Energy Efficiency (EE) in an Intelligent Reflecting Surface (IRS)-aided network was studied as the sole design objective of the network to be optimized. This chapter explores the dynamics of a multi-user Multiple-Input Single-Output (MISO) system enhanced by Machine Type Communication (MTC) technology and enabled with an IRS as a 'joint' optimization of EE and another network metric. Here, a multi-antenna Access Point (AP) is tasked with transmitting information symbols to numerous Internet of Things (IoT) users, all within the constraints of short packet transmission. The core objective revolves around simultaneously elevating the system's total EE and optimizing the number of IoT users that could be served fairly by jointly optimizing active and passive beamformers. This is achieved through a novel algorithm that employs alternating optimization (AO) to iteratively solve the main optimization problem.

To facilitate this optimization, the approach adopts properties of the Difference of Convex (DC) functions and the methodological precision of Successive Convex Approximation (SCA), building a concave-convex problem. The active beamformers at the AP, along with the strategic admittance of users into the system, are optimized through fractional programming techniques, specifically exploiting a quadratic form to reach a sub-optimal solution. For passive beamforming optimization, critical for manipulating Non-Line-of-Sight (NLoS) signals via the IRS, a novel combination of a penalty-based strategy and the SCA technique is employed. This duo effectively addresses the complexities introduced by the unit-modulus constraints integral to the IRS's operational framework.

Through extensive simulations, a trade-off emerges between EE and the system's capacity for user admissibility, highlighting the balancing required to optimize both active and passive beamformers. Moreover, the simulations underscore the significant impact of IRS deployment on the system's EE and its enhanced capability to incorporate a greater number of users. This exploration not only affirms the proposed algorithm's efficacy in solving the optimization problem at hand but also illuminates the transformative potential of IRS technology in improving the capabilities of MISO MTC-enabled frameworks in the IoT communication sphere.

The latter stages of this chapter venture into the evolving landscape of MTC, particularly

as they transition into the era of sixth-Generation (6G) networks. This transition marks a significant shift towards accommodating mission-critical applications that necessitate Ultra-Reliable and Low-Latency Communications (URLLCs). These applications, ranging from wireless industrial automation to healthcare, require not only unwavering reliability but also rigorous adherence to delay-Quality of Service (d-QoS) standards. This shift introduces fresh challenges in the radio access design for MTC networks, underscoring the need for novel solutions. Emphasizing the adoption of short packet transmission strategies with finite block-lengths emerges as crucial for meeting the strict low latency requirements central to URLLCs.

In response to these challenges, this chapter also delves into an energy-efficient resource allocation design algorithm for an IRS-assisted downlink URLLC network, building upon the foundational MTC framework established earlier. While employing similar techniques of decomposing the main non-convex problem into more manageable sub-problems and leveraging an AO approach enhanced with SCA, we introduce an innovative iterative rank relaxation method. This method allows for the formulation of a concave-convex objective function for each sub-problem. This approach aids in the precise optimization of system parameters and surpasses existing benchmarks through an iterative solution that methodically approaches rank-one solutions for both the active beamforming and IRS phase-shift sub-problems.

This chapter is based on:

J. Jalali, A. Khalili, A. Rezaei, R. Berkvens, M. Weyn and J. Famaey, "IRS-Based Energy Efficiency and Admission Control Maximization for IoT Users With Short Packet Lengths", *IEEE Transactions on Vehicular Technology*, vol. 72, no. 9, pp. 12 379–12 384, Sept. 2023. <https://doi.org/10.1109/TVT.2023.3266424>

J. Jalali, F. Lemic, H. Tabassum, R. Berkvens, and J. Famaey, "Toward Energy Efficient Multiuser IRS-Assisted URLLC Systems: A Novel Rank Relaxation Method", in *GLOBECOM 2023 - 2023 IEEE Global Communications Conference - 6G Communication Workshop*, Kuala Lumpur, Malaysia, Dec. 2023, pp. 1–7.

5.1 Introduction

Intelligent Reflective Surfaces (IRSs) have rapidly evolved to become a pivotal element in the evolution of modern wireless communication networks, attracting widespread attention from the research community for their remarkable ability to simplify deployment processes while substantially improving the quality of wireless signal propagation [157, 61, 65]. In a practical scenario, the IRS utilizes metamaterial reflection elements to redirect incoming signals toward desired directions, thereby enhancing performance in terms of coverage and achievable data-rate. As such, it is considered a crucial technique for deploying millimeter (mmWave) and sub-mmWave wave bands in future mobile networks [10]. IRSs reflect signals passively through beamforming without any need for signal decoding or amplification [64]. This characteristic sets IRSs apart from traditional relays, such as Decode-and-Forward (DF) or Amplify-and-Forward (AF) relays, which actively process and re-transmit received signals [176].

Distinguished by one of its many capabilities to operate in a full-duplex mode, IRS technology has proven instrumental in boosting both Spectral Efficiency (SE) and Energy Efficiency (EE), promising a significant leap forward for the development of Beyond-fifth-generation (B5G) wireless networks. The efficacy of IRS in elevating network performance through these dimensions has been thoroughly investigated, with a significant body of research dedicated to refining both active and passive beamforming techniques. Such endeavors aim to optimize SE and the weighted sum-rate, necessitating meticulous calibration of active beamforming at the Base Station (BS) and passive beamforming at the IRS to fully leverage the technology's benefits [160, 83].

Moreover, the synergy between Simultaneous Wireless Information and Power Transfer (SWIPT) and IRS technologies has catalyzed breakthroughs in energy efficiency. Notably, the research highlighted in [136] studies into the joint optimization of IRS phase shifts, BS active beamformers, and power-splitting ratios for users, illustrating the potential for an enhanced energy-efficient operation within IRS-enhanced SWIPT systems. Such studies not only illuminate the IRS's vital contribution to advancing wireless communications towards more ecologically sustainable practices but also reveal the intricate tradeoff required between amplifying signal quality and optimizing energy consumption. This comprehensive exploration of IRS technology underscores its transformative impact on the wireless communications landscape, positioning it as a key enabler for the future of connectivity.

Machine Type Communication (MTC) emerges as a cornerstone for the forthcoming surge in wireless communication advancements, with its applications broadly divided into massive MTC (mMTC) and ultra-reliable MTC (uMTC) domains. It is set to play a major role in the proliferation of next-generation technologies, including the Internet of Things (IoT), Internet of Vehicles (IoV), and Internet of Everything (IoE). These developments are anticipated to revolutionize the way connectivity is experienced across an extensive network of devices and platforms, pushing the boundaries of traditional wireless communication frameworks. The mMTC segment, in particular, is designed to support the expansion of future networks, enabling them to support a massive number of devices. This expansion is critical for ensuring efficient connectivity for countless devices that communicate via shorter packets [177]. However, this shift towards shorter packet transmissions presents a notable challenge to the conventional Shannon capacity formula, which struggles to accurately de-

pict the performance of these services [163]. This discrepancy has spurred the development of novel resource allocation strategies aimed at accommodating networks with devices that are delay-tolerant or those that meet the strict ultra-reliable low-latency (URLLC) criteria.

The exploration of resource allocation within MTC networks has led to significant research efforts. For example, a study achieved global optimal resource allocation for a URLLC system, focusing on optimizing bandwidth, power allocation, and antenna configuration to minimize the combined average power consumption for DownLink (DL) and UpLink (UL) communications [164]. Further investigations have sought to enhance network performance through various means, such as maximizing the sum throughput in a Multiple-Input Single-Output (MISO) Orthogonal Frequency Division Multiple Access (OFDMA) system via active beamforming vectors at the BS [165]. Other notable efforts include the precoder design at a BS for maximizing EE in multi-user Multiple-Input Multiple-Output (MIMO) networks employing finite blocklength codes [178], and the optimal design for energy-efficient MIMO-aided UL URLLC grant-free access systems [179]. Additionally, research has considered a hybrid approach of puncturing and superposition policies to simultaneously maximize the minimum average throughput for enhanced mobile broadband (eMBB) users and the number of supported URLLC users [180].

The integration of IRS into delay-insensitive systems stands as yet another groundbreaking approach to reducing computational latency, marking a significant stride towards heightened efficiency and enhanced performance in wireless networks. In particular, within OFDMA systems designed for URLLC services, the employment of IRS technology has been instrumental in substantially improving the weighted sum throughput. This improvement is achieved through a concerted effort in jointly optimizing active beamforming vectors along with phase shifts at both the BS and the IRS itself [166]. The utility of IRS extends into the realm of mobile edge computing systems as well, where its application has been explored with the objective of minimizing latency. This is accomplished by fine-tuning edge computing resources, computation offloading strategies, and beamforming matrices, demonstrating the IRS's capability to significantly impact system performance by reducing latency [167]. A further exploration of IRS technology within MTC, particularly in settings such as factory automation, has provided valuable insights. This research has shed light on the improvements in average data rates and the reduction in decoding error probabilities, especially when considering the transmission of short packets — a critical consideration in industrial applications where reliability and quick data transmission are paramount [111].

The strategic enablement of MTC services necessitates a concentrated effort to improve reliability while simultaneously expanding the capacity for a larger number of MTC/IoT users within networks. In this pursuit, the deployment of IRS has been identified as a promising avenue, offering a novel approach to enhance network capacity to accommodate more users significantly. Despite the potential benefits, there exists a noticeable research void concerning the application of IRS technology in systems enabled for MTC, particularly those systems characterized by the transmission of short packets. The specific goal of optimizing EE alongside increasing the count of fairly admitted IoT users within such frameworks remains an underexplored facet in the scholarly domain [165, 166, 178, 179, 180, 167, 111, 2]. This oversight signals a ripe opportunity for academic inquiry, pointing to the need for comprehensive studies that not only aim to integrate IRS platforms into MTC systems but also seek to balance EE improvements with the expansion of system capacity for IoT

users. Moreover, the challenge of upholding Quality of Service (QoS) standards within an IRS-supported IoT architecture, especially when dealing with the constraints of short packet communications, poses an intriguing research question. The inherent limitations associated with short packet transmissions — such as reduced data rates and potentially increased error probabilities — complicate the attainment of optimal QoS levels. Thus, investigating how IRS technology can be leveraged to mitigate these challenges, ensuring that QoS standards are not just maintained but enhanced, represents a critical area for future research endeavors.

This chapter is dedicated to tackling these issues, presenting a comprehensive study on resource allocation algorithm design for a DL MISO MTC-enabled IRS system. Our model encompasses a multi-antenna AP that serves multiple single-antenna IoT users, leveraging a smart reconfigurable reflector to facilitate communication. The innovation of our approach is dual-faceted. Firstly, we strive to achieve an optimal balance in maximizing the total EE of the system. Secondly, we integrate an effective admission control mechanism. This dual pursuit is both an academic exercise and a practical endeavor to explore the potential of IRS technology in MTC services. Through this comprehensive investigation, we aspire to illuminate the path forward in system design, highlighting how IRS technology can substantially enhance both the capacity and reliability of MTC services. By studying resource allocation in the context of IRS-assisted MTC systems, we seek to bridge the existing research gap, highlighting the potential of the IRS in enhancing network performance. Our exploration aims at the practical implications of integrating IRS technology, offering an understanding of how it can be harnessed to meet the rigorous demands of IoT applications, especially those requiring short packet lengths. This chapter has two parts. Consequently, the main contributions of this part of the chapter can be summarized as follows:

- We maximize the system's total EE together with admission control by jointly optimizing active and passive beamformers at the AP and IRS, respectively, subject to the minimum required data rate for each admitted IoT user with a short packet and unit-modulus constraints at the IRS.
- This problem is formulated as a Multi-Objective Optimization Problem (MOOP) which is a non-convex Mixed Integer Non-Linear Programming (MINLP) problem, and it is Non-deterministic Polynomial-time (NP) hard. To tackle this issue, we first convert it into a Single-Objective Optimization Problem via a weighting coefficient. Then, we exploit an Alternating Optimization (AO) resource allocation algorithm to solve the formulated optimization problem iteratively, which improves the objective function in each step. For the active beamformers at the BS, we first define a lower bound of the Signal-to-Interference-plus-Noise Ratio (SINR) and then apply the Difference of Convex (DC) functions and successive convex approximation (SCA) technique is adopted to make a concave-convex function and then the fractional problem is solved based on the quadratic transform which obtains a sub-optimal solution. Second, a penalty-based approach is adopted along with the SCA technique to handle the unit-modulus constraints at the IRS. In addition, an explicit objective is proposed to design a more efficient phase shift and provide a better convergence. In addition, an explicit objective is proposed to design a more efficient phase shift and provide a better convergence.
- The simulation results reveal that deploying an IRS can increase the system's EE and admission control of the IoT users with a short packet length. Results also reveal an

interesting tradeoff region between EE and user admissibility.

In summary, to stress the novelty of the first part of this chapter, we restate that we investigate a MISO MTC-enabled IRS system in this chapter. A multi-antenna AP transmits information symbols to a set of IoT users by taking into account short packet transmission. We serve a *target group of IoT users*, not any general user equipment (UE). We study a *new Multi-Objective Optimization Problem* that has never been analyzed in the literature to the best of the authors' knowledge. In particular, the total EE, together with the number of IoT users that could be served, is maximized by jointly optimizing active and passive beamformers at the AP and the IRS, respectively. An efficient algorithm based on AO is proposed to solve the main optimization problem iteratively. By the end of this chapter, we anticipate offering valuable insights and guidelines that underscore the efficacy of IRS technology in revolutionizing MTC systems. Our exploration is grounded in the belief that the strategic application of the IRS can meet the exacting demands of modern IoT applications.

As this chapter progresses, we delve into the evolution of MTC systems as they prove beneficial to sixth-Generation (6G) network paradigms. This transition is characterized by a heightened focus on supporting mission-critical applications that demand URLLCs, spanning diverse sectors such as wireless industrial automation and healthcare. The unique requirements of these applications — notably, their demand for exceptional reliability and strict adherence to delay-Quality of Service (d-QoS) standards — present new and complex challenges in radio access network design for MTC systems. These developments signal a critical need for innovative approaches that can help the demands of URLLC.

This chapter is structured as follows: In Section 5.2, we introduce the system model of an MTC-enabled IRS-assisted network finite blocklength transmission. Next, we explain the practical overhead design and channel estimation issues in Section 5.3. In Section 5.4, we propose our multi-objective maximization problem of total EE and admission control. The proposed problem will be solved in the subsequent Section 5.5. The computational complexity and the solution convergence of the MTC-enabled IRS-assisted network are discussed in Section 5.6, with results and a brief summary outlined in Section 5.7 and Section 5.8, respectively. In Section 5.9, we introduce the system model and outline the proposed EE optimization problem of a URLLC network. The resource allocation algorithm to solve the EE problem is detailed in Section 5.10 to enable URLLC service in an IRS-aided network. In Section 5.11, we assess the performance of our novel rank-one relaxation algorithm. Finally, Section 5.12 draws conclusions.

Notations: Matrices and vectors are denoted by boldface capital letters \mathbf{A} and lower case letters \mathbf{a} , respectively. For a square matrix \mathbf{A} , \mathbf{A}^T , \mathbf{A}^H , $\text{rank}(\mathbf{A})$, $\text{Tr}(\mathbf{A})$, and $\|\mathbf{A}\|_*$ are transpose, Hermitian conjugate transpose, rank of a matrix, trace, norm of a matrix, respectively. \mathbf{I}_N denotes the N -by- N identity matrix. $\text{diag}(\cdot)$ is the diagonalization operation. $\text{diag}(\mathbf{A})$ indicates a vector whose elements are extracted from the main diagonal elements of matrix \mathbf{A} . $\Re\{\cdot\}$ is used to denote the real part of a complex number. The absolute value of a complex scalar, and the Euclidean norm of a complex vector are expressed by $|\cdot|$ and $\|\cdot\|$, respectively. $\mathcal{CN}(\boldsymbol{\mu}, \mathbf{C})$ denotes the distribution of a circularly symmetric complex Gaussian (CSCG) random vector with mean $\boldsymbol{\mu}$ and covariance matrix \mathbf{C} . The largest eigenvalue of matrix \mathbf{X} is denoted by $\lambda_{\max}(\mathbf{X})$. $Q^{-1}(\cdot)$ stands for the inverse of the Gaussian Q-function. Moreover, $\mathbb{C}^{M \times N}$ represents an $M \times N$ dimensional complex matrix and ∇_x expresses the

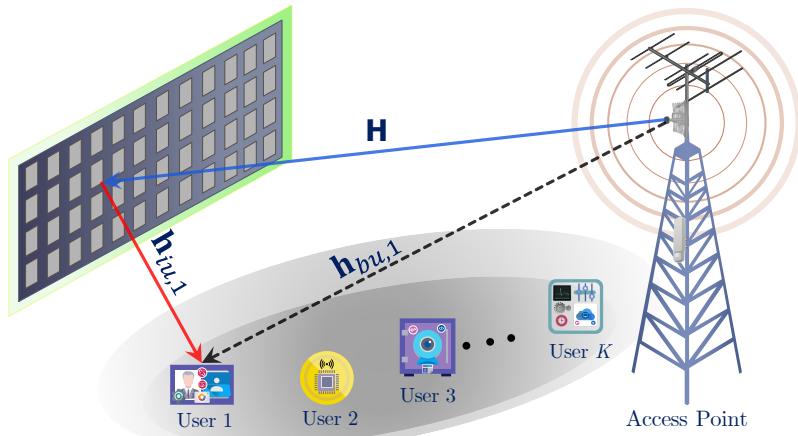


Figure 5.1: Energy efficiency and admission control of IoT users in an MTC-enabled IRS-assisted network with finite blocklength transmission.

gradient vector with respect to \mathbf{x} . Finally, we express $N \times N$ positive semidefinite matrices as $\mathbf{A} \in \mathbb{S}_+^N$ and read $\mathbf{A} \succeq \mathbf{0}$.

5.2 System Model of an MTC-enabled IRS-assisted network

In our system model, we consider a scenario comprising IoT users operating within an environment enhanced by an IRS system capable of supporting finite blocklength communications. This setup is depicted in Fig. 5.1 and includes an IRS equipped with N reflective elements, an AP that has M antennas, and K IoT users each having a single antenna. The ensemble of IRS elements is represented by the set $\mathcal{N} = \{1, \dots, N\}$, the array of AP antennas by $\mathcal{M} = \{1, \dots, M\}$, and the set of users by $\mathcal{K} = \{1, \dots, K\}$. For each user k within this network, a specific number of information bits, denoted as B_k , is allocated. These bits are then encoded by the AP into a codeword comprised of m_d symbols. The symbol sequence designated for user k is symbolized as $z_k[l]$, with l indexing the symbols within the sequence $\mathcal{L} = 1, 2, \dots, m_d$. The formulation of the transmit signal emanating from the AP, intended for sequential broadcast across the communication channel, can mathematically be written as:

$$\mathbf{s}[l] = \sum_{k \in \mathcal{K}} u_k \mathbf{w}_k z_k[l], \quad \forall l \in \mathcal{L}, \quad (5.1)$$

where $\mathbf{w}_k \in \mathbb{C}^{M \times 1}$ represents the beamforming vector for user k . This vector steers the transmitted signal toward the intended recipient, thereby optimizing the signal's integrity and ensuring efficient utilization of the available spectral resources. In our system model, the channel links between the AP, the IRS, and the IoT users are assumed to exhibit time-invariant characteristics, indicative of a slow fading environment. This assumption allows us to simplify our model by eliminating the time index l , thereby adopting a quasi-static flat-fading channel. In this context, the wireless channels are considered to remain consistent

throughout the duration of each transmission block, ensuring that the channel properties do not fluctuate within the scope of a single codeword transmission. The presence of a user in the system is indicated by setting $u_k = 1$ for the k -th IoT user, signifying that this user is actively being served. Conversely, a value of $u_k = 0$ implies the absence or dropping of the k -th user from the system's consideration in the current transmission block.

In our exploration, we account for a scenario wherein both the Channel State Information (CSI) and the specific delay requirements are precisely known at the AP. This knowledge base provides a strategic advantage, allowing for an in-depth understanding of the system's performance capabilities under the assumption of perfect CSI. It also paves the way for extrapolating insights into the theoretical upper-performance bounds for systems that operate under conditions of imperfect CSI. This consideration is vital as it acknowledges the real-world challenges and limitations encountered in wireless communication systems, where perfect CSI is often unattainable due to dynamic environmental conditions and inherent system noises. The significance of having accurate CSI and understanding delay constraints at the AP cannot be overstated. It enables the formulation of more efficient resource allocation strategies, beamforming vector optimizations, and reflective surface configurations, all tailored to meet the stringent requirements of IoT applications. Moreover, this approach facilitates a comprehensive evaluation of how deviations from the ideal CSI assumptions impact system performance, thereby offering a holistic view of potential performance thresholds. (see [83, 136, 160, 165, 2])

The system defines the baseband equivalent channel responses to demonstrate the complex interactions between the AP, the IRS, and each user. Specifically, $\mathbf{H} \in \mathbb{C}^{N \times M}$ represents the channel response from the AP to the IRS. The channel from the IRS to user k is denoted as $\mathbf{h}_{iu,k} \in \mathbb{C}^{N \times 1}$, and the direct channel from the AP to user k is represented by $\mathbf{h}_{bu,k} \in \mathbb{C}^{M \times 1}$. These definitions allow for a comprehensive description of the signal propagation paths within the system. Furthermore, the IRS's reflection-coefficients matrix is denoted as:

$$\Theta = \text{diag}(\beta_1 e^{j\alpha_1}, \beta_2 e^{j\alpha_2}, \dots, \beta_N e^{j\alpha_N}). \quad (5.2)$$

This matrix modulates the reflected signal's properties, with $\beta_n \in [0, 1]$ corresponding to the reflection amplitude and $\alpha_n \in (0, 2\pi]$, $\forall n \in \mathcal{N}$, indicating the phase shift imposed by the IRS¹. This mechanism of reflection-coefficient modulation is central to the IRS's ability to enhance the communication system's performance by optimally redirecting the incident signals toward the intended IoT users, thereby maximizing the system's overall efficiency and reliability in a slow-fading environment.

To model the impact of both the direct and the IRS-assisted paths on signal transmission, we define the equivalent channel link for each user k as follows:

$$\mathbf{h}_k^H \triangleq \mathbf{h}_{iu,k}^H \Theta \mathbf{h} + \mathbf{h}_{bu,k}^H, \quad \forall k \in \mathcal{K}. \quad (5.3)$$

This equation, (5.3), encapsulates the aggregate channel effect by combining the direct AP-to-user link and the AP-to-IRS-to-user reflected path, thereby creating a comprehensive representation of the communication channel. Given this equivalent channel, the signal

¹We consider continuous phase shifts, as discrete shifts cause misalignment of IRS-reflected and non-IRS-reflected signals, which degrades performance [157].

received by each user k can be expressed as:

$$y_k = \mathbf{h}_k^H \mathbf{s} + n_k \triangleq \sum_{k \in \mathcal{K}} u_k \mathbf{h}_k^H \mathbf{w}_k z_k + n_k, \forall k \in \mathcal{K}, \quad (5.4)$$

where the noise at the receiver is modeled as an Additive White Gaussian Noise (AWGN) variable, n_k , with zero mean and variance σ_k^2 and follows a circularly symmetric Gaussian distribution, denoted by $n_k \sim \mathcal{CN}(0, \sigma_k^2)$. This model takes into account the random nature of the communication channel's noise and its impact on the received signal. The MIMO experienced by user k is thus formulated as:

$$\gamma_k = \frac{u_k |\mathbf{h}_k^H \mathbf{w}_k|^2}{\sum_{i \neq k, i \in \mathcal{K}} u_i |\mathbf{h}_k^H \mathbf{w}_i|^2 + \sigma_k^2}, \forall k \in \mathcal{K}. \quad (5.5)$$

This equation, (5.5), highlights the balance between the desired signal power and the aggregate interference and noise, which is crucial for ensuring reliable communication. Addressing the need for low-latency and high-reliability in wireless communication for MTC-type IoT terminals, the concept of finite and short blocklength transmission is introduced. This approach is vital for applications requiring timely and dependable data exchange. The achievable data rate for each user, accounting for the finite blocklength regime, is precisely approximated as follows [163]:

$$R_k(u_k, \mathbf{w}_k, \Theta) = F(u_k, \mathbf{w}_k, \Theta) - G(u_k, \mathbf{w}_k, \Theta), \forall k \in \mathcal{K}, \quad (5.6)$$

where

$$F_k(u_k, \mathbf{w}_k, \Theta) = \log_2(1 + \gamma_k), \quad \forall k \in \mathcal{K}, \quad (5.7)$$

$$G_k(u_k, \mathbf{w}_k, \Theta) = Q^{-1}(\epsilon_k) \sqrt{\frac{1}{m_d} V_k}, \quad \forall k \in \mathcal{K}. \quad (5.8)$$

Furthermore, the decoding error probability for each user is denoted by ϵ_k , while the term m_d specifies the length of the block used in the transmission, serving as a key parameter in the context of finite blocklength communications. The concept of channel dispersion, V_k , further enriches our analysis by quantifying the variability of the channel's capacity and is mathematically expressed as:

$$V_k = a^2 (1 - (1 + \gamma_k)^{-2}), \forall k \in \mathcal{K}, \quad (5.9)$$

with $a = \log_2(e)$ acting as a scaling factor to translate natural logarithms to the base-2 logarithms, aligning with the bits measurement used in information theory. To ensure the QoS for each user, relevant to the aspects of reliability, latency, and the requisite number of received bits, we introduce a critical parameter: a *minimum threshold data rate*, R_{th}^k . This threshold guarantees that the communication service meets the specified performance criteria for each user, formalized as:

$$R_k(u_k, \mathbf{w}_k, \Theta) \geq R_{\text{th}}^k, \forall k \in \mathcal{K}. \quad (5.10)$$

This condition ensures that the system's design and operational protocols are aligned to satisfy the users' essential communication needs. Advancing our discussion to the domain of EE, we define it as the quotient of the total system data rate by the overall network power

consumption, measured in [bits/Joule]. This metric offers a holistic view of the system's performance, balancing throughput against the energy expenditure:

$$\mathcal{E}_{eff}(u_k, \mathbf{w}_k, \Theta) = \frac{\sum_{k \in \mathcal{K}} R_k(u_k, \mathbf{w}_k, \Theta)}{\sum_{k \in \mathcal{K}} u_k \|\mathbf{w}_k\|^2 + P_s + NP_d + P_c^{AP}}, \quad (5.11)$$

where P_s represents the static power consumption, necessary for sustaining the basic circuit operations of the IRS, and P_d accounts for the dynamic power dissipation per reflecting component, indicative of the energy required for adjusting the IRS elements. The term P_c^{AP} denotes the circuit power at the AP, encompassing the energy consumption intrinsic to the AP's operation.

This comprehensive framework is designed to study the complex balance between achieving high data rates and maintaining energy efficiency within the system, underlining the importance of sophisticated resource allocation and system design to meet the demands of modern wireless communication networks. In the subsequent analysis, our primary objective is to formulate an optimization problem that aims to maximize the EE of the system with an admission control mechanism. This formulation is carefully designed to accommodate the minimum data rate requirements of IoT users, particularly in scenarios characterized by short packet lengths. The core challenge lies in devising a strategy that not only enhances the EE of the network but also ensures that the QoS criteria, specifically in terms of data rate thresholds for each user, are met under the constraints of finite blocklength communications. Before delving deeper into the optimization problem, let's discuss the practicality of such a design.

5.3 Practical Overhead and CSI Estimation

The domain of IRS channel estimation is split into two predominant methodologies, contingent upon the IRS's inherent configuration capabilities—specifically, whether the IRS is equipped with sensing devices, such as receive RF chains. This distinction gives rise to two classifications: *semi-passive* IRS and *fully passive* IRS. The delineation between these configurations significantly influences the approach and feasibility of channel estimation strategies, as detailed in an array of studies [35, 181, 182, 183, 184, 157].

Semi-Passive IRS Channel Estimation

The semi-passive IRS model incorporates sensors that can directly acquire pilot signals from information users or the AP, facilitating the direct estimation of their corresponding channels to the IRS. Since IRS reflecting components and sensors are supposed to be in each other's proximity, the links between the information users or the AP and IRS elements can roughly be reconstructed from the estimated CSIs with sensors. The estimation process is carried out using the strong spatial correlation between IRS elements and sensors, employing advanced signal processing methods such as machine learning, data interpolation, and compressed sensing. This approach culminates in the application of the channel reciprocity theorem to

determine the CSI of the reverse channels, thereby providing a more direct and potentially accurate channel estimation method.

Fully Passive IRS Channel Estimation

In contrast, the fully passive IRS lacks any form of sensing capabilities, rendering direct channel estimation methods inapplicable. This means estimating the channels between AP and IRS separately from the links between IRS elements and the information users is not viable for a fully passive IRS. This configuration demands innovative strategies to estimate the cascaded channels involving the AP, the IRS, and the information users. One practical method involves sending training signals from the AP or information users and varying the IRS elements' reflection patterns over time to deduce the composite channel characteristics. However, this method faces a significant hurdle due to the large number of reflecting elements in typical IRS deployments, which could result in prohibitively high pilot/training signal overhead.

To mitigate this challenge, the concept of IRS element grouping emerges as a pragmatic solution. By grouping adjacent IRS elements into subsurfaces, the channel estimation process is simplified to determining the effective cascaded channel for each subsurface rather than each individual element. This strategy effectively reduces the channel estimation overhead, making it a viable approach for fully passive IRS configurations. These differing methodologies underscore the adaptability required in channel estimation techniques to align with the specific IRS configurations. While semi-passive IRS allows for a more straightforward estimation process through its onboard sensors, fully passive IRS necessitates creative solutions to overcome the limitations imposed by its lack of direct sensing capabilities. Both approaches, however, are instrumental in advancing the potential of IRS technology to enhance wireless communication networks, albeit through distinct pathways significantly. Generally, no RF chain is needed at the IRS for cascaded channel estimation since channel estimation is done at the AP. Otherwise, the IRS needs to be equipped with RF chains to estimate channels directly.

In the context of IRS-aided communication systems, especially those operating under finite blocklength regimes, the protocol for channel estimation and signal transmission is designed to ensure system efficiency and accuracy. This process, as delineated in studies such as [64, 185, 54], highlights the operational dynamics between the AP, the IRS, and the user devices, emphasizing the critical role of CSI in optimizing the system's performance. Here's an overview of the transmission protocol for an IRS-equipped system with receive RF chains:

Transmission Protocol Overview

1. **Uplink Pilot Transmission:** Initially, in a manner akin to Time Division Duplexing (TDD) protocols seen in massive MIMO systems, all receivers (i.e., user devices) send orthogonal pilot signals simultaneously to the AP in the uplink phase. This step is pivotal for acquiring the necessary CSI without interference.
2. **Channel Estimation:** Upon receiving these signals, the AP estimates the channels

between itself and the users (BS-user channels), while the IRS, equipped with receive RF chains, independently estimates the channels between itself and the users (IRS-user channels).

3. **Information and Channel Feedback:** Subsequently, the AP begins transmitting information symbols to the users. Concurrently, it communicates the estimated BS-user channels to the IRS using either optical links (suitable in indoor settings) or dedicated wireless links (for broader contexts, like outdoor environments or when an IRS controller is involved). This dual communication enables the IRS to perform joint optimization of the active beamforming vectors at the AP and the phase shifts on the IRS.
4. **Optimization and Adjustment:** The IRS controller, upon receiving the channel information and requirements, transmits the optimized active beamforming vectors back to the AP and adjusts the IRS's phase shifts accordingly to maximize system performance.
5. **Cooperative Transmission:** Finally, with the optimized parameters in place, the AP and the IRS cooperatively transmit information symbols to the receivers, ensuring enhanced signal quality and system efficiency through the joint beamforming effort.

Signaling Overhead and Channel Estimation Challenges

The signaling overhead in this IRS-aided system primarily consists of the complex numbers representing channel information that needs to be exchanged between the AP and the IRS for effective optimization. Accurately obtaining the CSI at both the AP and the IRS is paramount to harnessing the full potential of the IRS-aided system. However, practical challenges in achieving precise CSI necessitate reliance on channel reciprocity in TDD systems, allowing for downlink channel estimation based on uplink channel information by varying IRS reflection patterns. For multi-user systems, channel estimation overhead can be mitigated by utilizing the IRS to reflect simultaneously transmitted pilot signals from all users through the same IRS-AP channel, thus streamlining the process [183, 186]. Despite these advancements, the chapter adopts a simplified approach for clarity, assuming sequential user channel estimation in the uplink ². The beamforming designs proposed within this chapter remain adaptable to various channel estimation techniques, conditional on the availability of first and second-order statistics of the channel estimation errors. This adaptability underscores the robustness of the proposed methods in accommodating different operational environments and channel estimation accuracies.

In the following paragraphs, we calculate the overhead for a simple scenario of not having receiver noise at the AP. We avoid adding this part to our manuscript as it may hide away the main contribution of our study due to limited space.

²In this chapter, we assume that the downlink-uplink channel reciprocity holds; thus, the downlink channel can be learned by estimating its counterpart in the uplink by varying the IRS reflection patterns [187, 188].

5.3.1 Channel Estimation for an IRS-Assisted IoT Users with Finite Block Length

In line with the foundational assumptions established for our principal system model and depicted in Fig. 5.2 (modified version of Fig. 5.1), our examination of channel estimation revolves around a configuration where a set K single-antenna IoT users, with a finite block length communication capability, engage in simultaneous uplink interactions with an AP that has M antennas. This interaction is significantly enhanced by the strategic deployment of an N -element IRS, specifically introduced to amplify the communication efficacy across the user spectrum. Within the scope of this model, a critical assumption is made regarding the channel conditions, which are presumed to remain invariant over each fading block, typified by a length of T symbols.

The dynamics of the communication channels within this setup are encapsulated through several key definitions. The channel connecting each user k to the AP is precisely defined by $\mathbf{h}_{k,bu} \in \mathbb{C}^{M \times 1}$, offering a clear depiction of the direct communication link. Concurrently, the interaction between each user k and the IRS is characterized by $\mathbf{h}_{k,iu} = [h_{k,1}^{iu}, \dots, h_{k,N}^{iu}]^T \in \mathbb{C}^{N \times 1}$, while the IRS-to-AP channel is delineated as $\mathbf{R} = [\mathbf{r}_1, \dots, \mathbf{r}_N] \in \mathbb{C}^{M \times N}$. These channel representations are foundational to understanding the communication network's structure and the IRS's role within it.

The presence of the IRS introduces a unique dimension to the AP's received uplink signal at any given time instant i , within the range $1 \leq i \leq T$, manifesting as a composite of both the direct signals emanating from the users and those reflected via the IRS. This dual nature of signal reception underscores the impact of the IRS in mediating and enhancing the communication flow between the users and the AP. A crucial component of the IRS's operational mechanism is embodied in the reflection coefficients, $\theta_{n,i}$, which define the state of each IRS element at any given time instant i within the observed coherence block. These coefficients are binary, with a value of $|\theta_{n,i}| = 1$ indicating an active (on) state of the n^{th} IRS element, thereby enabling it to modify the phase of the incident signal, and $|\theta_{n,i}| = 0$ denoting an inactive (off) state. This binary framework for the IRS elements' operation is central to our understanding of how the IRS modulates the incident signals, enhancing the system's overall communication performance.

Expanding upon our system model for IRS-assisted uplink communications, we incorporate the conventional two-stage transmission protocol applicable within each coherence block, extending over T symbols. This protocol characterizes the coherence block into two distinct phases: a *channel estimation phase* that spans $\tau < T$ symbols, followed by a *data transmission phase* occupying the remaining $T - \tau$ symbols. This structured approach ensures a systematic allocation of resources towards both accurate channel estimation and efficient data transmission. During the channel estimation phase, an arrangement is set forth where each user k , across the spectrum of K users, is allocated a unique pilot sequence composed of τ symbols:

$$\mathbf{a}_k = [a_{k,1}, \dots, a_{k,\tau}]^T, \quad k = 1, \dots, K. \quad (5.12)$$

Here, the pilot symbols, the norm of $a_{k,i}$, is either zero or one, $\forall k, i$, ensuring a straightforward modulation scheme for each pilot symbol transmitted by the users, $\forall k, i$. One needs to properly design user pilot symbols $a_{n,i}$ and IRS reflection coefficients $\theta_{n,i}$'s so that the

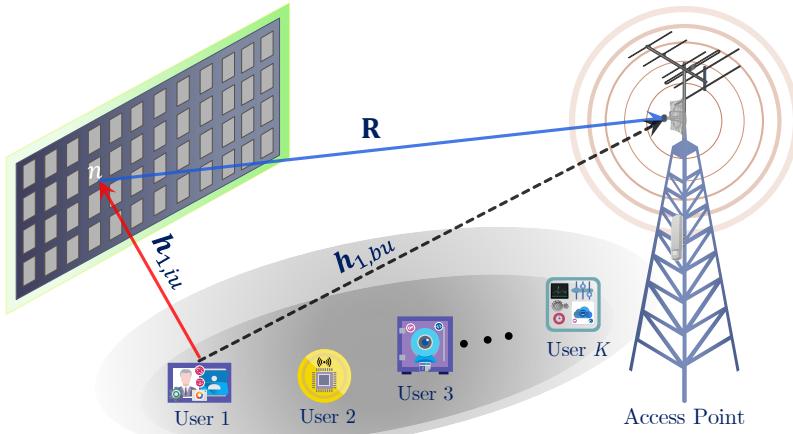


Figure 5.2: Channel estimation for an IRS-Assisted IoT users with finite block length.

AP can estimate the CSI. Therefore, we can propose a novel three-stage channel estimation protocol for IRS-assisted multi-user communications based on [183, 189, 190].

Initial Stage: Direct Channel Estimation

The protocol initiates with the first stage, where τ_1 symbols, denoted by:

$$\mathbf{a}_k = [a_{k,1}, \dots, a_{k,\tau_1}]^T, \quad k = 1, \dots, K, \quad (5.13)$$

are transmitted as pilot symbols. During this preliminary phase, the IRS assumes a passive stance, with all its elements deactivated ($\theta_{n,i} = 0$ for each element n , throughout the intervals $i = 1, \dots, \tau_1$). This strategic deactivation ensures the AP's received signal is devoid of IRS-mediated reflections, thereby isolating and accurately estimating the direct channels, $\mathbf{h}_{k,bu}$, from each user to the AP based on the received signals.

Second Stage: Reflective Channel Estimation of User '1'

Progressing to the second stage, a new set of τ_2 symbols, denoted by:

$$\mathbf{\ddot{a}}_k = [a_{k,\tau_1+1}, \dots, a_{k,\tau_1+\tau_2}]^T, \quad k = 1, \dots, K, \quad (5.14)$$

is employed as the pilot sequence for each user k . Unlike the initial stage, here, all IRS elements are activated ($\theta_{n,i} = 1$ for all n , across $i = \tau_1 + 1, \dots, \tau_1 + \tau_2$), setting the stage for the reflective channel estimation. User '1' is singled out to transmit non-zero pilot symbols, while the pilot sequences for users '2' through 'K' are nullified ($\mathbf{\ddot{a}}_k = \mathbf{0}$ for $k = 2, \dots, K$). This focused approach allows the AP, armed with the previously estimated $\mathbf{h}_{1,bu}$, to adeptly ascertain the IRS-reflected channels associated with user '1', symbolized as $\mathbf{g}_1 = h_{1,n}^U \mathbf{r}_1, \forall n$.

Final Stage: Comprehensive Reflective Channel Estimation

The concluding stage introduces τ_3 symbols, denoted by:

$$\tilde{\mathbf{a}}_k = [ak, \tau_1 + \tau_2 + 1, \dots, a_{k,\tau_1+\tau_2+\tau_3}]^T, \quad k = 1, \dots, K, \quad (5.15)$$

as the pilot sequence for each user k . In a departure from the previous stage, this phase is dedicated to users '2' through 'K', who transmit non-zero pilot sequences to the AP, facilitating the estimation of the remaining IRS-reflected channels, that is $\tilde{\mathbf{a}}_k = \mathbf{0}$. It appears that $(K - 1)MN$ unknowns exist that need to be estimated in $\mathbf{g}_{k,n}$'s, $k \geq 2$. Nonetheless, the number of unknowns can be highly reduced, drawing on the relationship between the user-IRS-AP reflected channels of user '1' and the other users.

Finally, for the simple scenario without receiver noise at the AP, the minimum theoretical overhead sequence length to perfectly estimate all the channel coefficients can be easily calculated to be [183]:

$$K + N + \max(K - 1, \lceil (K - 1)N/M \rceil). \quad (5.16)$$

Specifically, it can be proved that K and N time slots are sufficient to estimate the direct channels of all the users and IRS reflected channels of the user '1' in stage one and stage two, respectively, while $\max(K - 1, \lceil (K - 1)N/M \rceil)$ time slots are adequate for perfect channel estimation in the last stage. Surprisingly, the minimum overhead size reduces with M generally. Such a result strongly contrasts with the traditional multi-user channel estimation results without IRS, where the minimum overhead size is independent of the number of receiving antennas at the AP [189, 190].

5.3.2 Discrete IRS Phase Shifts

The implementation of IRS in enhancing wireless communication networks introduces a complex interaction between cost-efficiency and performance optimization. A critical aspect of this interplay revolves around the realization that, due to hardware limitations, the phase shifts facilitated by the reflecting elements of an IRS cannot feasibly exhibit continuous variability [191]. Instead, these phase shifts must be quantized into discrete levels, a necessity that aligns more closely with practical deployment scenarios and cost considerations, as highlighted by numerous studies [192, 185, 193, 188, 194].

The transition from an idealized model with continuous phase shifts to a more realistic framework incorporating discrete phase shifts brings forth certain performance implications. Specifically, the quantization of phase shifts into a finite number of discrete states can lead to a misalignment between the signals reflected by the IRS and those that bypass it, directly reaching the receivers. This misalignment, in turn, manifests as a degradation in the overall system performance.

The extent of this performance degradation, particularly in scenarios where the IRS comprises a large number of elements (theoretically extending towards infinity, $N \rightarrow \infty$), is quantifiably linked to the resolution of the phase shifters. The power loss associated with employing

IRS elements equipped with b -bit discrete phase shifters can be mathematically expressed as [188]:

$$1/\left(\frac{2^b}{\pi} \sin\left(\frac{\pi}{2^b}\right)\right)^2 \quad (5.17)$$

Notably, the utilization of one-bit phase shifters results in a power loss of approximately 3.9 dB in comparison to a scenario with near-ideal continuous phase shifts.

This scenario unveils an exciting cost-performance tradeoff, wherein the resolution of the phase shifters (2^b) and the quantity of reflecting elements (N) emerge as critical variables. Specifically, network designers are presented with the option to either increase the number of reflecting elements while compromising on phase shifter resolution or, conversely, to reduce the number of elements while opting for higher resolution shifters. This flexibility allows for strategic system design choices that aim to balance the received power at the IoT user against the economic and practical realities of IRS deployment.

In essence, the parameters N and b serve as instrumental levers in the design and optimization of IRS-assisted communication systems. By wisely selecting these parameters, system designers can solve the tradeoffs between manufacturing costs, the complexity of the IRS, and the desired level of communication performance. This nuanced approach to system design implies the importance of considering practical hardware constraints in the pursuit of enhancing wireless networks through IRS technology.

In our exploration of IRS-assisted wireless communication systems, we have bypassed the incorporation of discrete phase shifts within the scope of our current study. This decision is underpinned by a key consideration that directly influences the focus and outcomes of our exploration in this chapter.

The inclusion of discrete phase shifts, despite their practical relevance and cost-efficiency in real-world deployments, is deemed to offer minimal additional insight into the core objectives of our study. Drawing upon comprehensive analyses provided in existing literature [192, 185, 193, 188, 194], it becomes evident that the adoption of discrete phase shifts would invariably lead to a decrease in both the EE of the system and the number of users that can be effectively admitted into the network. Moreover, these adverse effects are not just speculative; they are quantifiable, with the extent of EE reduction and user admission capacity being directly correlated to the granularity of the phase shift resolution and the scale of the IRS deployment. Given this predictable outcome, we realize that the detailed examination of discrete phase shifts would not significantly enrich the primary thrust of our analysis of this chapter, which aims to determine the potential enhancements achievable through IRS technology under idealized conditions.

5.4 Multi-Objective Problem Formulation of EE and Admission Control

In this section, we investigate the formulation of a Multi-Objective Optimization Problem (MOOP) that seeks to concurrently maximize the total EE of the system and the number of fairly admitted IoT users. This ambitious goal necessitates a simultaneous optimization of

the active beamforming strategies at the AP and the phase shifts at the IRS. The formulation of this maximization optimization problem integrates a variety of constraints and objectives aimed at refining the system's performance in terms of EE and user admission control, embodying a comprehensive approach to system optimization. Accordingly, this problem as a MOOP can be mathematically formulated as follows:

$$P_1 : \max_{\mathbf{u}, \mathbf{w}_k, \Theta} \mathcal{E}_{eff}(u_k, \mathbf{w}_k, \Theta) \quad (5.18a)$$

$$\max_{\mathbf{u}, \mathbf{w}_k, \Theta} \sum_{k \in \mathcal{K}} u_k$$

$$s.t. : R_k(u_k, \mathbf{w}_k, \Theta) \geq u_k R_{th}^k, \quad \forall k \in \mathcal{K}, \quad (5.18b)$$

$$|\Theta_{nn}| = 1, \quad \forall n \in \mathcal{N}, \quad (5.18c)$$

$$\sum_{k \in \mathcal{K}} u_k \|\mathbf{w}_k\|^2 \leq p_{max}, \quad (5.18d)$$

$$u_k \in \{0, 1\}, \quad \forall k \in \mathcal{K}. \quad (5.18e)$$

where constraint (5.18b) ensures the reliability of each admitted MTC-type finite-blocklength user, a key requirement for maintaining the integrity of communication within the system. Constraint (5.18c) guarantees that the phase shift matrix, comprising N unit-modulus elements, adheres to the physical limitations and operational capabilities of the IRS, that is, N unit-modulus elements in the diagonal phase shift matrix. Moreover, constraint (5.18d) describes the transmission power budget limitation in which p_{max} is the maximum allowable transmission power. This constraint is crucial for ensuring that the system's operations remain within feasible and sustainable power consumption levels. Constraint (5.18e) indicates that u_k is a binary variable, where $\mathbf{u} = [u_1, \dots, u_K]$ constitutes the optimization decision vector, representing the admission status of all users within the system.

The optimization problem P_1 ³ is a non-convex MINLP due to the non-convexity of the objective function and the constraints, as well as incorporating binary variables in the objective and constraints. In general, finding an optimal solution for such a problem is impossible. However, in the next section, we adopt an approach to find an efficient sub-optimal solution.

Despite the non-convex MINLP nature of P_1 , the subsequent sections will explore a methodological approach designed to approximate an efficient sub-optimal solution. This approach will uncover algorithmic strategies to navigate the problem's complexity, aiming to identify solutions that significantly enhance the system's EE while maximizing user admission under the specified constraints. Through this endeavor, we aspire to achieve a balanced and optimized operational paradigm for IRS-assisted wireless communication systems, particularly in contexts characterized by finite blocklength communications and EE requirements.

³Please note P_1 ensures user fairness for a subset of the users, i.e., the total number of admitted users.

5.5 Proposed AO Solution

The optimization problem P_1 , characterized by its non-convex nature due to the highly coupled optimization variables, presents significant challenges for direct solution methodologies. Given the complexities inherent in P_1 , conventional optimization techniques may fall short in offering feasible solutions. To tackle this complexity, we introduce a novel solution strategy based on Alternating Optimization (AO), renowned for its efficacy in addressing problems marred by non-convexity and variable coupling with manageable computational demands.

Transitioning to a Single-Objective Optimization Problem (SOOP)

Our methodology begins by transforming the MOOP into a more tractable Single-Objective Optimization Problem (SOOP). This transformation is facilitated by the introduction of weighting coefficients, which serve to merge the distinct objectives of the MOOP into a unified objective function. These coefficients are carefully chosen to reflect the prioritization or preference between the original objectives, enabling a balanced consideration of EE and user admission within a singular optimization framework.

Decomposition into Sub-Problems

Following this consolidation into a SOOP, we further dissect the overarching problem into two manageable sub-problems, each focusing on specific aspects of the optimization:

1. **Active Beamformers and User Admission Optimization:** The first sub-problem concentrates on optimizing the active beamforming strategies at the AP and determining the admission status of potential users. To address the non-convexity and mixed-integer nature of this sub-problem, we employ a combination of the Big-M method, Semi-Definite Programming (SDP), and fractional programming techniques grounded in the quadratic transform. This approach allows for an effective optimization of the active beamformers while sensibly selecting the subset of users to be admitted, aligning with the system's capacity and EE objectives.
2. **Phase Shift Optimization:** The second sub-problem is dedicated to the optimization of the IRS phase shifts, a critical component in maximizing the system's reflective enhancement capabilities. The resolution of this sub-problem is achieved through the utilization of the penalty method and the Successive Convex Approximation (SCA) technique. These methods are proficient at navigating the unit-modulus constraints associated with the IRS phase shifts, facilitating an iterative refinement process that converges towards an optimal set of phase shift values.

Implementation and Iterative Refinement

The AO approach operates by iteratively solving these sub-problems, progressively refining the solution with each iteration. By alternately optimizing the active beamforming and user

admission strategy and then the IRS phase shifts, the method effectively decouples the originally intertwined variables, making the problem more tractable. This iterative process continues until convergence criteria are met, indicating that further iterations yield negligible improvements in the objective function.

Through this structured AO framework, we posit that a sub-optimal yet highly efficient solution to P_1 can be attained. This solution strategy not only overcomes the formidable barriers posed by the problem's non-convexity but also ensures that the system's performance is optimized across both EE and user admission dimensions.

5.5.1 First-stage: Optimizing \mathbf{w}_k with Fixed Θ

At this stage, when we fix the phase shifts at the IRS, denoted by Θ , our focus shifts to optimizing the active beamforming vectors, \mathbf{w}_k , at the AP, alongside the admission control decisions for the IoT users. The utilization of SDP facilitates this optimization by allowing us to represent the beamforming vectors in a semidefinite matrix form, with $\mathbf{W}_k = \mathbf{w}_k \mathbf{w}_k^H$ and subsequently the channels, $\mathbf{H}_k = \mathbf{h}_k \mathbf{h}_k^H$, $\forall k \in \mathcal{K}$.

To effectively manage the product of the binary variable u_k and the beamforming matrix \mathbf{W}_k , we introduce an auxiliary variable $\tilde{\mathbf{W}}_k$. This is achieved through the big-M method, which imposes additional constraints to overcome the complexities associated with binary optimization variables. These constraints are outlined as follows:

$$0 \preceq \tilde{\mathbf{W}}_k \preceq p_{\max} \mathbf{I}_M u_k, \quad \forall k \in \mathcal{K}, \quad (5.19a)$$

$$\mathbf{W}_k - (1 - u_k) p_{\max} \mathbf{I}_M \preceq \tilde{\mathbf{W}}_k \preceq \mathbf{W}_k, \quad \forall k \in \mathcal{K}. \quad (5.19b)$$

In the subsequent phase of optimization, the binary nature of the variable u_k is relaxed to a continuous spectrum, facilitating a more tractable optimization process. This relaxation is governed by the following constraints:

$$\sum_{k \in \mathcal{K}} u_k - \sum_{k \in \mathcal{K}} (u_k)^2 \leq 0, \quad (5.20)$$

$$0 \leq u_k \leq 1, \quad \forall k \in \mathcal{K}. \quad (5.21)$$

Also note that the data rate can be rewritten as: In the optimization of the active beamforming matrices with the IRS phase shifts held constant, the data rate for each user, denoted by $R_k(\tilde{\mathbf{W}}_k)$, is recast as the difference between two terms, $F_k(\tilde{\mathbf{W}}_k)$ and $G_k(\tilde{\mathbf{W}}_k)$ as follows:

$$R_k(\tilde{\mathbf{W}}_k) = F_k(\tilde{\mathbf{W}}_k) - G_k(\tilde{\mathbf{W}}_k), \quad \forall k \in \mathcal{K}. \quad (5.22)$$

Thus, the SINR γ_k , crucial to determining the quality and reliability of the communication link for each user, is subsequently formulated as:

$$\gamma_k = \frac{\text{Tr}(\mathbf{h}_k \tilde{\mathbf{W}}_k)}{\sum_{i \in \mathcal{K}, i \neq k} \text{Tr}(\mathbf{h}_k \tilde{\mathbf{W}}_i) + \sigma_k^2}, \quad \forall k \in \mathcal{K}. \quad (5.23)$$

Given that the constraint (5.18b) in P_1 — belonging to the reliability constraint of each admitted user based on their SINR levels — is inherently non-concave, we address this

challenge by introducing auxiliary variables $\xi_k, \forall k \in \mathcal{K}$. These variables serve as a lower bound to the SINR, facilitating a reformulation of the SINR constraint to accommodate the optimization process. Consequently, the SINR in (5.23) is redefined in a manner that enables its decomposition into a more tractable form, captured by the inequality:

$$0 \leq \xi_k \leq \gamma_k = \frac{f_k(\tilde{\mathbf{W}}_k)}{g_k(\tilde{\mathbf{W}}_k)}, \forall k \in \mathcal{K}, \quad (5.24)$$

where $f_k(\tilde{\mathbf{W}}_k)$ represents the numerator of the SINR equation, denoting the trace of the product between the user's channel matrix and their corresponding beamforming matrix. Meanwhile, $g_k(\tilde{\mathbf{W}}_k)$ encompasses the denominator, accounting for the aggregate interference from other users plus the noise floor, σ_k^2 , for each user k . The nominator and denominator of (5.24) can be expressed as:

$$f_k(\tilde{\mathbf{W}}_k) = \text{Tr}(\mathbf{h}_k \tilde{\mathbf{W}}_k), \quad \forall k \in \mathcal{K}, \quad (5.25)$$

$$g_k(\tilde{\mathbf{W}}_k) = \sum_{i \in \mathcal{K}, i \neq k} \text{Tr}(\mathbf{h}_k \tilde{\mathbf{W}}_i) + \sigma_k^2, \forall k \in \mathcal{K}, \quad (5.26)$$

respectively.

The optimization problem P_1 can now be restated with the introduction of a weighting coefficient $0 < \alpha < 1$, which indicates the relative importance of EE versus user admission. This weighted approach reinforces the multi-objective nature of our study, allowing for a harmonized optimization that does not singularly prioritize one objective over the other but seeks a balanced improvement across both dimensions. The constraints encapsulated within the modified optimization problem denoted as P_{10} encompass a wide array of system requirements, from ensuring minimum data rate thresholds for user connectivity to adhering to the AP's power budget. Moreover, the rank constraint on \mathbf{W}_k serves as a testament to the solution's feasibility within the physical constraints of beamforming technology. By exploiting the lower bound in (5.24), SDP, big-M, and the introduction of coefficient $0 < \alpha < 1$ that indicates the importance of the different objectives, the main optimization problem in the first stage can be recast as:

$$P_2 : \max_{u_k, \tilde{\mathbf{W}}_k, \mathbf{W}_k, \xi_k} \alpha \mathcal{E}_{eff}(\tilde{\mathbf{W}}_k, u_k) + (1 - \alpha) \sum_{k \in \mathcal{K}} u_k \quad (5.27a)$$

$$s.t. : 0 \leq \xi_k \leq \frac{f_k(\tilde{\mathbf{W}}_k)}{g_k(\tilde{\mathbf{W}}_k)}, \quad \forall k \in \mathcal{K}, \quad (5.27b)$$

$$R_k(\xi_k) \geq R_{th}^k, \quad \forall k \in \mathcal{K}, \quad (5.27c)$$

$$\text{rank}(\mathbf{W}_k) \leq 1, \quad \forall k \in \mathcal{K}, \quad (5.27d)$$

$$\sum_{k \in \mathcal{K}} \text{Tr}(\tilde{\mathbf{W}}_k) \leq p_{\max}, \quad (5.27e)$$

$$0 \preceq \tilde{\mathbf{W}}_k \preceq p_{\max} \mathbf{I}_M u_k, \quad \forall k \in \mathcal{K}, \quad (5.27f)$$

$$\mathbf{W}_k - (1 - u_k) p_{\max} \mathbf{I}_M \preceq \tilde{\mathbf{W}}_k \preceq \mathbf{W}_k, \forall k \in \mathcal{K}, \quad (5.27g)$$

$$\sum_{k \in \mathcal{K}} u_k - \sum_{k \in \mathcal{K}} (u_k)^2 \leq 0, \quad (5.27h)$$

$$0 \leq u_k \leq 1, \quad \forall k \in \mathcal{K}, \quad (5.27i)$$

where

$$R_k(\xi_k) = F_k(\xi_k) - G_k(\xi_k), \quad \forall k \in \mathcal{K}, \quad (5.28)$$

in which

$$F_k(\xi_k) = \log(1 + \xi_k), \quad \forall k \in \mathcal{K}, \quad (5.29)$$

$$G_k(\xi_k) = Q^{-1}(\epsilon_k) \sqrt{\frac{a^2}{m_d} \left(1 - (1 + \xi_k)^{-2}\right)}, \forall k \in \mathcal{K}. \quad (5.30)$$

P_2 is still a non-convex optimization problem. Addressing the inherent non-convexity of P_2 presents a significant challenge, requiring sophisticated mathematical maneuvers to reach a viable solution. In an effort to make this complex problem more tractable, we first reformulate the optimization problem into a canonical form amenable to DC programming. This reformulation is a strategic step that enables us to apply convex approximation techniques to the non-convex components of the problem, thereby simplifying the optimization landscape.

A key aspect of this approach involves addressing the constraint represented in (5.27b) by employing the first-order Taylor expansion to approximate the non-convex terms, thus converting them into a convex framework. Specifically, the constraint is re-expressed in a manner that clearly expresses the relationship between the auxiliary variable ξ_k and the function $g_k(\tilde{\mathbf{W}}_k)$, as follows:

$$\begin{aligned} \xi_k g_k(\tilde{\mathbf{W}}_k) &\leq f_k(\tilde{\mathbf{W}}_k) \\ \Rightarrow \xi_k A_k(\tilde{\mathbf{W}}_k) &\leq f_k(\tilde{\mathbf{W}}_k) - \xi_k \sigma_k^2, \forall k \in \mathcal{K}, \end{aligned} \quad (5.31)$$

where

$$A_k(\tilde{\mathbf{W}}_k) = \sum_{i \in \mathcal{K}, i \neq k} \text{Tr}(\mathbf{h}_k \tilde{\mathbf{W}}_i), \quad \forall k \in \mathcal{K}. \quad (5.32)$$

This representation in (5.31) shows the interaction between ξ_k and the summation of the trace operations across the non-self user beamforming matrices, denoted by $A_k(\tilde{\mathbf{W}}_k)$, within

the system's operational context. The non-convex nature of (5.31), stemming from the product of optimization variables, $\tilde{\mathbf{W}}_i$ and ξ_k , $\forall i, k \in \mathcal{K}$, poses a considerable challenge. However, we do this by decomposing this product into a DC form as follows:

$$\xi_k A_k(\tilde{\mathbf{W}}_k) = P_k(\xi_k, \tilde{\mathbf{W}}_k) - Q_k(\xi_k, \tilde{\mathbf{W}}_k), \forall k \in \mathcal{K}, \quad (5.33)$$

where

$$P_k(\xi_k, \tilde{\mathbf{W}}_k) = \frac{1}{2} \left(\xi_k + A_k(\tilde{\mathbf{W}}_k) \right)^2, \quad \forall k \in \mathcal{K}, \quad (5.34)$$

$$Q_k(\xi_k, \tilde{\mathbf{W}}_k) = \frac{1}{2} (\xi_k)^2 + \frac{1}{2} \left(A_k(\tilde{\mathbf{W}}_k) \right)^2, \forall k \in \mathcal{K}, \quad (5.35)$$

in which $P_k(\xi_k, \tilde{\mathbf{W}}_k)$ and $Q_k(\xi_k, \tilde{\mathbf{W}}_k)$ represent the convex and concave components, respectively. By denoting $\Omega_k = \{\xi_k, \mathbf{W}_k, \tilde{\mathbf{W}}_k, u_k\}$ as a set of optimization variables, we have:

$$U_k(\Omega_k) = P_k(\Omega_k) - Q_k(\Omega_k), \forall k \in \mathcal{K}. \quad (5.36)$$

Thus, P_2 can be recast as follows:

To effectively address the challenges posed by the non-convex optimization problem P_2 , we further refine our approach and introduce P_3 , a recast version that used the principles of DC programming to facilitate a more tractable solution approach. Thus, the updated optimizing problem can be given as follows:

$$P_3 : \max_{\Omega_k} \alpha \frac{\sum_{k \in \mathcal{K}} R_k(\xi_k)}{\sum_{k \in \mathcal{K}} \text{Tr}(\tilde{\mathbf{W}}_k) + P_s + NP_d + P_c^{\text{AP}}} \quad (5.37a)$$

$$+ (1 - \alpha) \sum_{k \in \mathcal{K}} u_k - \lambda \left(\sum_{k \in \mathcal{K}} (u_k - u_k^2) \right)$$

$$s.t. : U_k(\Omega_k) \leq f_k(\tilde{\mathbf{W}}_k, u_k) - \xi_k \sigma_k^2, \quad \forall k \in \mathcal{K}, \quad (5.37b)$$

$$R_k(\xi_k) \geq u_k R_{\text{th}}^k, \quad \forall k \in \mathcal{K}, \quad (5.37c)$$

$$\xi_k \geq 0, \quad \forall k \in \mathcal{K}, \quad (5.37d)$$

$$\text{rank}(\mathbf{W}_k) \leq 1, \quad \forall k \in \mathcal{K}, \quad (5.37e)$$

$$\sum_{k \in \mathcal{K}} \text{Tr}(\tilde{\mathbf{W}}_k) \leq p_{\text{max}}, \quad (5.37f)$$

$$0 \preceq \tilde{\mathbf{W}}_k \preceq p_{\text{max}} \mathbf{I}_M u_k, \quad \forall k \in \mathcal{K}, \quad (5.37g)$$

$$\mathbf{W}_k - (1 - u_k) p_{\text{max}} \mathbf{I}_M \preceq \tilde{\mathbf{W}}_k \preceq \mathbf{W}_k, \forall k \in \mathcal{K}, \quad (5.37h)$$

$$\sum_{k \in \mathcal{K}} u_k - \sum_{k \in \mathcal{K}} (u_k)^2 \leq 0, \quad (5.37i)$$

$$0 \leq u_k \leq 1, \quad \forall k \in \mathcal{K}, \quad (5.37j)$$

where λ is a large constant that acts as a penalty factor.

The objective function in P_3 seamlessly integrates the EE and user admission control objectives, complemented by a penalty term regulated by λ , to enforce user fairness and manage

the binary nature of user admission decisions. The constraints in P_3 are designed to encapsulate the requirements of the system. It should be noted that the objective function and constraints (5.37b) and (5.37c) belong to the class of DC problems. Thus, the SCA technique can be directly applied to approximate the non-convex problem in each iteration. Indeed, the objective function and constraints (5.37b) and (5.37c) are approximated by a more tractable one at a given local point. To this end, we use first-order Taylor expansion to obtain a globally lower-bound of functions $G_k(\xi_k)$ and $Q_k(\Omega_k)$, $\forall k \in \mathcal{K}$. By denoting ∇_{\square} as representing the gradient with respect to \square , the lower-bounds of these functions at iteration t are respectively given by:

$$G_k(\xi_k) \leq \tilde{G}_k(\xi_k) \triangleq G_k(\xi_k^t) + \partial_{\xi_k}^T G_k(\xi_k^t)(\xi_k - \xi_k^t), \forall k \in \mathcal{K}, \quad (5.38)$$

$$\begin{aligned} Q_k(\Omega_k) \geq \tilde{Q}_k(\Omega_k) \triangleq Q_k(\Omega_k^t) + \partial_{\Omega_k}^T Q_k(\Omega_k^t)(\Omega_k - \Omega_k^t) \\ + \text{Tr} \left(\nabla_{\tilde{\mathbf{W}}_k}^H Q_k(\Omega_k^t) (\tilde{\mathbf{W}}_k - \tilde{\mathbf{W}}_k^t) \right), \forall k \in \mathcal{K}. \end{aligned} \quad (5.39)$$

Therefore, we have:

$$\tilde{R}_k(\xi_k) = F_k(\xi_k) - \tilde{G}_k(\xi_k), \quad \forall k \in \mathcal{K}, \quad (5.40)$$

$$\tilde{U}_k(\Omega_k) = P_k(\Omega_k) - \tilde{Q}_k(\Omega_k), \quad \forall k \in \mathcal{K}. \quad (5.41)$$

Then, P_4 with any given local point at iteration t can be approximated as:

$$P_4 : \max_{\Omega_k} \alpha \frac{\sum_{k \in \mathcal{K}} \tilde{R}_k(\xi_k)}{E} + (1 - \alpha) \sum_{k \in \mathcal{K}} u_k \quad (5.42a)$$

$$- \lambda \left(\sum_{k \in \mathcal{K}} u_k - \left((u_k^t)^2 - 2u_k^t (u_k - u_k^t) \right) \right)$$

$$s.t. : \tilde{U}_k(\Omega_k) \leq f_k(\mathbf{W}_k, u_k) - \xi_k \sigma_k^2, \quad \forall k \in \mathcal{K}, \quad (5.42b)$$

$$\tilde{R}_k(\xi_k) \geq u_k R_{\text{th}}^k, \quad \forall k \in \mathcal{K}, \quad (5.42c)$$

$$\xi_k \geq 0, \quad \forall k \in \mathcal{K}, \quad (5.42d)$$

$$\text{rank}(\mathbf{W}_k) \leq 1, \quad \forall k \in \mathcal{K}, \quad (5.42e)$$

$$\sum_{k \in \mathcal{K}} \text{Tr}(\tilde{\mathbf{W}}_k) \leq p_{\max}, \quad (5.42f)$$

$$0 \preceq \tilde{\mathbf{W}}_k \preceq p_{\max} \mathbf{I}_M u_k, \quad \forall k \in \mathcal{K}, \quad (5.42g)$$

$$\mathbf{W}_k - (1 - u_k) p_{\max} \mathbf{I}_M \preceq \tilde{\mathbf{W}}_k \preceq \mathbf{W}_k, \quad \forall k \in \mathcal{K}, \quad (5.42h)$$

$$0 \leq u_k \leq 1, \quad \forall k \in \mathcal{K}, \quad (5.42i)$$

where,

$$E = \sum_{k \in \mathcal{K}} \text{Tr}(\tilde{\mathbf{W}}_k) + P_s + N P_d + P_c^{\text{AP}}. \quad (5.43)$$

The optimizing problem P_4 is still non-convex due to existence of a fractional term in the objective function. A common approach to handling fractional objective functions, like

the one in P_4 , is Dinkelbach's method. Dinkelbach's method, renowned for its efficacy in addressing optimization problems with fractional objective functions, transforms such functions into a more tractable subtractive form through the introduction of an auxiliary variable. This transformation facilitates the iterative solution of the problem by alternating between solving the transformed subtractive problem in an inner loop and updating the auxiliary variable in an outer loop. While Dinkelbach's method guarantees convergence to a global optimum under certain conditions, its applicability is limited to scenarios where the transformed subtractive problem is convex and involves only a single fractional objective function.

Nevertheless, suppose the transformed optimization problem in subtractive form is likewise nonconvex. Consequently, only a suboptimal solution to the inner optimization problem can be achieved with reasonable computational complexity, e.g., by employing the Weighted Sum Mean Square Error (WSMSE) method or Successive Convex Approximation (SCA). This is where the convergence of Dinkelbach's method cannot be assured. Furthermore, Dinkelbach's method can only tolerate a single fractional objective function. For example, $R_k(\mathbf{W}_k)$ in $\mathcal{E}_{eff}(\mathbf{W}_k)$ is a sum of logarithms of fractional functions, and therefore, Dinkelbach's method cannot be applied.

Thus, we exploit the fractional programming procedure [195], which can handle the fractional/multiplicative functions and even the function of fractional/multiplicative functions more flexibly. Similar to Dinkelbach's method, fractional programming also introduces auxiliary parameters to decouple the optimization variables and updates the optimization variables and auxiliary parameters iteratively. However, the adopted quadratic transformation for fractional programming is more flexible such that the resultant inner optimization problem is usually convex. Therefore, the fractional programming method is guaranteed to converge to a stationary point of the original optimization problem and enjoys a polynomial-time computational complexity. In the following, we first present a solution methodology for the optimization problem $\mathcal{E}_{eff}(\mathbf{W}_k)$ based on the quadratic transformation [195, 196], which can handle the severe variable coupling and can be readily used for developing a concrete algorithm for resource allocation policy.

The objective function in P_4 is in a format of concave-convex in which we use semidefinite relaxation (SDR) to remove the rank-one constraint (5.27d). In order to solve P_4 , we use the fractional programming method based on the quadratic transformation, which introduces an auxiliary parameter to transform a fractional form function into an equivalent subtractive form. To do so, we utilize the result of **Corollary 1** in [195] as follows:

Corollary 1 Consider f as a non-decreasing function, then the sum-of-ratio problem

$$\max_{\mathbf{x}} \frac{f_{Obj}(\mathbf{x})}{g_{Obj}(\mathbf{x})} \quad (5.44a)$$

$$s.t. : \mathbf{x} \in \mathcal{X}, \quad (5.44b)$$

is equivalent to the following problem

$$\max_{\mathbf{x}, m_{Obj}} 2m_{Obj} \sqrt{f_{Obj}(\mathbf{x}) - m_{Obj}^2 g_{Obj}(\mathbf{x})} \quad (5.45a)$$

$$s.t. : \mathbf{x} \in \mathcal{X}, m_{Obj} \in \mathbb{R}, \quad (5.45b)$$

where m_{Obj} is an auxiliary variable. The proof of the equivalence between (5.44) and (5.45) is provided in [195]. When $f_{Obj}(\mathbf{x})$ is a concave function with respect to \mathbf{x} in a convex set \mathcal{X} , the subtractive function $2m_{Obj}\sqrt{f_{Obj}(\mathbf{x})} - m_{Obj}^2 g_{Obj}(\mathbf{x})$ would be a concave function with respect to \mathbf{x} . Consequently, the resulting problem in (5.45) is a convex optimization problem for a given m_{Obj} . Finally, we note that the optimal auxiliary variable is given by:

$$m_{Obj} = \frac{\sqrt{f_{Obj}(\mathbf{x})}}{g_{Obj}(\mathbf{x})}. \quad (5.46)$$

Thus, we can develop an iterative algorithm with a polynomial-time computational complexity to update \mathbf{x} and m_{Obj} alternately. However, the algorithm is only guaranteed to converge to a sub-optimal solution of the main problem in (5.45) if the transformed problem in (5.44) can globally be solved [195].

In the following, we demonstrate how to execute the quadratic transformation to achieve a sub-optimal solution of P_4 . The problem P_4 can be transformed into the following equivalent optimization problem by adopting the quadratic transformation in (5.44) and (5.45):

$$P_5 : \max_{\Omega_k, m_{Obj}} \alpha \left(2m_{Obj} \sqrt{\sum_{k \in \mathcal{K}} \tilde{R}_k(\xi_k)} - m_{Obj}^2 E \right) \quad (5.47a)$$

$$+ (1 - \alpha) \sum_{k \in \mathcal{K}} u_k - \lambda \left(\sum_{k \in \mathcal{K}} u_k - \left((u_k^t)^2 - 2u_k^t (u_k - u_k^t) \right) \right)$$

$$s.t. : \tilde{U}_k(\Omega_k) \leq f_k(\mathbf{W}_k, u_k) - \xi_k \sigma_k^2, \quad \forall k \in \mathcal{K}, \quad (5.47b)$$

$$\tilde{R}_k(\xi_k) \geq u_k R_{th}^k, \quad \forall k \in \mathcal{K}, \quad (5.47c)$$

$$\xi_k \geq 0, \quad \forall k \in \mathcal{K}, \quad (5.47d)$$

$$\sum_{k \in \mathcal{K}} \text{Tr}(\tilde{\mathbf{W}}_k) \leq p_{\max}, \quad (5.47e)$$

$$0 \preceq \tilde{\mathbf{W}}_k \preceq p_{\max} \mathbf{I}_M u_k, \quad \forall k \in \mathcal{K}, \quad (5.47f)$$

$$\mathbf{W}_k - (1 - u_k) p_{\max} \mathbf{I}_M \preceq \tilde{\mathbf{W}}_k \preceq \mathbf{W}_k, \forall k \in \mathcal{K}, \quad (5.47g)$$

$$0 \leq u_k \leq 1, \quad \forall k \in \mathcal{K}, \quad (5.47h)$$

where m_{Obj} denotes the new auxiliary variable corresponding to the objective function of the optimization problem in P_5 and can be updated globally as:

$$m_{Obj} = \frac{\sqrt{\sum_{k \in \mathcal{K}} \tilde{R}_k(\xi_k)}}{E}. \quad (5.48)$$

The resulting subtractive function in (5.47) is concave with respect to the optimization variables for given auxiliary variables. Generally, P_5 yields a solution with a rank higher than one due to constraint (5.27d). Therefore, to solve (5.47) for a given m_{Obj} , we use the SDR to remove constraint (5.27d)⁴. The resulting problem is now a convex Semi-Definite

⁴In essence, the quadratic transformation methodology provides a robust framework for tackling the

Programming (SDP) problem that numerical convex solvers, such as CVX, can solve. Finally, the solution to the relaxed SDP problem is optimal if it satisfies constraint (5.27d), i.e., $\text{rank}(\mathbf{W}_k) \leq 1$.

To validate our methodology, we investigate the SDP relaxation's tightness in the following theorem.

Theorem 1 For a given p_{\max} and m_{Obj} , and by assuming that the channel vectors of all users, \mathbf{h}_k , $\forall k \in \mathcal{K}$ are mutually statistically independent, the optimal beamforming matrices \mathbf{W}_k^* of the relaxed version of the problem in P_5 are rank-one, i.e., $\text{rank}(\mathbf{W}_k) \leq 1$, $\forall k \in \mathcal{K}$ with probability one.

Proof 2 The SDP relaxed problem is in P_5 with respect to the remaining optimization variables and satisfies Slater's constraint qualification. Therefore, strong duality holds, and solving the dual problem is equivalent to solving the primal problem. Therefore, one can prove **Theorem 1** by exploiting the Karush–Kuhn–Tucker (KKT) conditions of P_5 [197]. ■

We can now rewrite the constraint in a mathematically tractable form via the DC method represented as:

$$\|\mathbf{W}\|_* - \|\mathbf{W}\|_2 \leq 0. \quad (5.49)$$

Note that $\|\mathbf{W}\|_* = \sum_i \sigma_i \geq \|\mathbf{W}\|_2 = \max_i \{\sigma_i\}$ holds for any given $\mathbf{W} \in \mathbb{H}^{M \times M}$, where σ_i is the i -th singular value of \mathbf{W} . The equality holds if and only if \mathbf{W} achieves rank one i.e., $\text{rank}(\mathbf{W}) = 1$ [2]. Now, we take the first-order Taylor approximation of $\|\mathbf{W}\|_2$ as:

$$\|\mathbf{W}\|_2 \geq \overbrace{\|\mathbf{W}^{(t)}\|_2 + \text{Tr} \left(\lambda_{\max} \left(\mathbf{W}^{(t)} \right) \lambda_{\max}^H \left(\mathbf{W}^{(t)} \right) (\mathbf{W} - \mathbf{W}^t) \right)}^{=\phi(\mathbf{W})}. \quad (5.50)$$

By resorting to (5.50), a convex approximation can be obtained for (5.49) which is given by:

$$\tilde{\phi}^t(\mathbf{W}) \triangleq \|\mathbf{W}\|_* - \phi(\mathbf{W}) \leq 0. \quad (5.51)$$

As a result, by augmenting $\tilde{\phi}^t(\mathbf{W})$ to the objective function of P_6 with $\psi \gg 1$ as a penalty factor to penalize any non-rank-one matrix, the optimization problem in the $(t+1)$ -iteration can be written as follows:

optimization challenges presented in P_4 . By transforming the problem into P_5 and iteratively solving for the auxiliary variable m_{Obj} , we establish a concrete algorithmic methodology to achieving a sub-optimal solution that effectively balances EE with the practical considerations of user admission and system constraints. This approach illustrates the potential of fractional programming in optimizing complex communication systems.

$$P_6 : \max_{\Omega_k} \alpha \left(2m_{Obj} \sqrt{\sum_{k \in \mathcal{K}} \tilde{R}_k(\xi_k)} - m_{Obj}^2 E \right) + (1 - \alpha) \sum_{k \in \mathcal{K}} u_k - \lambda \left(\sum_{k \in \mathcal{K}} u_k - \left((u_k^t)^2 - 2u_k^t (u_k - u_k^t) \right) \right) - \psi(\tilde{\phi}^t(\mathbf{W})) \quad (5.52a)$$

$$s.t. : \tilde{U}_k(\Omega_k) \leq f_k(\mathbf{W}_k, u_k) - \xi_k \sigma_k^2, \quad \forall k \in \mathcal{K}, \quad (5.52b)$$

$$\tilde{R}_k(\xi_k) \geq u_k R_{th}^k, \quad \forall k \in \mathcal{K}, \quad (5.52c)$$

$$\xi_k \geq 0, \quad \forall k \in \mathcal{K}, \quad (5.52d)$$

$$\sum_{k \in \mathcal{K}} \text{Tr}(\tilde{\mathbf{W}}_k) \leq p_{\max}, \quad (5.52e)$$

$$0 \preceq \tilde{\mathbf{W}}_k \preceq p_{\max} \mathbf{I}_M u_k, \quad \forall k \in \mathcal{K}, \quad (5.52f)$$

$$\mathbf{W}_k - (1 - u_k) p_{\max} \mathbf{I}_M \preceq \tilde{\mathbf{W}}_k \preceq \mathbf{W}_k, \quad \forall k \in \mathcal{K}, \quad (5.52g)$$

$$0 \leq u_k \leq 1, \quad \forall k \in \mathcal{K}. \quad (5.52h)$$

Consequently, P_6 is a convex optimization problem and can be efficiently solved.

5.5.2 Second-stage: Optimizing Θ

In the second stage of our optimization process, the focus is on the optimization of the phase shifts at the IRS, denoted as Θ , with the active beamformers $\tilde{\mathbf{W}}_k$ being held constant from the previous stage. This stage of optimization specifically targets the maximization of the data rate, a key performance indicator for the effectiveness of IRS-assisted communication systems. However, the optimization of Θ is notably challenged by the unit-modulus constraint, as expressed in constraint (5.18c), which presents the feasible set for the phase shifts and significantly complicates the optimization process.

To address the complexities introduced by this constraint, we employ a strategic reparameterization of the problem. This approach involves defining a new vector \mathbf{v} that encapsulates the phase shifts at the IRS, with each element being the exponential representation of the phase shift. Accordingly, we first define:

$$\mathbf{v} = (e^{j\alpha_1}, \dots, e^{j\alpha_N})^H \in \mathbb{C}^{N \times 1}, \quad (5.53)$$

$$\tilde{\mathbf{v}} = [\mathbf{v}^T \ \tau]^T \in \mathbb{C}^{(N+1) \times 1}, \quad (5.54)$$

where $\tau \in \mathbb{C}$ is a dummy variable with $|\tau| = 1$, to accommodate the unit-modulus constraint within our optimization framework in a more tractable manner. We note that the augmented vector $\tilde{\mathbf{v}}$ is defined to combine the phase shift vector \mathbf{v} and the dummy variable τ . By extending the dimensionality of the problem with the inclusion of τ , we create additional flexibility that can be exploited during the optimization process, thereby enhancing the potential for finding a viable solution to the originally intractable problem. To facilitate the solution design, we also define:

$$\mathbf{V} = \tilde{\mathbf{v}} \tilde{\mathbf{v}}^H \in \mathbb{C}^{(N+1) \times (N+1)}, \quad (5.55)$$

which indicates that the matrix \mathbf{V} is semi-definite and satisfies $\text{rank}(\mathbf{V}) \leq 1$. Thus, we obtain:

$$\begin{aligned} \left| (\mathbf{h}_{iu,k}^H \Theta \mathbf{h} + \mathbf{h}_{bu,k}^H) \tilde{\mathbf{W}}_k \right|^2 &\triangleq \text{Tr}(\mathbf{V} \mathbf{X}_k \tilde{\mathbf{W}}_k \mathbf{X}_k^H) \\ &= \text{Tr}(\tilde{\mathbf{W}}_k \mathbf{Y}_k), \quad \forall k \in \mathcal{K}, \end{aligned} \quad (5.56)$$

where

$$\begin{aligned} \mathbf{X}_k &= \left[(\text{diag}(\mathbf{h}_{iu,k}^H) \mathbf{H})^T \quad \mathbf{h}_{bu,k}^* \right]^T, \quad \forall k \in \mathcal{K}, \\ \mathbf{Y}_k &= \mathbf{X}_k^H \mathbf{V} \mathbf{X}_k, \quad \forall k \in \mathcal{K}. \end{aligned} \quad (5.57)$$

In addressing the challenges posed by the non-convex data rate constraint and the new objective function for the optimization of the phase shifts at the IRS, Θ , we embark on a strategic approach similar to that employed in the optimization of the active beamforming vectors. This strategy involves the clever use of auxiliary variables, in this case, denoted as (γ_k) , to facilitate the application of the SCA method, thus enabling the iterative refinement of the solution towards optimality. The application of auxiliary variables serves to effectively linearize the non-linear aspects of the optimization problem, thereby rendering it more tractable. Specifically, the data rate for each user k , now denoted as $\tilde{R}_k(\gamma_k)$, is expressed as a difference between two terms: This means:

$$\tilde{R}_k(\gamma_k) = F_k(\gamma_k) - \tilde{G}_k(\gamma_k), \quad \forall k \in \mathcal{K}. \quad (5.58)$$

Given this foundation, the optimization problem concerning the IRS phase shifts can be reformulated, taking into consideration the simplifications and assumptions applicable to this stage of the optimization process. Specifically, contributions from user admission variables u_k and total power considerations, which have been addressed in the preceding sub-problem, are omitted from the objective function to focus solely on the optimization of Θ . Now, we restate the optimization problem as follows:

$$P_7 : \max_{\mathbf{V}, \gamma_k} \sum_{k \in \mathcal{K}} \tilde{R}_k(\gamma_k) \quad (5.59a)$$

$$\text{s.t.} : \gamma_k \geq 0, \mathbf{V} \succeq \mathbf{0}, \Omega_k = \{\gamma_k, \mathbf{V}\}, \quad \forall k \in \mathcal{K}, \quad (5.59b)$$

$$\tilde{R}_k(\gamma_k) \geq u_k R_{\text{th}}^k, \quad \forall k \in \mathcal{K}, \quad (5.59c)$$

$$\tilde{U}_k(\Omega_k) \leq f_k(\mathbf{V}) - \xi_k \sigma_k^2, \quad \forall k \in \mathcal{K}, \quad (5.59d)$$

$$\text{rank}(\mathbf{V}) \leq 1. \quad (5.59e)$$

Similar to P_6 , P_7 usually does not give a rank-one solution because of constraint (5.59e). By rewriting (5.59e) as:

$$\|\mathbf{V}\|_* - \|\mathbf{V}\|_2 \leq 0, \quad (5.60)$$

and owing to (5.50), a convex approximation, $\tilde{\phi}^t(\mathbf{V}) \leq 0$, of rank-one constraint can be made. Thus, supplementing $\tilde{\phi}^t(\mathbf{V})$ to the objective function of P_8 with $\zeta \gg 1$ as a penalty

Algorithm 4 Proposed Algorithm for IRS-Based Energy Efficiency and Admission Control Maximization for IoT Users With Short Packet Lengths

Input: Set $m_{Obj}^{(0)}$, I_{\max} , and D_{\max} .

- 1: **repeat**
- 2: Calculate $\tilde{G}_k(\xi_k)$, $\tilde{Q}_k(\xi_k, \tilde{\mathbf{W}}_k)$, and $\tilde{\phi}^2(\mathbf{W})$ via a Successive Convex Approximation (SCA) structure.
- 3: Solve P_6 for a given Θ , and $m_{Obj}^{(d-1)}$.
- 4: **if** $|\sqrt{\sum_{k \in \mathcal{K}} \tilde{R}_k^{(d)}(\xi_k)} - m_{Obj}^{(d-1)} - E^{(d)}| \leq \varepsilon$
- 5: **return** $\Omega = \Omega^{(d)}$, $m_{Obj}^* = m_{Obj}^{(d-1)}$.
- 6: **else** Update $m_{Obj}^{(d)} = \frac{\sqrt{\sum_{k \in \mathcal{K}} \tilde{R}_k(\xi_k)}}{E}$, **end if**.
- 7: $d = d + 1$.
- 8: **until** $d = D_{\max}$
- 9: Calculate $\tilde{G}_k(\mathbf{Y}_k)$ and $\tilde{\phi}^2(\mathbf{V})$ via an SCA structure.
- 10: Solve P_8 for the obtained $\tilde{\mathbf{W}}$, u_k , from the previous steps.
- 11: $i = i + 1$
- 12: **until** $i = I_{\max}$.
- 13: **return** \mathbf{Y}, \mathbf{V} .

factor to penalize any non-rank-one matrix, the optimization problem in the $(t+1)$ -iteration can be written as follows:

$$P_8 : \max_{\mathbf{V}, \mathbf{Y}_k, k \in \mathcal{K}} \sum_{k \in \mathcal{K}} \tilde{R}_k(\mathbf{Y}_k) - \zeta(\tilde{\phi}^t(\mathbf{V})), \quad (5.61a)$$

$$\text{s.t.} : \mathbf{Y}_k \geq 0, \mathbf{V} \succeq \mathbf{0}, \Omega_k = \{\mathbf{Y}_k, \mathbf{V}\}, \quad \forall k \in \mathcal{K}, \quad (5.61b)$$

$$\tilde{R}_k(\mathbf{Y}_k) \geq u_k R_{\text{th}}^k, \quad \forall k \in \mathcal{K}, \quad (5.61c)$$

$$\tilde{U}_k(\Omega_k) \leq f_k(\mathbf{V}) - \xi_k \sigma_k^2, \quad \forall k \in \mathcal{K}, \quad (5.61d)$$

The optimization problem P_8 , structured to address the phase shift optimization at the IRS, Θ , with considerations for the non-convex constraints and auxiliary variables, mirrors the analytical strategy employed in solving P_6 . This mirrored approach underscores the versatility of our algorithmic framework, which can overcome the complexities inherent in both active beamforming and phase shift optimization within IRS-assisted MTC-enabled wireless communication systems. The algorithmic implementation of this strategy is summarized in **Algorithm 4**, a procedural blueprint that gives the iterative steps required to achieve convergence to a locally optimal solution.

5.6 Computational Complexity and Convergence Analysis

In the computational complexity and convergence analysis of our proposed algorithm, we study the mathematical intricacies that define the computational demand of solving the optimization problems, specifically P_6 and P_8 . This analysis provides insight into the efficiency and scalability of the algorithm, crucial aspects when considering its practical application.

The computational complexity of solving an SDP problem, which is central to our optimization approach, is determined by several factors, including the size of the problem and the desired solution accuracy. For an SDP problem characterized by m constraints and involving an $n \times n$ Positive Semi-Definite (PSD) matrix, the computational complexity is generally expressed as:

$$\mathcal{O}(\sqrt{n} \log(1/\zeta)(mn^3 + m^2n^2 + m^3)), \quad (5.62)$$

where $\zeta > 0$ is the solution accuracy [2]. This formula shows the connection between the problem size, the number of constraints, and the accuracy of the solution, offering a comprehensive view of the computational demand. Applying (5.62) to the optimization problem P_6 , with $m = 6K + 1$ constraints and an $n = M$ PSD matrix, yields the following complexity order:

$$O_1 = \mathcal{O} \log \left(\frac{1}{\zeta_1} \right) (6K + 1) ((M)^{3.5} + (6K + 1)^{2.5} M^2 + (6K + 1)^2), \quad (5.63)$$

where ζ_1 is the solution accuracy specific to P_6 . This provides a quantitative measure of the computational resources required to solve the problem, highlighting the impact of the number of users (K) and the size of the active beamforming matrix (M). Similarly, for the optimization problem P_8 , which focuses on the phase shift optimization at the IRS, the complexity can be formulated as:

$$O_2 = \mathcal{O} \log \left(\frac{1}{\zeta_2} \right) (4K + 1) ((N)^{3.5} + (4K + 1)^{2.5} N^2 + (4K + 1)^2), \quad (5.64)$$

with ζ_2 representing the solution accuracy for P_8 . This expression determines the computational demands associated with optimizing the IRS phase shifts, emphasizing the role of the number of IRS elements (N) and users (K).

The overall computational complexity of the proposed solution approach is thus a function of the complexities of solving P_6 and P_8 , scaled by the number of iterations, I_{iter} , required for the AO method to converge. This yields an aggregate complexity of:

$$\mathcal{O}_{\text{tot}} = \mathcal{O}(I_{\text{iter}}(O_1 + O_2)), \quad (5.65)$$

offering a comprehensive overview of the computational demands of the algorithm across both stages of optimization. Through this analysis, we gain valuable insights into the computational implications of employing the proposed AO-based algorithm in IRS-assisted MTC-enabled systems.

In the following, we also prove that our algorithm is convergent.

Proposition 2 *The objective function value of P_1 would be improved via this iterative algorithm.*

Proof 3 Let us consider $\{\mathbf{W}^{(j)}, \mathbf{u}^{(j)}, \Theta^{(j+1)}\}$ as the feasible solution set to P_8 . Then, the feasible solution set of P_8 is a feasible solution to P_1 as well. Therefore, $\{\mathbf{W}^{(j)}, \mathbf{u}^{(j)}, \Theta^{(j)}\}$ and $\{\mathbf{W}^{(j+1)}, \mathbf{u}^{(j+1)}, \Theta^{(j+1)}\}$ are feasible to P_1 in the (j) -th and $(j+1)$ -th iterations, respectively. Now, we define $f_{P_1}(\mathbf{W}^{(j)}, \mathbf{u}^{(j)}, \Theta^{(j)})$, $f_{P_8}(\Theta^{(j)})$, and $f_{P_6}(\mathbf{W}^{(j)}, \mathbf{u}^{(j)})$ as the objective functions of problem P_1 , P_8 and P_6 in the (j) -th iteration, respectively. Thus, we have

$$\begin{aligned} & f_{P_1}(\mathbf{W}^{(j+1)}, \mathbf{u}^{(j+1)}, \Theta^{(j+1)}) \\ & \stackrel{(a)}{=} f_{P_8}(\Theta^{(j+1)}) \stackrel{(b)}{\geq} f_{P_8}(\Theta^{(j)}) \\ & = f_{P_1}(\mathbf{W}^{(j+1)}, \mathbf{u}^{(j+1)}, \Theta^{(j)}), \end{aligned} \quad (5.66)$$

where (a) follows the fact that problem P_1 is equivalent to problem P_8 for optimal \mathbf{W} and \mathbf{u} , and (b) holds since $f_{P_8}(\Theta^{(j+1)}) \geq f_{P_8}(\Theta^{(j)})$ according to sub-problem 2 (that is the second-stage: Optimizing Θ). Similarly, for a given $\Theta^{(j)}$, we have

$$\begin{aligned} & f_{P_1}(\mathbf{W}^{(j+1)}, \mathbf{u}^{(j+1)}, \Theta^{(j)}) \\ & \stackrel{(a)}{=} f_{P_6}(\mathbf{W}^{(j+1)}, \mathbf{u}^{(j+1)}) \stackrel{(b)}{\geq} f_{P_6}(\mathbf{W}^{(j)}, \mathbf{u}^{(j)}) \\ & = f_{P_1}(\mathbf{W}^{(j)}, \mathbf{u}^{(j)}, \Theta^{(j)}). \end{aligned} \quad (5.67)$$

From the above two inequalities, we can conclude the following inequality holds

$$f_{P_1}(\mathbf{W}^{(j+1)}, \mathbf{u}^{(j+1)}, \Theta^{(j+1)}) \geq f_{P_1}(\mathbf{W}^{(j)}, \mathbf{u}^{(j)}, \Theta^{(j)}). \quad (5.68)$$

Thus, we have shown that the objective function of P_1 is monotonically non-decreasing after each iteration. \blacksquare

5.7 Numerical Results for the MTC-enabled IRS-aided Network

In this section, we demonstrate the proposed algorithm's effectiveness for maximizing EE and IoT user admission in IRS-enabled systems with short packet lengths. We consider a simulation setup within a (100, 100) meter rectangular area. Here, the AP is positioned at the coordinates (0, 0) meters, and the IRS is at (50, 0) meters, with the assumption that all users are distributed randomly throughout this defined area. The model for path loss employed in this study is 3gpp-compliant [55] and is given by: $35.3 + 37.6 \log_{10}(d_k)$ [dB], with d_k representing the meter-measured distance between the AP and IoT user k .

For the purpose of achieving a fine-grained optimization, the AO method's convergence tolerance is set to 10^{-2} , and the thermal noise spectral density is -174 [dBm/Hz]. Moreover, the decoding error probability, a critical parameter in the context of short packet communications, is standardized across users at $\epsilon_k = 10^{-7}$. Additional simulation parameters are uniformly applied across all scenarios, including a total of $K = 20$ IoT users, $M = 5$ antennas at the AP, a packet length (m_d) of 250 symbols, and a threshold data rate (R_{th}^k) set at 1.6 [bits/Sec/Hz].

Table 5.1: Simulation Parameters for Multi-user MTC-enabled IRS Systems.

Parameter	Value
Area dimensions	(100, 100) meters
AP location	(0, 0) meters
IRS location	(50, 0) meters
Path loss model	$35.3 + 37.6 \log_{10}(d_k)$ dB
AO convergence tolerance	10^{-2}
Thermal noise density	-174 dBm/Hz
Decoding error probability, ϵ_k	10^{-7}
Number of users, K	20
Number of AP antennas, M	5
Number of IRS elements, N	50
Minimum transmit power, p_{\max}	30 dBm
Block code length, m_d	250 symbols
Threshold rate, R_{th}^k	1.6 bits/Second (Sec)/Hz

These simulation conditions are carefully chosen to reflect realistic operational environments and constraints, providing a robust framework within which the performance and implications of the proposed algorithm can be comprehensively assessed [165, 178]. The detailed setup, as shown in *Table 5.1*, encompassing both the physical deployment of network components and the algorithmic parameters, enables a thorough investigation into the algorithm's potential to significantly improve system performance in terms of EE and the capacity to support a higher number of IoT users within IRS-enabled wireless communication networks.

Fig. 5.3 illustrates the average EE as a function of varying maximum transmit power levels, p_{\max} , with $\alpha = 1$. To contextualize the performance of the proposed algorithm, we compare it against three baseline schemes: The first baseline focuses on optimizing the network's data rate as per [165]. The second assumes arbitrary passive beamforming at the IRS, and the third baseline envisions a scenario without IRS assistance (cf., Chapter 3). Across all schemes, a common trend emerges where the EE initially rises with increasing p_{\max} but eventually plateaus. Specifically, in the first baseline scenario, a subsequent decline in EE is noted, underscoring an insight that enhancing the data rate beyond a certain point, where the system achieves peak EE, leads to increased overall network power consumption, thereby diminishing EE.

This analysis further demonstrates the significant impact of phase shift optimization, illustrating that EE escalates with the increase of the reflecting elements at the IRS. Additionally, Fig. 5.3 explores the correlation between the average number of admitted IoT users and p_{\max} , when setting $\alpha = 0$. An upward trajectory in user admission rates is observed as p_{\max} escalates, attributable to the network's enhanced capability to support an expanded user base while adhering to the strict quality requirements necessitated for IoT users with finite blocklength in MTC-enabled networks.

Crucially, the figure shows the superior performance of the proposed scheme over the baseline alternatives, attributed to the strategic deployment of IRS and the concurrent optimization of active and passive beamforming matrices at the AP and IRS. This comprehensive comparison not only highlights the advantages of integrating IRS into the network but also

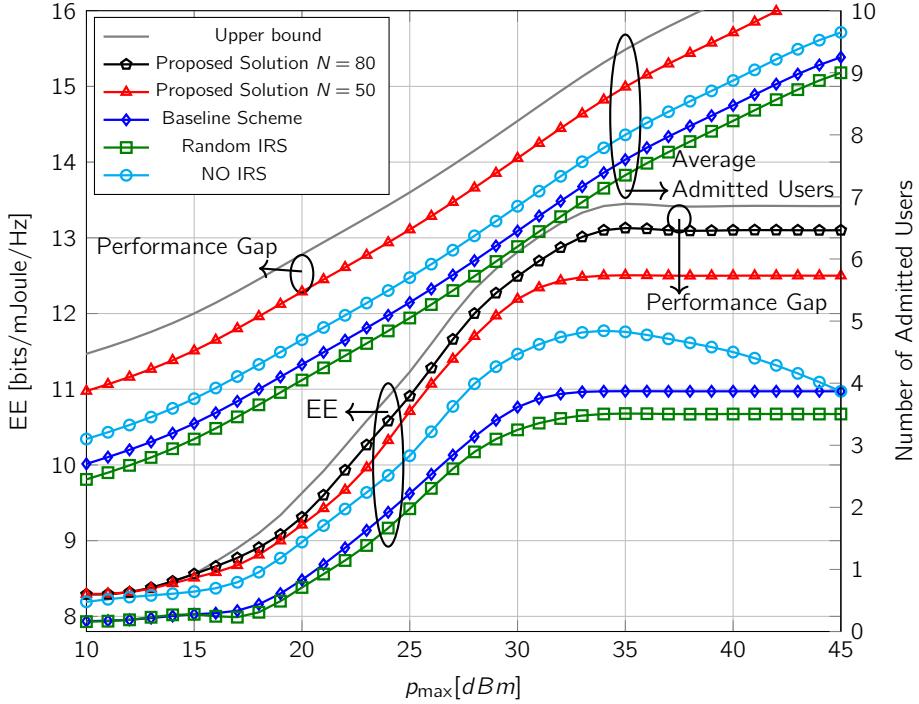


Figure 5.3: *EE and the average number of admitted users vs. p_{\max}*

illustrates the proposed scheme's efficacy in optimizing network parameters to achieve superior performance metrics.

In a further evaluative step of Fig. 5.3, the proposed scheme's performance is benchmarked against an idealized (and unattainable) performance upper bound represented by Shannon's capacity formula, achieved by setting the channel dispersion V_k in (5.8) to zero. This comparison serves as a theoretical touchstone, emphasizing the practical efficiency and effectiveness of the proposed algorithm in addressing the complex landscape of IRS-assisted wireless networks for IoT MTC-enabled applications. Through these comparative analyses, the proposed algorithm's role in pushing the boundaries of what is achievable in terms of EE and IoT user admission in contemporary wireless systems is vividly observed.

5.7.1 Tradeoff between EE and the Average Number of Admitted Users

Fig. 5.4 plots the tradeoff region between EE and the number of IoT admitted users, across a spectrum of values for the weighting coefficient $0 < \alpha < 1$, incremented in steps of 0.05. This illustration brings to light a significant tradeoff between EE and user admission: as EE is optimized, the number of users that can be admitted into the network conversely diminishes. This phenomenon underscores a fundamental principle where the pursuit of maximal EE inherently restricts the network's ability to accommodate an increasing number

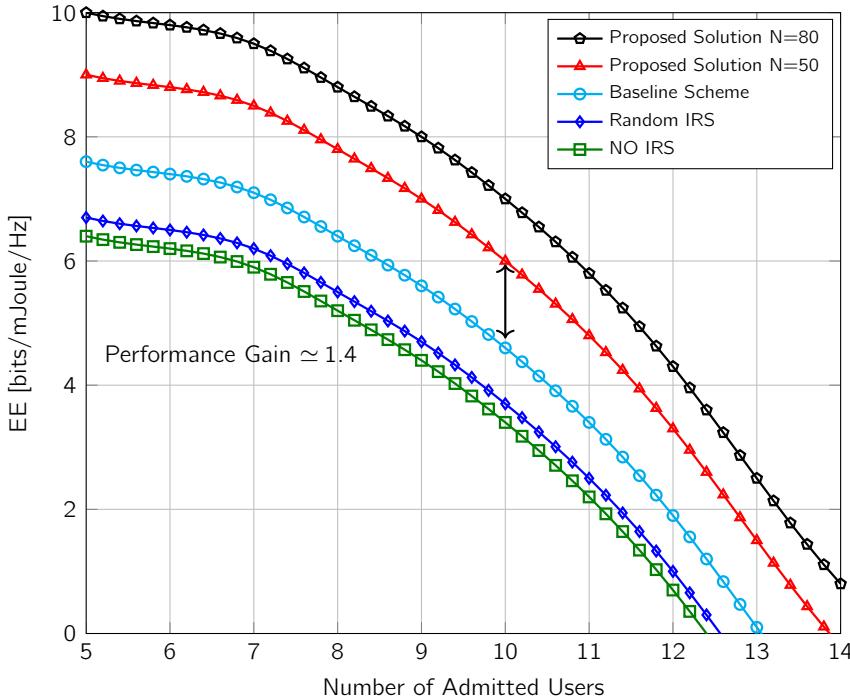


Figure 5.4: *EE vs. the average number of admitted users for IoT users with short packet lengths in an IRS-assisted MTC network.*

of users, establishing EE as a monotonically decreasing function in relation to user admission rates. Another insight from this analysis is the distinct performance dynamics influenced by varying α values. Specifically, when α is set towards higher values, the optimization leans favorably towards EE, albeit at the cost of admitting a relatively limited user base. This scenario implies that while the network may deliver high data rate services, it does so for a narrower number of users. On the contrary, as α values are reduced, there is a noticeable shift in optimization focus towards enhancing the number of admitted users. This adjustment naturally entails a more inclusive network that can support a larger user base, adhering to the minimum QoS standards required by users.

The modification of α effectively directs the optimization problem's emphasis between maximizing EE and expanding user admission. Consequently, in scenarios where α is lowered to prioritize user inclusivity, although the fairness within the network improves, it is observed that EE performance experiences a decline. This exploration into the tradeoff region highlights the inherent challenges in balancing EE with user admission objectives and proves the strategic role that the parameter α plays in navigating this balance.

5.7.2 User Admissibility vs. User Serviceability

In this chapter, we aim exclusively towards enhancing the framework of an IRS-assisted MTC-enabled system, specifically focusing on maximizing the number of users the network can viably support. Central to our proposition is the strategic allocation of resources to those users whom the system is capable of admitting, striving for fairness in distribution among this subset. It is critical to understand that our algorithm *does not* purport to accommodate every potential user⁵. Instead, our intention is rooted in determining the extent of user support achievable, particularly through the lens of maximizing EE.

The above discussion presents two seemingly contradictory inquiries that cannot be concurrently addressed:

- The capacity of our designed network to fairly admit a certain number of users; known as user admissibility.
- The network's ability to extend services to all potential users; referred to as user serviceability.

In this chapter, we have answered the user admissibility question. It is imperative to recognize that the attempt to maximize the total number of users the system can support does not inherently guarantee the serviceability of all users. In scenarios where maximizing user admissibility coincides with the capability to serve all users, such alignment represents a special case rather than a standard expectation. This distinction underscores that, should the network manage to provide service to every user, it effectively admits them all, establishing a benchmark of user admissibility⁶. Further investigation of our system's design reveals an embedded fairness in resource allocation, ensuring that admitted users receive equitable consideration.

In what follows, we explain in more detail why our design is fair (i.e., we have inherent user fairness) and explore the ways we could tackle user serviceability.

User Admissibility

Within the domain of MTC-enabled services, the necessity for short packet lengths inherently aligns with the requirement for modest data rates among URLLC users. This foundational aspect of MTC-enabled applications leads us to a preliminary conclusion regarding the equitable treatment of 'admitted' MTC-enabled IoT users within the framework of our study. Given the unique constraints and needs of these users, fairness in resource allocation emerges not just as an objective but as an intrinsic characteristic of our system design.

To critically assess and quantify this inherent fairness, our analysis employs Jain's fairness index, a methodologically robust and universally recognized metric for evaluating fairness

⁵User fairness *across time* over all users is out of the scope of this chapter. We do not claim that we support all users; rather, we want to answer the question how many users can be supported (in the milieu of maximizing energy efficiency)

⁶If we can give service to all users, then we are admitting them all.

in resource allocation scenarios [198]. Jain's fairness index is particularly relevant to our discussion for its ability to deliver a quantifiable measure of how equitably resources are distributed among users, including those within MTC contexts. The mathematical expression for Jain's fairness index in our scenario is formulated as:

$$I_J = \frac{\left(\sum_{k \in \mathcal{K}} u_k R_k(u_k, \mathbf{w}_k, \Theta) \right)^2}{\sum_{k \in \mathcal{K}} u_k \sum_{k \in \mathcal{K}} (R_k(u_k, \mathbf{w}_k, \Theta))^2}. \quad (5.69)$$

This index operates within a range from 0 to 1, where a value of 1 means perfect fairness — every user benefits from identical throughput levels — while a score of 0 denotes absolute unfairness. Intriguingly, a fairness index value of 0.5 is indicative of a perfectly equitable resource allocation among the users.

The crux and primary innovation of our study pertains to the maximization of the overall number of admitted users while concurrently guaranteeing a minimum data rate for each accepted user. In particular, the proposed problem formulation allows a network designer to access all pertinent parameters, of which the most salient are the aggregate / achievable count of admitted users and the energy efficacy provided through our methodology

The essence and novelty of our investigation lie in the strategic objective, as in P_1 , to enhance the total number of users that can be admitted into the system, ensuring simultaneously that each admitted user is guaranteed a minimum requisite data rate. This multi-objective role is well thought of in our problem formulation P_1 , which equips network designers with the capability to evaluate critical system parameters effectively. These include the collective number of users that can feasibly be supported by the system, and the level of EE achieved through the implementation of our proposed methodologies. Through this approach, our study not only addresses the question of user admissibility but also sheds light on the broader implications of resource allocation fairness, particularly as it applies to MTC-enabled IoT users within an IRS-assisted communication framework.

The implication of our findings is that by employing Jain's fairness index, fairness can be ensured while achieving the highest possible number of admissible users. Specifically, our results demonstrate that it is feasible to optimize resource allocation in a communication network such that each user receives an equitable share of resources while concurrently maximizing the total number of admissible users. In essence, our study establishes the possibility of achieving both high system performance and fairness.

By examining the Pareto efficiency graph depicted in Fig 5.5, it is possible to infer that nine users can be fully admitted into the network, with an EE of approximately 8 [bits/million Joule (mJoule)/Hz]. Notably, the (8,9) point lies on the Pareto optimal boundary of the graph, signifying that any point inside the Pareto region represents a feasible solution to the optimization problem, albeit non-optimal.

Figure 5.5 displays the dynamics between the fairness indices achieved through our proposed scheme and the number of users, particularly when the fairness constraint, as defined by (5.69), is integrated into the primary optimization framework P_1 . The figure illustrates a notable trend: with the increase of users, the arena of competition becomes increasingly fierce. This phenomenon is attributed to the inevitable rise in users facing less favorable channel conditions and those designated with lower priority levels, who subsequently face

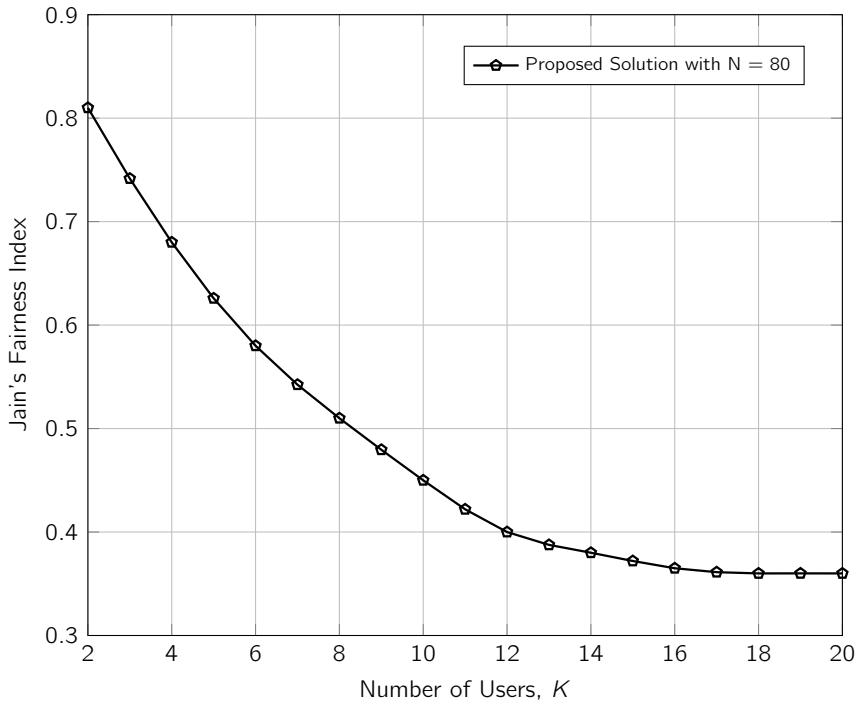


Figure 5.5: Jain's fairness index versus the total number of users in for IoT users with short packet lengths in an IRS-assisted MTC network.

challenges in securing access to communication services. The ambition is to enhance the total number of users the system can feasibly support. It is imperative to acknowledge, however, that while our focus is on maximizing user admission, the aspiration to optimize fairness across the board or to guarantee service provision to every prospective user lies beyond our design objectives. Despite this, our analysis reveals a compelling insight: the algorithm we propose does not overlook fairness entirely but rather ensures a degree of equity among the set of users it admits.

This balance illustrates a deliberate design consideration within our algorithm, aiming to strike a compromise between broadening user admission and maintaining fairness among those admitted. Such a compromise is reflective of the complex trade-offs inherent in optimizing communication systems, especially those with the challenges of accommodating a growing user base while striving to uphold a pinch of fairness.

User Serviceability

To address the challenge of ensuring service to all users within a network, a shift in perspective is required, moving away from merely maximizing the total user count. This transition is crucial because in a scenario where all users are guaranteed admission, the goal of maximizing user numbers loses its relevance. The issue of user serviceability, highlighted earlier, can be approached from multiple angles, drawing upon existing strategies within the literature

to ensure service distribution over time.

One potential strategy involves redefining the optimization problem based on a Max-Min fairness methodology. This approach focuses on equalizing the minimum service time necessary for each user to receive adequate service, a concept that has been thoroughly investigated in prior studies, such as in [199]. Alternatively, adopting a Medium Access Control (MAC) layer perspective offers another pathway to achieving temporal user fairness. By assigning time slots (or minislots) to MTC-enabled IoT users, it becomes feasible to ensure that various user types are authorized their respective services in alignment with MTC-enabled IoT constraints, as explored in [200].

The Max-Min fairness approach provides another distinct viewpoint on network management, diverging from the perspective centered on the optimization of total user admission. Moreover, our current investigation, which seeks to optimize user admission alongside EE at the Physical (PHY) layer, lays a foundational framework that can be effectively integrated into MAC layer protocols. This integration is crucial for ensuring cross-temporal fairness among all user types, as demonstrated in Fig. 2 of [200]. Such synergy between the PHY layer algorithm and MAC layer scheduling underscores the applicability and relevance of our study within broader communication system architectures, emphasizing the contribution of our findings to enhancing network efficiency and fairness.

In summary, while our study deliberately sidelines the consideration of user serviceability to maintain a focus on maximizing user admissibility, it presents a comprehensive framework for understanding and addressing the complexities of equitable resource allocation. The insights gleaned from our analysis not only contribute to optimizing network performance but also underscore the potential for implementing our findings within integrated PHY and MAC layer strategies, thereby reinforcing the importance of our contributions to the field.

5.8 Insights and Practical Implications of MTC-enabled services for IRS-aided Networks

Thus far, our exploration within this chapter has centered around MTC-enabled, IRS-assisted networks, focusing particularly on a DL multi-user MISO configuration with short packet transmission. Our journey took us through the complexities of designing resource allocation strategies that optimize active and passive beamforming, with the goal of enhancing EE while also increasing the admittance of IoT users, all within the scope of meeting stringent QoS criteria for each MTC-enabled user. Given the non-convex nature of the core problem, our approach was hierarchical and structured; we initially segmented the main problem into two distinct sub-problems, active and passive beamforming, before delving into the SCA and penalty-based methodologies to untangle the intricacies of beamforming matrices. Our simulations illustrated the key role of the IRS in improving system EE and facilitating QoS compliance for users constrained by short packet lengths, marking an improvement over traditional approaches.

In the ensuing sections of this chapter, we introduce and study Ultra-Reliable Low-Latency Communication (URLLC) systems, a facet of modern wireless communication. URLLC

can be thought of as an essential feature for various MTC services, including a myriad of emergent and mission-critical applications [201]. In URLLC systems, reliability, and minimal latency are of significant importance, which is not necessarily the same as conventional requirements in MTC scenarios. This shift is necessitated by the evolving demands of next-generation wireless systems. We develop a system model and analysis of how URLLC services can benefit from IRS. Before that, let's talk about URLLC in a bit more detail.

5.8.1 Ultra-Reliable Low-Latency Communication (URLLC) in IRS-aided Networks

The integration of IRS into the communication system helps to enhance reliability, reduce packet retransmission, and minimize delay. Thus, the IRS can be a potential and cost-effective solution to realize URLLC. To facilitate low-latency communications, the packet size in URLLC must be extremely small [161]. This mode of operation, referred to as short packet transmission⁷, does not align well with conventional Shannon's capacity theorem, which assumes that coding is performed on an infinite blocklength [8]. However, URLLC functions within a finite blocklength regime, which requires a different approach. Polyanskiy *et al.* in [163] determined the achievable regime as a complex function of the Signal-to-Noise (SNR) ratio, blocklength, and the probability of decoding errors.

As data traffic demand increases, so does the energy consumption of wireless networks. Therefore, EE becomes a pivotal aspect of future network design [3]. In [178], Singh *et al.* sought to maximize EE in downlink multi-user Multi-Input Single-Output (MISO) networks by jointly optimizing precoders at the Base Station (BS) and the Decoding Error Probability (DEP) with finite blocklength codes. Further, Nasir *et al.* developed a unique class of conjugate beamforming and a new path-following algorithm to enhance URLLC rates and EE in cell-free massive Multi-Input Multi-Output (cfm-MIMO) settings in [202]. However, [178] and [202] only consider the downlink system, while a more comprehensive full-duplex system is not accounted for. Ghanem *et al.* closed this gap in [203, 165] and aimed to maximize system EE through a joint effort of power control and sub-carrier assignment in downlink and uplink communication, ensuring the Quality of Service (QoS) requirements for users in both directions are met.

Nonetheless, not all these strategies, as referenced in [3, 178, 202, 203, 165], are exclusively applicable or can be extended to IRS-assisted URLLC services. As such, it is necessary to examine IRS-assisted URLLC networks in their own context. Hashemi *et al.* authors presented the performance analysis of the average achievable data-rate and error probability over an IRS-aided URLLC transmission with/without phase noise in [168]. Considering non-linear energy harvesting, the end-to-end performance of the IRS-assisted wireless system was analyzed in [204] for industrial URLLC applications, and the approximate closed-form expression of the block error rate was derived. Xie *et al.* studied an IRS-assisted downlink multi-user URLLC system and jointly optimized the user grouping and the blocklength allocation at the BS, as well as the reflective beamforming at the IRS for latency minimization in [90].

The majority of current research in the field falls still short of thoroughly examining the

⁷We extensively use this in the previous sections of our studies on MTC-enabled services.

advantages of utilizing the IRS platform in URLLC networks with the aim of optimizing EE [205, 8, 111, 167, 3, 206]. The IRS can assist URLLC systems in lowering the anticipated worst-case Signal-to-Interference-plus-Noise Ratio (SINR) in applications requiring low latency. Furthermore, it would be intriguing to assess whether the Quality of Service (QoS) is upheld in the context of short packet transmission in an IRS-supported URLLC system. In this chapter, we aim to design a resource allocation algorithm for a downlink URLLC-enabled IRS system, wherein a multi-antenna BS serves multiple single-antenna URLLC receivers via a smart, reconfigurable reflector. As a result, our primary focus in the following sections is on enhancing the total EE in the proposed network, which offers valuable insights into the system's design. To extend our study of MTC to URLLC, we do the following:

- We formulate a non-convex EE problem, subject to the minimum required data-rate for each URLLC user and unit-modulus constraints at the IRS.
- We employ an AO algorithm to solve the EE maximization problem, optimizing both active and passive beamformers at the BS and IRS, respectively. Using an innovative iterative rank relaxation and SCA method, we approach the IRS rank-one constrained problem. We then apply the Dinkelbach algorithm and a penalty-based approach to obtain a suboptimal solution.
- We use simulation to indicate that the integration of IRSs and a multi-antenna BS enhances the performance of URLLC users with short packet transmissions.

Building upon the foundation of MTC-enabled IRS-aided networks in our earlier discussions, we now introduce a system model incorporating URLLC services in an IRS-assisted network.

5.9 URLLC-enabled System Model and Problem Formulation

We consider a downlink MISO multi-user URLLC-enabled IRS-assisted system, represented in Fig. 5.6, which comprises an IRS with N elements, a BS equipped with M antennas, and K users, each with a single antenna. The group of IRS elements, BS antennas, and users are respectively denoted by the sets $\mathcal{N} = \{1, \dots, N\}$, $\mathcal{M} = \{1, \dots, M\}$, and $\mathcal{K} = \{1, \dots, K\}$. Additionally, we suppose that B_k bits of information are assigned to user k . In this setup, the BS transforms these information bits into a block code in time slot l , characterized by a length of m_d symbols. We decode the k -th user block code as $x_{k,l}$, with l being part of the set $\mathcal{L} = \{1, 2, \dots, m_d\}$. Following these assumptions, the transmission signal emanating from the BS can be expressed as

$$\mathbf{s}[l] = \sum_{k \in \mathcal{K}} \boldsymbol{\omega}_{k,l} x_{k,l}, \forall l \in \mathcal{L}, \quad (5.70)$$

where $\boldsymbol{\omega}_{k,l} \in \mathbb{C}^{M \times 1}$ represents the beamforming vector for user k . The channel links exhibit time invariance (slow fading). Additionally, we assume that the BS possesses full access to the Channel State Information (CSI) and has comprehensive knowledge of all URLLC users'

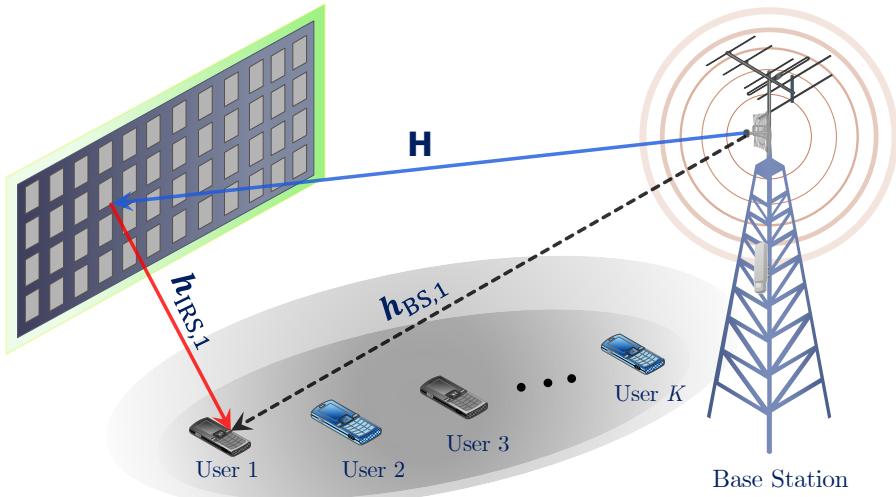


Figure 5.6: Illustration of a multi-user IRS-assisted MISO downlink network comprising one BS and K URLLC users employing finite block length transmissions. \mathbf{h} is the channel matrix between the BS and the IRS, $\mathbf{h}_{\text{IRS},k}$ denotes the channel response vector from the IRS to user k , and $\mathbf{h}_{\text{BS},k}$ indicates the channel between the BS and user k .

delay requirements. As a result, our proposed algorithm is positioned as the theoretical upper limit of performance when contrasted with strategies devised under the constraints of partial or no CSI availability [83, 165, 174, 9]. The baseband equivalent channel responses for BS-to-IRS, IRS-to-user k , and BS-to-user k are denoted as $\mathbf{h} \in \mathbb{C}^{N \times M}$, $\mathbf{h}_{\text{IRS},k} \in \mathbb{C}^{N \times 1}$, and $\mathbf{h}_{\text{BS},k} \in \mathbb{C}^{M \times 1}$, respectively. Also, we define the matrix of reflection coefficients at the IRS as $\Psi_I = \text{diag}(\alpha_{1,I} e^{j\phi_{1,I}}, \alpha_{2,I} e^{j\phi_{2,I}}, \dots, \alpha_{N,I} e^{j\phi_{N,I}})$, where $\alpha_{n,I} \in [0, 1]$ and $\phi_{n,I} \in (0, 2\pi]$, $\forall n \in \mathcal{N}, \forall I \in \mathcal{L}$ are the reflection amplitude and phase shift of the n -th reflection coefficient at the IRS during time slot I , respectively. Now, by defining the combined channel link seen by the k -th URLLC user as:

$$\mathbf{h}_k^H \triangleq \mathbf{h}_{\text{IRS},k}^H \Psi_I \mathbf{h} + \mathbf{h}_{\text{BS},k}^H, \quad \forall k \in \mathcal{K}, \forall I \in \mathcal{L}, \quad (5.71)$$

the SINR of that user can be expressed as:

$$\Gamma_{k,I} = \frac{|\mathbf{h}_k^H \mathbf{\omega}_{k,I}|^2}{\sum_{i \neq k, i \in \mathcal{K}} |\mathbf{h}_k^H \mathbf{\omega}_{i,I}|^2 + \sigma_k^2}, \quad \forall k \in \mathcal{K}, \forall I \in \mathcal{L}, \quad (5.72)$$

where σ_k^2 represents the noise variance received at user k . In the context of URLLC systems, ensuring low-latency and high-reliability wireless communication requires finite and short blocklength transmissions. Thus, an accurate estimate of the achievable data-rate for each user can be delineated as follows [163]:

$$\mathcal{R}_k(\mathbf{\omega}_{k,I}, \Psi_I) = \mathcal{U}_k(\mathbf{\omega}_{k,I}, \Psi_I) - \mathcal{V}_k(\mathbf{\omega}_{k,I}, \Psi_I), \quad \forall k \in \mathcal{K}, \quad (5.73)$$

where

$$\mathcal{U}_k(\mathbf{\omega}_{k,I}, \Psi_I) = \sum_{I \in \mathcal{L}} \log_2(1 + \Gamma_{k,I}), \quad \forall k \in \mathcal{K}, \quad (5.74)$$

$$\mathcal{V}_k(\boldsymbol{\omega}_{k,I}, \boldsymbol{\Psi}_I) = \sum_{I \in \mathcal{L}} \beta_k \sqrt{\Delta_{k,I}}, \quad \forall k \in \mathcal{K}, \quad (5.75)$$

$$\beta_k = \frac{Q^{-1}(\epsilon_k)}{\sqrt{m_d}}, \quad \forall k \in \mathcal{K}. \quad (5.76)$$

Here, ϵ_k refers to the decoding error probability, m_d , as defined earlier, is representative of the blocklength, and $\Delta_{k,I}$ the channel dispersion, computed as:

$$\Delta_{k,I} = (\log_2 e)^2 \left(1 - \frac{1}{(1 + \Gamma_{k,I})^2} \right), \quad \forall k \in \mathcal{K}, \forall I \in \mathcal{L}. \quad (5.77)$$

To ensure users' Quality of Service (QoS) regarding the received number of bits, the reliability, and the latency, a minimum data-rate denoted by $R_{\min,k}$ which should be satisfied for each user as follows:

$$\mathcal{R}_k(\boldsymbol{\omega}_{k,I}, \boldsymbol{\Psi}_I) \geq R_{\min,k}, \quad \forall k \in \mathcal{K}. \quad (5.78)$$

Next, we describe the EE as the ratio of the total system data-rate over the associated network power consumption in [bits/Joule/Hz]:

$$\eta_{\text{eff}}(\boldsymbol{\omega}_{k,I}, \boldsymbol{\Psi}_I) = \frac{\sum_{k \in \mathcal{K}} \mathcal{R}_k(\boldsymbol{\omega}_{k,I}, \boldsymbol{\Psi}_I)}{\sum_{k \in \mathcal{K}} \sum_{I \in \mathcal{L}} \|\boldsymbol{\omega}_{k,I}\|^2 + P_s + N P_d + P_c^{\text{BS}}}, \quad (5.79)$$

where P_s represents the static power consumption necessary for sustaining the basic circuit operations of the IRS, while P_d is the dynamic power expended per reflecting component, and P_c denotes the circuit power at the BS.

Following this, we begin by formulating a resource allocation problem to optimize energy efficiency while accounting for feasibility constraints in a downlink MISO IRS-aided URLLC system with short packet transmissions. Subsequently, we propose an iterative algorithm to work out this optimization problem. Our objective is to maximize the total EE of the system under consideration. We aim to achieve this by concurrently optimizing the active beamformers at the BS and the phase shifts at the IRS. Consequently, the task of maximizing the overall energy efficiency of the system can be mathematically articulated as follows:

$$\mathcal{P}_9 : \max_{\boldsymbol{\omega}_{k,I}, \boldsymbol{\Psi}_I} \eta_{\text{eff}}(\boldsymbol{\omega}_{k,I}, \boldsymbol{\Psi}_I) \quad (5.80a)$$

$$\text{s.t.} : \mathcal{R}_k(\boldsymbol{\omega}_{k,I}, \boldsymbol{\Psi}_I) \geq R_{\min,k}, \quad \forall k \in \mathcal{K}, \quad (5.80a)$$

$$|\boldsymbol{\Psi}_{nn,I}| = 1, \quad \forall n \in \mathcal{N}, \forall I \in \mathcal{L}, \quad (5.80b)$$

$$\sum_{I \in \mathcal{L}} \sum_{k \in \mathcal{K}} \|\boldsymbol{\omega}_{k,I}\|^2 \leq P_{\max}, \quad (5.80c)$$

$$\boldsymbol{\omega}_{k,I} = 0, \quad \forall k \in \mathcal{K}, \forall I \geq T_k, \quad (5.80d)$$

where (5.80a) is the minimum data-rate requirement ($R_{\min,k}$) of each URLLC user k . The constraint (5.80b) seeks that the diagonal phase shift matrix contains N unit-modulus elements along its main diagonal. The constraint (5.80c) defines the limitation of the BS

transmission power budget, with p_{\max} signifying the maximum permissible BS transmission power. The constraint (5.80d) is imposed to safeguard the real-time URLLC service functionality, ensuring that user k receives service within the first T_k time slots to satisfy its delay requirements.

Given the non-convex nature of the objective function and constraints, the optimization problem P_9 is non-convex and NP-hard. As a rule, finding an optimal solution for such problems is typically non-trivial and out of the question. However, in the following section, we employ a strategy to find an efficient polynomial time suboptimal solution.

5.10 Proposed Solution: A Novel Rank Relaxation Method

Problem P_9 poses a significant challenge due to its non-convex nature and the strong coupling of optimization variables. As such, there is a lack of a standard, well-structured method to solve P_9 . In response to this, we propose an Alternating Optimization (AO) approach based on Successive Convex Approximation (SCA) with reasonable computational complexity aimed at securing a suboptimal solution. This AO approach involves two distinct non-convex sub-problems. The first sub-problem focuses on determining the optimal active beamformers at the BS, while the second sub-problem concentrates on optimizing the phase shifts at the IRS. For the latter, we incorporate a modified objective function to eliminate feasibility concerns. In more specific terms, we first optimize the beamforming vectors at the BS and then design the phase shifts at the IRS alternatively, taking into account the previously found beamforming vectors at the BS. Moving forward, we adopt the SCA technique to initially derive a lower bound for $\mathcal{R}_k(\omega_{k,l}, \Psi_l)$, which will subsequently be utilized when addressing each subproblem.

Lemma 2 *Let (j) denote the subscript associated with the feasible solution procured in the j -th iteration of the SCA algorithm. A lower bound approximation of $\mathcal{R}_k(\omega_{k,l}, \Psi_l)$ is given by:*

$$\mathcal{R}_k(\omega_{k,l}, \Psi_l) \geq \tilde{\mathcal{R}}_k(\omega_{k,l}, \Psi_l) = \sum_{l \in \mathcal{L}} \tilde{\mathcal{R}}_{k,l}^{(j)}(\omega_{k,l}, \Psi_l), \quad (5.81)$$

$$\begin{aligned} \tilde{\mathcal{R}}_{k,l}^{(j)}(\omega_{k,l}, \Psi_l) &\triangleq -\gamma_{k,l} b_{k,l} + 2\rho_{k,l} \Re \left\{ \sum_{i \neq k, i \in \mathcal{K}} a_{i,k,l}^{(j)} a_{i,k,l}^H + \sigma_k^2 \right\} \\ &\quad + 2\Re \left\{ \frac{a_{k,k,l}^{(j)} a_{k,k,l}^H}{b_{k,l}^{(j)} \left(b_{k,l}^{(j)} - |a_{k,k,l}^{(j)}|^2 \right)} \right\} + \xi_{k,l}, \\ &\quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \end{aligned} \quad (5.82)$$

where

$$\mathbf{a}_{k,i,l} = \mathbf{h}_k^H \boldsymbol{\omega}_{i,l}, \forall i, k \in \mathcal{K}, \forall l \in \mathcal{L}, \quad (5.83)$$

$$\mathbf{b}_{k,l} = \sum_{i \in \mathcal{K}} |\mathbf{h}_k^H \boldsymbol{\omega}_{i,l}|^2 + \sigma_k^2, \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \quad (5.84)$$

$$\rho_{k,l} = \frac{\beta_k (\log_2 e)^2}{\mathbf{b}_{k,l} \sqrt{\left(\Gamma_{k,l}^{(j)}\right)^2 + 2\Gamma_{k,l}^{(j)}}}, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \quad (5.85)$$

$$\gamma_{k,l} = \frac{|\mathbf{a}_{k,k,l}^{(j)}|^2}{\mathbf{b}_{k,l}^{(j)} \left(\mathbf{b}_{k,l}^{(j)} - |\mathbf{a}_{k,k,l}^{(j)}|^2 \right)} + \frac{\rho_{k,l} \left(\mathbf{b}_{k,l}^{(j)} - |\mathbf{a}_{k,k,l}^{(j)}|^2 \right)}{\mathbf{b}_{k,l}^{(j)}}, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \quad (5.86)$$

$$\begin{aligned} \xi_{k,l} = & \log_2 \left(1 + \Gamma_{k,l}^{(j)} \right) - \frac{|\mathbf{a}_{k,k,l}^{(j)}|^2}{\mathbf{b}_{k,l}^{(j)} \left(\mathbf{b}_{k,l}^{(j)} - |\mathbf{a}_{k,k,l}^{(j)}|^2 \right)} \\ & - \frac{\beta_k \left(2\Delta_{k,l}^{(j)} + (\log_2 e)^2 \right)}{2\sqrt{\Delta_{k,l}^{(j)}}} - \frac{\rho_{k,l} \mathbf{b}_{k,l}^{(j)}}{2 \left(1 + \Gamma_{k,l}^{(j)} \right)}. \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}. \end{aligned} \quad (5.87)$$

In the $(j+1)$ -th iteration, the data-rate function $\tilde{\mathcal{R}}_{k,l}(\boldsymbol{\omega}_{k,l}, \boldsymbol{\Psi}_l)$ exhibits concavity with respect to each variable, reaching its boundary at the point $(\boldsymbol{\omega}_{k,l}^{(j)}, \boldsymbol{\Psi}_l^{(j)})$.

Proof 4 According to the data-rate equation in (5.73), we can drive a lower bound on $\mathcal{U}_k(\boldsymbol{\omega}_{k,l}, \boldsymbol{\Psi}_l)$ and an upper bound on $\mathcal{V}_k(\boldsymbol{\omega}_{k,l}, \boldsymbol{\Psi}_l)$. In order to approximate the non-convex logistic function, we utilize the MM algorithm [207]. This is achieved by constructing a surrogate function using a first-order Taylor approximation, which is expressed as follows:

$$\mathcal{U}_k(\mathbf{a}) \simeq \mathcal{U}_k(\mathbf{a}^{(j)}) + \nabla_{\mathbf{a}} \mathcal{U}_k(\mathbf{a}^{(j)}).(\mathbf{a} - \mathbf{a}^{(j)}) \triangleq \tilde{\mathcal{U}}_k(\mathbf{a}), \forall k \in \mathcal{K}, \quad (5.88)$$

$$\mathcal{V}_k(\mathbf{b}) \simeq \mathcal{V}_k(\mathbf{b}^{(j)}) + \nabla_{\mathbf{b}} \mathcal{V}_k(\mathbf{b}^{(j)}).(\mathbf{b} - \mathbf{b}^{(j)}) \triangleq \tilde{\mathcal{V}}_k(\mathbf{b}), \forall k \in \mathcal{K}, \quad (5.89)$$

where $\mathbf{a}, \mathbf{b} = \{\boldsymbol{\omega}_{k,l}, \boldsymbol{\Psi}_l\}$ and j is the iteration number. Moreover, $\mathbf{a}^{(j)}$ and $\mathbf{b}^{(j)}$ denote the solutions of the problem at (j) th iteration. Based on (5.83)-(5.87), we first rewrite the SINR formula as follows:

$$\Gamma_{k,l} = \frac{|\mathbf{a}_{k,k,l}|^2}{\mathbf{b}_{k,l} - |\mathbf{a}_{k,k,l}|^2}, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}. \quad (5.90)$$

Equivalently, we can rewrite the channel dispersion as:

$$\Delta_{k,l} = (\log_2 e)^2 \left(1 - \left(\frac{\mathbf{b}_{k,l} - |\mathbf{a}_{k,k,l}|^2}{\mathbf{b}_{k,l}} \right)^2 \right), \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}. \quad (5.91)$$

Subsequently, we have:

$$\begin{aligned}\mathcal{U}_k(\boldsymbol{\omega}_{k,l}, \boldsymbol{\Psi}_l) &= -\sum_{l \in \mathcal{L}} \log_2 \left(1 - \frac{|\mathbf{a}_{k,k,l}|^2}{\mathbf{b}_{k,l}} \right) \\ &= \sum_{l \in \mathcal{L}} \mathcal{U}_{k,l}(\boldsymbol{\omega}_{k,l}, \boldsymbol{\Psi}_l), \quad \forall k \in \mathcal{K},\end{aligned}\quad (5.92)$$

$$\begin{aligned}\mathcal{V}_k(\boldsymbol{\omega}_{k,l}, \boldsymbol{\Psi}_l) &= \sum_{l \in \mathcal{L}} \beta_k \sqrt{\Delta_{k,l}} \\ &= \sum_{l \in \mathcal{L}} \mathcal{V}_{k,l}(\boldsymbol{\omega}_{k,l}, \boldsymbol{\Psi}_l), \quad \forall k \in \mathcal{K}.\end{aligned}\quad (5.93)$$

Thus, employing the MM technique allows us to derive the following lower limit for $\mathcal{U}_{k,l}(\boldsymbol{\omega}_{k,l}, \boldsymbol{\Psi})$:

$$\begin{aligned}\mathcal{U}_{k,l}(\boldsymbol{\omega}_{k,l}, \boldsymbol{\Psi}) &\geq -\frac{|\mathbf{a}_{k,k,l}^{(j)}|^2}{\mathbf{b}_{k,l}^{(j)} \left(\mathbf{b}_{k,l}^{(j)} - |\mathbf{a}_{k,k,l}^{(j)}|^2 \right)} \mathbf{b}_{k,l} \\ &\quad + 2\Re \left\{ \frac{\mathbf{a}_{k,k,l}^{(j)} \mathbf{a}_{k,k,l}^H}{\mathbf{b}_{k,l}^{(j)} - |\mathbf{a}_{k,k,l}^{(j)}|^2} \right\} + \xi_{k,l}^{\mathcal{U}}, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L},\end{aligned}\quad (5.94)$$

where

$$\xi_{k,l}^{\mathcal{U}} = \log_2 \left(1 + \Gamma_{k,l}^{(j)} \right) - \frac{|\mathbf{a}_{k,k,l}^{(j)}|^2}{\mathbf{b}_{k,l}^{(j)} \left(\mathbf{b}_{k,l}^{(j)} - |\mathbf{a}_{k,k,l}^{(j)}|^2 \right)}, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}. \quad (5.95)$$

Similarly, we use the first-order Taylor approximation to derive the following upper limit for $\mathcal{V}_{k,l}(\boldsymbol{\omega}_{k,l}, \boldsymbol{\Psi})$:

$$\begin{aligned}\mathcal{V}_{k,l}(\boldsymbol{\omega}_{k,l}, \boldsymbol{\Psi}) &\leq \beta_k \sqrt{\Delta_{k,l}^{(j)}} + \frac{(\log_2 e)^2 \beta_k}{2\sqrt{\Delta_{k,l}^{(j)}}} \left(1 - \frac{1}{(1 + \Gamma_{k,l})^2} \right), \\ &\quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}.\end{aligned}\quad (5.96)$$

Now, by defining $\Xi_{k,l} \triangleq 1/(1 + \Gamma_{k,l})$ and according to [3], we can rewrite:

$$\frac{1}{(1 + \Gamma_{k,l})^2} = \Xi_{k,l}^2 \quad (5.97)$$

$$\geq 2\Xi_{k,l}^{(j)}\Xi_{k,l} - (\Xi_{k,l}^{(j)})^2 \quad (5.98)$$

$$= \frac{2}{(1 + \Gamma_{k,l}^{(j)})(1 + \Gamma_{k,l})} - \frac{1}{(1 + \Gamma_{k,l}^{(j)})^2}, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \quad (5.99)$$

where the inequality holds due to convexity of $(\Xi_{k,l})^2$. The term $1/(1 + \Gamma_{k,l})$ in (5.99) is

non-convex due to its fractional form. We rewrite this term as follows and then apply MM:

$$\begin{aligned} \frac{1}{1 + \Gamma_{k,I}} &= \frac{\mathbf{b}_{k,I} - |\mathbf{a}_{k,k,I}|^2}{\mathbf{b}_{k,I}} \geq - \frac{\left(\mathbf{b}_{k,I}^{(j)} - |\mathbf{a}_{k,k,I}^{(j)}|^2 \right) \mathbf{b}_{k,I}}{|\mathbf{b}_{k,I}^{(j)}|^2} \\ &+ 2\Re \left\{ \frac{\sum_{i \neq k, i \in \mathcal{K}} \mathbf{a}_{i,k,I}^{(j)} \mathbf{a}_{i,k,I}^H + \sigma_k^2}{\mathbf{b}_{k,I}^{(j)}} \right\}, \forall k \in \mathcal{K}, \forall I \in \mathcal{L}, \end{aligned} \quad (5.100)$$

where the inequality holds due to the convexity with respect to the $\mathbf{a}_{i,k,I}$ and $\mathbf{b}_{k,I}$. Thus, we arrive at an upper bound for as follows: $\mathcal{V}_{k,I}(\boldsymbol{\omega}_{k,I}, \boldsymbol{\Psi})$:

$$\begin{aligned} \mathcal{V}_{k,I}(\boldsymbol{\omega}_{k,I}, \boldsymbol{\Psi}) &\leq \frac{(\log_2 e)^2 \beta_k}{\sqrt{(\Gamma_{k,I}^{(j)})^2 + 2\Gamma_{k,I}^{(j)}}} \left(\frac{\left(\mathbf{b}_{k,I}^{(j)} - |\mathbf{a}_{k,k,I}^{(j)}|^2 \right) \mathbf{b}_{k,I}}{|\mathbf{b}_{k,I}^{(j)}|^2} \right. \\ &\quad \left. - 2\Re \left\{ \frac{\sum_{i \neq k, i \in \mathcal{K}} \mathbf{a}_{i,k,I}^{(j)} \mathbf{a}_{i,k,I}^H + \sigma_k^2}{\mathbf{b}_{k,I}^{(j)}} \right\} \right) + \xi_{k,I}^{\mathcal{V}}, \forall k \in \mathcal{K}, \forall I \in \mathcal{L}, \end{aligned} \quad (5.101)$$

where

$$\begin{aligned} \xi_{k,I}^{\mathcal{V}} &= \frac{\left(2\Delta_{k,I}^{(j)} + (\log_2 e)^2 \right) \beta_k}{2\sqrt{\Delta_{k,I}^{(j)}}} \\ &+ \frac{(\log_2 e)^2 \beta_k}{2 \left(1 + \Gamma_{k,I}^{(j)} \right) \sqrt{\left(\Gamma_{k,I}^{(j)} \right)^2 + 2\Gamma_{k,I}^{(j)}}}, \forall k \in \mathcal{K}, \forall I \in \mathcal{L}. \end{aligned} \quad (5.102)$$

Ultimately, the lower bound of the data-rate function in (5.73) can be approximated as:

$$\tilde{\mathcal{R}}_k(\boldsymbol{\omega}_{k,I}, \boldsymbol{\Psi}_I) \triangleq \sum_{I \in \mathcal{L}} (\mathcal{U}_{k,I}(\boldsymbol{\omega}_{k,I}, \boldsymbol{\Psi}) - \mathcal{V}_{k,I}(\boldsymbol{\omega}_{k,I}, \boldsymbol{\Psi})), \forall k \in \mathcal{K}. \quad (5.103)$$

This completes the proof. ■

The lower bound presented is more manageable compared to the original data-rate function $\tilde{\mathcal{R}}_k(\boldsymbol{\omega}_{k,I}, \boldsymbol{\Psi}_I)$ in (5.73). However, this bound still involves coupled optimization variables. To address this, we implement the AO approach. Specifically, we optimize $\boldsymbol{\omega}_{k,I}$ and $\boldsymbol{\Psi}_I$ by alternatively refining each variable while keeping the others constant.

5.10.1 Step One: Optimizing $\boldsymbol{\Omega}_{k,I}$ with Fixed $\boldsymbol{\Psi}_I$

In this sub-problem, we presume that the passive reflecting elements at the IRS, i.e., $\boldsymbol{\Psi}_I$ are fixed, and we proceed with designing the active beamformers, $\boldsymbol{\omega}_{k,I}$, at the BS. By employing the principles of Semi-Definite Programming (SDP), we derive the following:

$\Omega_{k,l} = \omega_{k,l}\omega_{k,l}^H \in \mathbb{S}_+^M$ and $\mathbf{H}_k = \mathbf{h}_k\mathbf{h}_k^H \in \mathbb{S}_+^M$, $\forall k \in \mathcal{K}$. Leveraging on the aforementioned SDP relaxation, we can rewrite the SINR in (5.72) as follows:

$$\Gamma_{k,l} = \frac{\text{Tr}(\mathbf{h}_k^H \Omega_{k,l})}{\sum_{i \in \mathcal{K}, i \neq k} \text{Tr}(\mathbf{h}_k^H \Omega_{i,l}) + \sigma_k^2}, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \quad (5.104)$$

Consequently, the data-rate function defined in (5.73) can be reshaped as:

$$\mathcal{R}_k(\Omega_{k,l}) = \mathcal{U}_k(\Omega_{k,l}) - \mathcal{V}_k(\Omega_{k,l}), \quad \forall k \in \mathcal{K}. \quad (5.105)$$

With the aid of the SDP transformations, the EE optimization problem P_9 can be reformulated as follows:

$$P_{10} : \max_{\Omega_{k,l}} \eta_{\text{eff}}(\Omega_{k,l}) \quad (5.106a)$$

$$\text{s.t.} : \mathcal{R}_k(\Omega_{k,l}) \geq R_{\min,k}, \quad \forall k \in \mathcal{K}, \quad (5.106b)$$

$$\Omega_{k,l} \succeq \mathbf{0}, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \quad (5.106c)$$

$$\text{rank}(\Omega_{k,l}) \leq 1, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \quad (5.106d)$$

$$\sum_{l \in \mathcal{L}} \sum_{k \in \mathcal{K}} \text{Tr}(\Omega_{k,l}) \leq p_{\max}, \quad (5.106e)$$

$$\text{Tr}(\Omega_{k,l}) = 0, \quad \forall k \in \mathcal{K}, \forall l \geq T_k. \quad (5.106f)$$

It is important to note that the constraint (5.106b) in P_{10} does not exhibit concavity. To circumvent this non-concavity, we employ the outcome from *Lemma 2*. A proposed surrogate lower bound of $\mathcal{R}_k(\Omega_{k,l})$ is introduced, where it is guaranteed that $\mathcal{R}_k(\Omega_{k,l}) \geq \tilde{\mathcal{R}}_k(\Omega_{k,l})$. Hence, we can reexpress P_{10} as follows:

$$P_{11} : \max_{\Omega_{k,l}} \tilde{\eta}_{\text{eff}}(\Omega_{k,l}) = \frac{\sum_{k \in \mathcal{K}} \tilde{\mathcal{R}}_k(\Omega_{k,l})}{\mathcal{E}_{\text{tot}}(\Omega_{k,l})} \quad (5.107a)$$

$$\text{s.t.} : \tilde{\mathcal{R}}_k(\Omega_{k,l}) \geq R_{\min,k}, \quad \forall k \in \mathcal{K}, \quad (5.107b)$$

$$\Omega_{k,l} \succeq \mathbf{0}, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \quad (5.107c)$$

$$\text{rank}(\Omega_{k,l}) \leq 1, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \quad (5.107d)$$

$$\sum_{l \in \mathcal{L}} \sum_{k \in \mathcal{K}} \text{Tr}(\Omega_{k,l}) \leq p_{\max}, \quad (5.107e)$$

$$\text{Tr}(\Omega_{k,l}) = 0, \quad \forall k \in \mathcal{K}, \forall l \geq T_k, \quad (5.107f)$$

where

$$\mathcal{E}_{\text{tot}}(\Omega_{k,l}) = \sum_{l \in \mathcal{L}} \sum_{k \in \mathcal{K}} \text{Tr}(\Omega_{k,l}) + P_s + N P_d + P_c^{\text{BS}}. \quad (5.108)$$

The numerator of the objective function and constraint (5.107b) exhibit the form of convex-concave functions, which introduces favorable properties to the optimization problem. Nev-

ertheless, despite this advantageous characteristic, the overall problem remains non-convex due to the presence of the non-convex rank-one constraint (5.106d) and the fractional objective function. As a result, resolving this optimization problem necessitates the application of specialized techniques that can effectively address these non-convexities. Our forthcoming focus firstly centers on resolving the non-convex rank-one constraint (5.106d), which emerges within the context of our optimization problem. Fortunately, an innovative method has been proposed in [208], specifically devised to address these types of rank constraints. We hereby direct our attention to the ensuing proposition.

Proposition 3 Consider a nonzero positive semidefinite beamforming matrix, denoted as $\Omega_{k,l} \in \mathbb{S}_+^M$, where M represents the dimensionality of the matrix. We assert that $\Omega_{k,l}$ is a rank one matrix if and only if the inequality $\varpi_{k,l} \mathbf{I}_{M-1} - \mathcal{U}^T \Omega_{k,l} \mathcal{U} \succeq \mathbf{0}$ holds true, where $\varpi_{k,l} = 0$, \mathbf{I}_{M-1} denotes an identity matrix with dimension $M-1$, and $\mathcal{U} \in \mathbb{R}^{M \times (M-1)}$ corresponds to the eigenvectors corresponding to the $M-1$ smallest eigenvalues of $\Omega_{k,l}$.

Proof 5 Assuming the nonnegative eigenvalues of $\Omega_{k,l}$ are arranged in descending order as $[\kappa_M; \kappa_{M-1}; \dots; \kappa_1]$, we can exploit the relationship between an eigenvector's Rayleigh quotient and its associated eigenvalue. Consequently, the matrix $\varpi_{k,l} \mathbf{I}_{M-1} - \mathcal{U}^T \Omega_{k,l} \mathcal{U}$, $\forall k \in \mathcal{K}, \forall l \in \mathcal{L}$, takes on the form of a diagonal matrix, with its diagonal elements set as $[\varpi_{k,l} - \kappa_{M-1}; \varpi_{k,l} - \kappa_{M-2}; \dots; \varpi_{k,l} - \kappa_1]$. In light of this, we observe that $\Omega_{k,l}$ possesses $M-1$ smallest eigenvalues all being zero if and only if the conditions $\varpi_{k,l} \mathbf{I}_{M-1} - \mathcal{U}^T \Omega_{k,l} \mathcal{U} \succeq \mathbf{0}$ and $\varpi_{k,l} = 0, \forall k \in \mathcal{K}, \forall l \in \mathcal{L}$, are satisfied simultaneously. Consequently, $\Omega_{k,l}$ qualifies as a rank one matrix under these conditions [208]. \blacksquare

According to proposition 3, we can replace the rank constraint (5.107b) with a positive semidefinite constraint:

$$\varpi_{k,l} \mathbf{I}_{M-1} - \mathcal{U}^{(q)\top} \Omega_{k,l} \mathcal{U}^{(q)} \succeq \mathbf{0}, \forall k \in \mathcal{K}, \forall l \in \mathcal{L}. \quad (5.109)$$

Given the unavailability of $\mathcal{U}^{(q)\top}$, we resort to the SCA method and employ the smallest eigenvectors of $\Omega_{k,l}$. To achieve the goal of ultimately having $\varpi_{k,l} = 0$ while facilitating the attainment of an initial feasible point, we introduce a penalty term for $\varpi_{k,l}$ in the objective function. Thus, at SCA iteration (q) , the following convex problem is addressed:

$$P_{12} : \max_{\Omega_{k,l}, \varpi_{k,l}} \tilde{\eta}_{eff}(\Omega_{k,l}) - \varsigma^{(q)} \sum_{l \in \mathcal{L}} \sum_{k \in \mathcal{K}} \varpi_{k,l} \quad (5.110a)$$

$$s.t. : \tilde{\mathcal{R}}_k(\Omega_{k,l}) \geq R_{\min,k}, \quad \forall k \in \mathcal{K}, \quad (5.110b)$$

$$\Omega_{k,l} \succeq \mathbf{0}, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \quad (5.110c)$$

$$\varpi_{k,l} \mathbf{I}_{M-1} - \mathcal{U}^{(q)\top} \Omega_{k,l} \mathcal{U}^{(q)} \succeq \mathbf{0}, \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \quad (5.110d)$$

$$\sum_{l \in \mathcal{L}} \sum_{k \in \mathcal{K}} \text{Tr}(\Omega_{k,l}) \leq p_{\max}, \quad (5.110e)$$

$$\text{Tr}(\Omega_{k,l}) = 0, \quad \forall k \in \mathcal{K}, \forall l \geq T_k, \quad (5.110f)$$

We emphasize that the penalty factor's updates follow an interior iterative sub-algorithm with the formula $\varsigma^{(q)} = \min(\nu\varsigma^{(q-1)}, \vartheta_{\max})$, where ν represents a positive coefficient factor, and ϑ_{\max} denotes the maximum allowable penalty factor.

The objective function in P_{12} is a fractional function, rendering the optimization problem to be still non-convex. To tackle this final issue, we adopt a strategy to handle fractional programming problems. Specifically, we employ the widely recognized Dinkelbach method [209], with (d) as its iteration, to address the fractional form of the optimization problem. Denoting $\tilde{\eta}_{\text{eff}}^*(\boldsymbol{\Omega}_{k,l})$ as the optimal energy efficiency point within the feasible solution set defined by its constraints, we then transform the problem into a non-fractional optimization problem for further resolution. Thus, the problem can be restated as follow:

$$P_{13} : \max_{\boldsymbol{\Omega}_{k,l}, \varpi_{k,l}} \sum_{k \in \mathcal{K}} \tilde{\mathcal{R}}_k(\boldsymbol{\Omega}_{k,l}^{(d)}) - \varrho^{(d)} \mathcal{E}_{\text{tot}}(\boldsymbol{\Omega}_{k,l}^{(d)}) - \varsigma^{(d)} \sum_{l \in \mathcal{L}} \sum_{k \in \mathcal{K}} \varpi_{k,l} \quad (5.111a)$$

$$\text{s.t.} : \tilde{\mathcal{R}}_k(\boldsymbol{\Omega}_{k,l}) \geq R_{\min,k}, \quad \forall k \in \mathcal{K}, \quad (5.111b)$$

$$\boldsymbol{\Omega}_{k,l} \succeq \mathbf{0}, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \quad (5.111c)$$

$$\varpi_{k,l} \mathbf{I}_{M-1} - \mathcal{V}^{(q)\top} \boldsymbol{\Omega}_{k,l} \mathcal{V}^{(q)} \succeq \mathbf{0}, \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \quad (5.111d)$$

$$\sum_{l \in \mathcal{L}} \sum_{k \in \mathcal{K}} \text{Tr}(\boldsymbol{\Omega}_{k,l}) \leq p_{\max}, \quad (5.111e)$$

$$\text{Tr}(\boldsymbol{\Omega}_{k,l}) = 0, \quad \forall k \in \mathcal{K}, \forall l \geq T_k, \quad (5.111f)$$

where

$$\varrho^{(d)} = \max_{\boldsymbol{\Omega}_{k,l}^{(d)}} \tilde{\eta}_{\text{eff}}^*(\boldsymbol{\Omega}_{k,l}^{(d)}), \quad (5.112)$$

and $\varsigma^{(d)}$ is the new penalty factor update rules:

$$\varsigma^{(d)} = \min(\nu\varsigma^{(d-1)}, \vartheta_{\max}) \times \mathcal{E}_{\text{tot}}(\boldsymbol{\Omega}_{k,l}^{(d)}). \quad (5.113)$$

Finally, the convexity of the objective function in P_{13} with respect to the active beamforming variables can be demonstrated through a formal proof based on the following proposition.

Proposition 4 *The optimal energy efficiency, denoted as $\tilde{\eta}_{\text{eff}}^*(\boldsymbol{\Omega}_{k,l}^*)$, serves as a means to derive the resource allocation policy if and only if*

$$\begin{aligned} \max_{\boldsymbol{\Omega}_{k,l}} \sum_{k \in \mathcal{K}} \tilde{\mathcal{R}}_k(\boldsymbol{\Omega}_{k,l}^{(d)}) - \varrho^{(d)} \mathcal{E}_{\text{tot}}(\boldsymbol{\Omega}_{k,l}^{(d)}) - \varsigma^{(q)} \sum_{l \in \mathcal{L}} \sum_{k \in \mathcal{K}} \varpi_{k,l} = \\ \sum_{k \in \mathcal{K}} \tilde{\mathcal{R}}_k(\boldsymbol{\Omega}_{k,l}^*) - \varrho^* \mathcal{E}_{\text{tot}}(\boldsymbol{\Omega}_{k,l}^*) - \varsigma^{(q)} \sum_{l \in \mathcal{L}} \sum_{k \in \mathcal{K}} \varpi_{k,l} = 0, \end{aligned} \quad (5.114)$$

for $\sum_{k \in \mathcal{K}} \tilde{\mathcal{R}}_k(\boldsymbol{\Omega}_{k,l}^*) \geq 0$ and $\mathcal{E}_{\text{tot}}(\boldsymbol{\Omega}_{k,l}^*) \geq 0$, where $\boldsymbol{\Omega}_{k,l}^*$ provides the optimal solution to P_{13} .

Proof 6 Let's denote ϱ^* be the optimal solution, corresponding to the optimal resource allocation policy $\boldsymbol{\Omega}_{k,l}^*$ of the objective function in P_{13} , that is:

$$\varrho^* = \max_{\boldsymbol{\Omega}_{k,l}} \frac{\sum_{k \in \mathcal{K}} \tilde{\mathcal{R}}_k(\boldsymbol{\Omega}_{k,l})}{\mathcal{E}_{\text{tot}}(\boldsymbol{\Omega}_{k,l})} - \varsigma^{(q)} \sum_{l \in \mathcal{L}} \sum_{k \in \mathcal{K}} \varpi_{k,l} \quad (5.115)$$

The optimal EE can then be calculated as follows:

$$\begin{aligned}\varrho^* &= \frac{\sum_{k \in \mathcal{K}} \tilde{\mathcal{R}}_k^*(\boldsymbol{\Omega}_{k,I}^*)}{\mathcal{E}_{\text{tot}}^*(\boldsymbol{\Omega}_{k,I}^*)} - \varsigma^{(q)} \sum_{l \in \mathcal{L}} \sum_{k \in \mathcal{K}} \varpi_{k,l} \\ &\geq \frac{\sum_{k \in \mathcal{K}} \tilde{\mathcal{R}}_k(\boldsymbol{\Omega}_{k,I})}{\mathcal{E}_{\text{tot}}^{(d)}(\boldsymbol{\Omega}_{k,I})} - \varsigma^{(q)} \sum_{l \in \mathcal{L}} \sum_{k \in \mathcal{K}} \varpi_{k,l}.\end{aligned}\quad (5.116)$$

It becomes readily apparent that:

$$\sum_{k \in \mathcal{K}} \tilde{\mathcal{R}}_k(\boldsymbol{\Omega}_{k,I}) - \varrho^* \mathcal{E}_{\text{tot}}(\boldsymbol{\Omega}_{k,I}) - \varsigma^{(q)} \sum_{l \in \mathcal{L}} \sum_{k \in \mathcal{K}} \varpi_{k,l} \mathcal{E}_{\text{tot}}(\boldsymbol{\Omega}_{k,I}) \leq 0, \quad (5.117)$$

Thus, it can be inferred that:

$$\sum_{k \in \mathcal{K}} \tilde{\mathcal{R}}_k(\boldsymbol{\Omega}_{k,I}^*) - \varrho^* \mathcal{E}_{\text{tot}}(\boldsymbol{\Omega}_{k,I}^*) - \varsigma^{(q)} \sum_{l \in \mathcal{L}} \sum_{k \in \mathcal{K}} \varpi_{k,l} \mathcal{E}_{\text{tot}}(\boldsymbol{\Omega}_{k,I}^*) = 0, \quad (5.118)$$

Therefore, we have:

$$\begin{aligned}\max_{\boldsymbol{\Omega}_{k,I}} \sum_{k \in \mathcal{K}} \tilde{\mathcal{R}}_k(\boldsymbol{\Omega}_{k,I}) - \varrho^* \mathcal{E}_{\text{tot}}(\boldsymbol{\Omega}_{k,I}) \\ - \varsigma^{(q)} \sum_{l \in \mathcal{L}} \sum_{k \in \mathcal{K}} \varpi_{k,l} \mathcal{E}_{\text{tot}}(\boldsymbol{\Omega}_{k,I}) = 0,\end{aligned}\quad (5.119)$$

and this can be attained through the resource allocation policy. Thus, this concludes the proof. \blacksquare

The optimization problem P_{13} is now a convex problem and can be efficiently solved by standard convex optimization solvers such as CVX [210]. Ultimately, we detail our proposed algorithm in **Algorithm 5**.

5.10.2 Step Two: Optimizing $\boldsymbol{\Psi}_I$

In the second sub-problem, equipped with the optimal active beamforming matrices $\boldsymbol{\omega}_{k,I}$ from the preceding subproblem, we progress with the optimization of the passive reflecting elements at the IRS, denoted by $\boldsymbol{\Psi}_I$. Given that the semidefinite matrix $\boldsymbol{\omega}_{k,I}$ is provided, the optimization problem in P_9 shifts towards the maximization of the data-rate. The major hurdle in optimizing the phase shifts at the IRS arises due to the constraint (5.80b). Particularly, the constraint (5.80b) imposes a unit-modulus constraint, which presents a significant challenge in the quest to solve the problem. As such, we start by defining $\mathbf{e} = (e^{j\phi_1}, \dots, e^{j\phi_N})^H \in \mathbb{C}^{N \times 1}$ and $\tilde{\mathbf{e}} = [\mathbf{e}^T \psi]^T \in \mathbb{C}^{(N+1) \times 1}$, where $\psi \in \mathbb{C}$ is a dummy variable with $|\psi| = 1$. To aid in crafting the solution, we further define $\mathbf{E} = \tilde{\mathbf{e}} \tilde{\mathbf{e}}^H \in \mathbb{C}^{(N+1) \times (N+1)}$. Consequently, we derive the following:

$$\begin{aligned}|(\mathbf{h}_{\text{IRS},k}^H \boldsymbol{\Psi}_I \mathbf{h} + \mathbf{h}_{\text{BS},k}^H) \boldsymbol{\Omega}_{k,I}|^2 &\triangleq \text{Tr}(\mathbf{E} \mathbf{Z}_k \boldsymbol{\Omega}_{k,I} \mathbf{Z}_k^H) \\ &= \text{Tr}(\boldsymbol{\Omega}_{k,I} \mathbf{Y}_k), \quad \forall k \in \mathcal{K},\end{aligned}\quad (5.120)$$

Algorithm 5 Iterative SCA-based Resource Allocation Algorithm for Energy Efficient Multi-user IRS-Assisted URLLC Systems Based on a Novel Rank Relaxation Method

Input: Set $d = 0$ for the Dinkelbach procedure, set the maximum number of iteration D_{\max} , initialize the beamformer matrix $\Omega_{k,l} = \Omega_{k,l}^0$, initialize ς^0 , set $\vartheta_{\max} \gg 1$, $\nu > 1$, and set the tolerance $\varepsilon = 10^{-3}$.

- 1: **repeat**
- 2: Calculate $\tilde{\mathcal{R}}_k(\omega_{k,l}, \Psi_l)$ for a given Ψ_l .
- 3: Calculate the rank-one relaxation constraint in (5.109).
- 4: Solve P_{13} for $\varrho^{(d-1)}$.
- 5: **if** $\left| \sum_{k \in \mathcal{K}} \tilde{\mathcal{R}}_k(\Omega_{k,l}^{(d)}) - \varrho^{(d-1)} \mathcal{E}_{\text{tot}}(\Omega_{k,l}^{(d)}) \right| - \varsigma^{(d)} \sum_{l \in \mathcal{L}} \sum_{k \in \mathcal{K}} \varpi_{k,l} \leq \varepsilon$
- 6: **return** $\Omega_{k,l} = \Omega_{k,l}^{(d)}$, $\varrho^* = \varrho^{(d-1)}$.
- 7: **else** $\varrho^{(d)} = \tilde{\eta}_{\text{eff}}^{(d)}(\Omega_{k,l}^{(d)})$, **end if**.
- 8: Update $\varsigma^{(d)} = \min(\nu \varsigma^{(d-1)}, \vartheta_{\max})$.
- 9: $d \leftarrow d + 1$.
- 10: **until** $d = D_{\max}$.
- 11: **return** $\Omega_{k,l} = \Omega_{k,l}^*$.

where

$$\mathbf{Z}_k = \left[(\text{diag}(\mathbf{h}_{\text{IRS},k}^H) \mathbf{H})^T \mathbf{h}_{\text{BS},k}^* \right]^T, \quad \forall k \in \mathcal{K}, \quad (5.121)$$

$$\mathbf{Y}_k = \mathbf{Z}_k^H \mathbf{E} \mathbf{Z}_k, \quad \forall k \in \mathcal{K}. \quad (5.122)$$

In a manner akin to P_{11} , we address the non-convex constraint (5.80a) and the objective function. To achieve this, we employ *Lemma 2* and subsequently rewrite the data-rate function as:

$$\tilde{\mathcal{R}}_k(\Psi_l) = \mathcal{U}_k(\Psi_l) - \mathcal{V}_k(\Psi_l), \quad \forall k \in \mathcal{K}. \quad (5.123)$$

Now, we restate the optimization problem P_9 as follows:

$$\begin{aligned} P_{14} : & \max_{\mathbf{E}, \Psi_l} \sum_{k \in \mathcal{K}} \tilde{\mathcal{R}}_k(\Psi_l) \\ \text{s.t. :} & \tilde{\mathcal{R}}_k(\Psi_l) \geq R_{\min,k}, \quad \forall k \in \mathcal{K}, \\ & \text{diag}(\Psi_l) = \mathbf{1}_{N+1}, \quad \forall l \in \mathcal{L}, \\ & \text{rank}(\mathbf{E}) \leq 1, \\ & \Psi_l \succeq 0, \quad \forall l \in \mathcal{L}, \\ & \mathbf{E} \succeq \mathbf{0}. \end{aligned} \quad (5.124a)$$

Following a similar approach as in P_{14} , and drawing on the insights from *proposition 3*, we can substitute the rank constraint (5.124c) with a positive semidefinite constraint:

$$\zeta_l \mathbf{I}_N - \Upsilon^{(g)} \Upsilon^{(g)T} \succeq \mathbf{0}, \quad \forall l \in \mathcal{L}. \quad (5.125)$$

where $\Upsilon^{(g)}$ is an $N \times (N+1)$ matrix at the (g) -th SCA iteration with its columns corresponding to the smallest N eigenvectors of Ψ_I . Furthermore, to sustain a rank of one for Ψ_I , equation (5.125) must hold true, assuming $\zeta_I = 0$. Hence, by incorporating ζ_I as a penalty term into the objective function of P_{14} , we write the subsequent convex problem:

$$P_{15} : \max_{\mathbf{E}, \Psi_I} \sum_{k \in \mathcal{K}} \tilde{\mathcal{R}}_k(\Psi_I) - \varkappa^{(g)} \sum_{l \in \mathcal{L}} \zeta_l \quad (5.126a)$$

$$s.t. : \tilde{\mathcal{R}}_k(\Psi_I) \geq R_{\min, k}, \quad \forall k \in \mathcal{K}, \quad (5.126b)$$

$$\text{diag}(\Psi_I) = \mathbf{1}_{N+1}, \quad \forall l \in \mathcal{L}, \quad (5.126c)$$

$$\zeta_I \mathbf{I}_N - \Upsilon^{(g)\top} \Psi_I \Upsilon^{(g)} \succeq \mathbf{0}, \quad \forall l \in \mathcal{L} \quad (5.126d)$$

$$\Psi_I \succeq 0, \quad \forall l \in \mathcal{L}, \quad (5.126e)$$

$$\mathbf{E} \succeq \mathbf{0}. \quad (5.126f)$$

where $\varkappa^{(g)}$ is a sequences of increasing weights. The optimization problem P_{15} now can be efficiently solved just as P_{13} [210]. The solution of these two sub-problems yields the suboptimal solution of P_9 .

Proposition 5 *The objective function of P_9 is ensured to be monotonically non-decreasing throughout the iterations of the proposed algorithm.*

Proof 7 Let's denote the objective functions of P_9 , P_{13} , and P_{15} respectively as J_{P_9} , $\mathsf{J}_{P_{13}}$, and $\mathsf{J}_{P_{15}}$. Also, in consideration of $\{\Omega_{k,l}^s, \Psi_I^s\}$ and $\{\Omega_{k,l}^{s-1}, \Psi_I^{s-1}\}$ as the feasible solutions of P_9 in the s -th and $(s-1)$ -th iterations respectively, we can establish the following inequalities:

$$\begin{aligned} \mathsf{J}_{P_9}(\Omega_{k,l}^s, \Psi_I^s) &= \mathsf{J}_{P_{13}}(\Omega_{k,l}^s, \Psi_I^s) \\ &\geq \mathsf{J}_{P_{13}}(\Psi_I^{s-1}) = \mathsf{J}_{P_9}(\Omega_{k,l}^{s-1}, \Psi_I^{s-1}), \end{aligned} \quad (5.127)$$

$$\begin{aligned} \mathsf{J}_{P_9}(\Omega_{k,l}^s, \Psi_I^{s-1}) &= \mathsf{J}_{P_{15}}(\Omega_{k,l}^s, \Psi_I^*) \\ &\geq \mathsf{J}_{P_{15}}(\Omega_{k,l}^{s-1}) = \mathsf{J}_{P_9}(\Omega_{k,l}^{s-1}, \Psi_I^s). \end{aligned} \quad (5.128)$$

By utilizing inequalities (5.127) and (5.128), we can guarantee the improvement in the value of the objective function of P_9 after every iteration. ■

Ultimately, the final iterative AO algorithm that solves subproblems is presented in **Algorithm 6**.

Algorithm 6 Proposed Iterative AO algorithm Toward Energy Efficient Multi-user IRS-Assisted URLLC Systems Based on a Novel Rank Relaxation Method

Input: Set $s = 0$, maximum number of iteration S_{\max} , and initialize the beamformer matrix and phase shifts as $\Omega_{k,l} = \Omega_{k,l}^0$ and $\Psi_l = \Psi_l^0$, respectively.

1: **repeat**

2: Solve problem P_{13} for given Ψ_l^s , and obtain the optimal solution $\Omega_{k,l}^s$ based on **Algorithm 5**.

3: Solve problem P_{15} for given $\Omega_{k,l}^s$, and obtain the optimal solution Ψ_l^s .

4: $s \leftarrow s + 1$.

5: **until** $s = S_{\max}$

6: **return** $\{\Omega_{k,l}^*, \Psi_l^*\} = \{\Omega_{k,l}^s, \Psi_l^s\}$.

Table 5.2: Simulation Parameters for Multi-user IRS-Assisted URLLC Systems.

Parameter	Value
Area dimensions	(100, 100) meters
AP location	(0, 0) meters
IRS location	(50, 0) meters
Path loss model	$35.3 + 37.6 \log_{10}(d_k)$ dB
AO convergence tolerance	10^{-2}
Thermal noise density	-174 dBm/Hz
Decoding error probability, ϵ_k	10^{-7}
Number of users, K	4
Number of AP antennas, M	5
Number of IRS elements, N	20
Minimum transmit power, p_{\max}	30 dBm
Block code length, m_d	250 symbols
Threshold rate, R_{th}^k	1.6 bits/Sec/Hz

5.11 Numerical Results for the URLLC-enabled IRS-assisted Network

This section provides simulation results to investigate the efficiency of the proposed algorithm in downlink MISO URLLC systems enabled IRS, utilizing finite blocklength codes. A rectangular region with dimensions of (100, 100) meters is considered. The BS is positioned at (0, 0) meters; the IRS is located at (50, 0) meters and all users are scattered randomly within the boundaries of this rectangle region. The path loss model is expressed as $35.3 + 37.6 \log_{10}(d_k)$ [dBm], where d_k refers to the distance in meters between the BS and the k -th URLLC user. The convergence tolerance for the proposed AO algorithm, which is based on rank relaxation, SCA, and the Dinkelback method, is set at 10^{-2} . It is assumed that the thermal noise density stands at -174 [dBm/Hz]. In addition, the maximum probability of decoding error for URLLC user k , represented as ϵ_k , is set to be 10^{-7} . Furthermore, the simulation parameters are configured with $K = 4$, $M = 5$, and $R_{\min,k} = 1.6$ [bits/Sec/Hz], in accordance with the methodology outlined in references [8, 3]. All simulation parameters

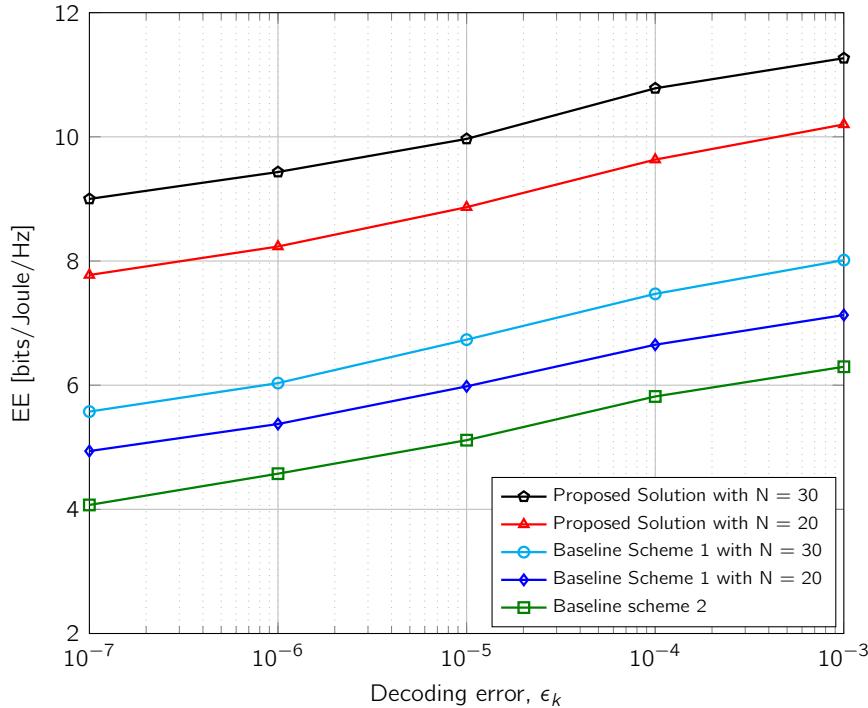


Figure 5.7: Impact of tolerable decoding error, ϵ_k , on the EE in an URLLC-enabled IRS-assisted network.

are also summarized in Table 5.2.

Fig. 5.7 visualizes the relationship between mean EE and various maximum decoding error probability thresholds, denoted by $\epsilon_{k,\max}$, while maintaining a constant block code length of 250 symbols ($m_d = 250$). The graph highlights a noticeable upward trajectory in EE as the tolerance for decoding error widens. This enhancement in EE can be directly linked to the inverse relationship between $Q^{-1}(\epsilon_k)$ and ϵ_k , where an increased allowable decoding error probability leads to a reduced rate loss $\mathcal{V}_k(\omega_{k,l}, \Psi_l)$ as specified in (5.73). Consequently, this dynamic facilitates the fulfillment of minimal data-rate requirements at lower transmission powers, effectively boosting EE. Furthermore, the graph illustrates an incremental gain in EE with an increase in the number of IRS's reflecting elements, showcasing the significant impact of IRS size on system efficiency.

In addition to showcasing these trends, the figure conducts a comparative analysis with two foundational models for further context. The first baseline scheme adheres to a static beamforming approach at the IRS, while the second baseline scheme operates under the assumption of an IRS-free environment. The empirical evidence, as delineated in the figure, unequivocally demonstrates the enhanced performance of our proposed approach over the IRS-less baseline scheme 2. This enhanced efficacy is attributed to the strategic deployment of IRS technology coupled with the synchronized optimization of both active and passive beamforming matrices at the BS and IRS, respectively. When comparing against baseline scheme 1, our method exhibits an improvement, thereby emphasizing the critical value of

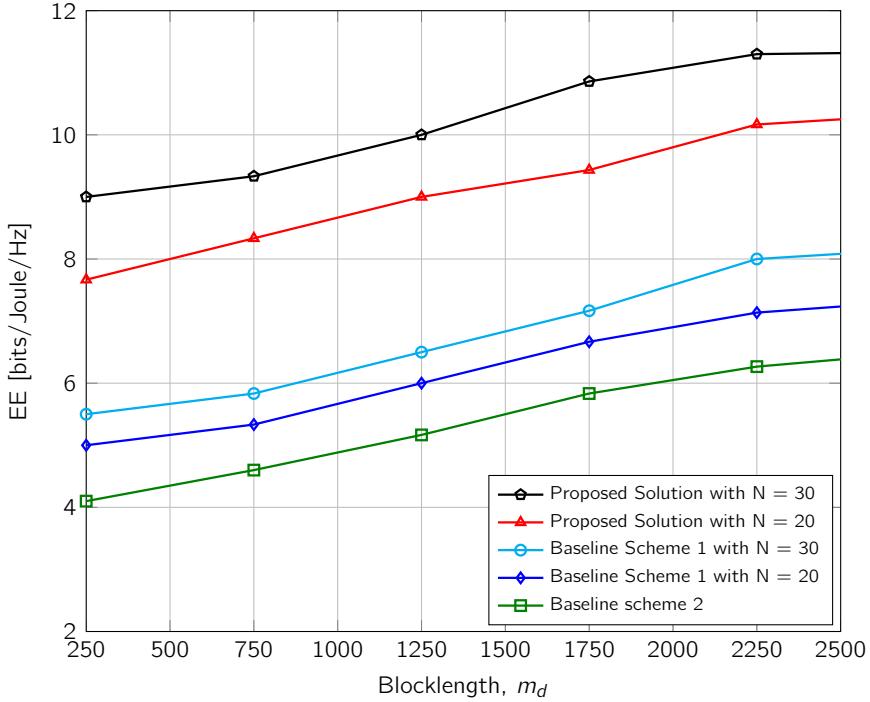


Figure 5.8: Impact of the blocklength, m_d , on the EE in an URLLC-enabled IRS-assisted network.

IRS integration and optimization within the system. This detailed comparison validates the superiority of our proposed methodology and reinforces the role of the IRS in elevating the EE of URLLC-enabled wireless communication systems, particularly under the stringent constraints of short packet transmissions and variable decoding error probabilities.

Fig. 5.8 illustrates the impact of the blocklength size, denoted as m_d , on the achievable EE at a constant $\epsilon_k = 10^{-7}$. With an increase in m_d , the EE initially experiences a moderate boost, followed by a slow elevation until it plateaus. In comparison to the scenario with an IRS, the scenario without it consistently exhibits a lower EE, independent of the quantity of reflecting elements. Smaller N values necessitate higher transmit power to meet QoS standards, leading to a reduction in EE. This reduction arises from the amplified interference in the data-rate function due to multi-user interference in the SINR function, resulting in an increase in transmit power, a decrease in data-rate, and consequently a negatively pruned EE.

Fig. 5.8 illustrates the impact of blocklength size, represented by m_d , on the EE of the system, maintaining a consistent decoding error probability of $\epsilon_k = 10^{-7}$. The figure shows a relationship between m_d and EE: as m_d increases, there is a discernible, albeit moderate, uplift in EE, which gradually ascends to a stable plateau. This pattern highlights the balance between blocklength size and system efficiency, where longer blocklengths, up to a certain threshold, contribute to enhancing EE.

The contrast between scenarios with and without the deployment of an IRS is starkly de-

picted. Invariably, configurations lacking an IRS lag in EE performance across all tested blocklengths, underscoring the IRS's role in energy optimization. Notably, the presence of fewer reflecting elements, denoted by smaller N values, compels the system to elevate transmit power levels to adhere to Quality of Service (QoS) criteria, inadvertently impacting EE negatively. This effect is primarily attributed to the escalated interference within the data-rate function, a consequence of multi-user interference affecting the SINR, which necessitates an upsurge in transmit power. Such an increase, while aimed at maintaining service standards, paradoxically diminishes data-rate efficiency and, by extension, truncates EE.

This phenomenon demonstrates a critical insight: the size of m_d and the strategic deployment of IRS elements play crucial roles in achieving optimal EE. Especially in multi-user environments where interference is a significant concern, the IRS emerges as a vital tool for mitigating adverse effects on EE. Through careful management of blocklength and leveraging the benefits of IRS technology, URLLC systems can help overcome the complex interplay between QoS demands and EE, optimizing performance across these metrics.

5.12 Conclusion

This chapter delved into the resource management problem within downlink multi-user configurations, facilitated by Intelligent Reflecting Surfaces (IRS) for enhancing both Machine-Type Communication (MTC) and Ultra-Reliable Low-Latency Communications (URLLC) systems. Central to our investigation was the strategic design of resource allocation mechanisms through active and passive beamforming, aimed at optimizing Energy Efficiency (EE) across these network models.

In the first part of this chapter which was dedicated to MTC-enabled networks, our analysis focused on achieving a dual objective: maximizing EE while also broadening user admittance, framed within the context of a Multi-Objective Optimization Problem (MOOP). This approach was tailored to address the scalable nature of MTC networks, which are characterized by their need for efficient connectivity across a multitude of devices.

Transitioning to URLLC systems, the complexity inherent in meeting URLLC standards necessitated a more focused objective, concentrating solely on maximizing EE. This refinement acknowledges the strict latency and reliability demands of URLLC applications, which significantly influence the formulation and optimization of resource allocation strategies.

Both the MTC and URLLC models were designed with a keen emphasis on meeting the diverse Quality of Service (QoS) requirements of different user types, incorporating the challenges associated with short packet transmissions. Given the non-convex nature of the primary problem, we employed the Alternating Optimization (AO) technique to decompose the optimization problem into two distinct, more tractable sub-problems: optimizing the active beamformers at the transmitter and adjusting the phase shifts at the IRS.

Further advancing our methodological approach, we introduced an innovative iterative semi-definite and rank relaxation strategy, coupled with the Successive Convex Approximation (SCA) technique and a penalty-based methodology, to effectively tackle each sub-problem.

Through simulation studies, we evaluated the efficacy of our proposed relaxation algorithms, highlighting the IRS's capability to significantly enhance EE and enable both MTC and URLLC systems to meet, and in many cases exceed, the set QoS standards—thereby outperforming existing conventional strategies.

The insights gained from our research illuminate the potential of IRS technology not just as an adjunct to existing networks but as an evolutionary mechanism that can significantly elevate system performance. These findings lay a robust foundation for future explorations into rank-constrained resource allocation strategies within IRS-aided systems. As we look ahead, the groundwork established here opens avenues for further investigations into the deployment of IRS technology, aiming to refine and extend the applicability of resource management techniques in the era of intelligent wireless communications.

Chapter 6

Joint Offloading and Resource Allocation in Uplink MEC-IRS Networks

In this chapter, we delve into the utilization of a Mobile Edge Computing (MEC) server placed at an Access Point (AP) side (or 5G Node B (gNB) side) to meet the stringent demands of delay sensitivity and reliability in multiuser Machine Type Communication (MTC) scenarios. The strategy involves offloading computational tasks to the MEC server in order to significantly reduce the latency for low-power IoT devices. Intelligent reflecting surfaces (IRSs) are employed to support the offloading process. These IRSs play a crucial role in enhancing the robustness of offloading, improving spectrum efficiency, and broadening network coverage. They achieve this by altering the propagation of incident radio-frequency waves through the adjustment of phase shifts using passive reflecting components.

This chapter represents a pivotal shift in focus from previous discussions, which were predominantly centered around downlink scenarios, to an in-depth exploration of uplink dynamics within the realm of MEC and multiuser MTC. Specifically, it highlights the strategic employment of a MEC server at an AP to tackle the critical challenges of delay sensitivity and reliability that are paramount in multiuser MTC environments. By offloading computational tasks in the uplink to the MEC server, the approach significantly alleviates latency for low-power IoT devices, thus optimizing the network's performance.

Consequently, our investigation zeroes in on the optimization of *joint* radio resource allocation and edge offloading decisions in a multiuser IRS-aided MEC network. This network is characterized by a multi-antenna AP that receives information symbols from a collection of Internet of Things (IoT) users, with these users transmitting short packets. Specifically, our objective is to minimize the overall power consumption of the system by formulating it as an optimization problem, taking into account various constraints. These include ensuring the Quality of Service (QoS) for MTC-enabled IoT users, adherence to transmit power feasibility, limitations on capacity, and restrictions related to IRS phase shifts.

The challenge lies in the non-convex nature of the problem we have formulated, which complicates the task of finding an effective solution. In response to this, we introduce a novel, efficient iterative algorithm. This algorithm leverages Successive Convex Approximation (SCA) techniques, combined with a penalty-based method to manage unit-modulus constraints that arise due to the passive reflecting elements present at the IRS. Through sim-

ulation, we demonstrate the enhanced efficiency and performance of our proposed algorithm when compared to other existing baseline methodologies. This proves the IRS's potential in substantially improving the efficiency and reliability of edge computing in multiuser MTC environments.

This chapter is based on :

J. Jalali, A. Khalili, R. Berkvens and J. Famaey, "Joint Offloading Policy and Resource Allocation in IRS-aided MEC for IoT Users with Short Packet Transmission", in *2023 IEEE 98th Vehicular Technology Conference (VTC2023-Fall)*, Hong Kong, Hong Kong, Oct. 2023, pp. 1–7. <https://doi.org/10.1109/VTC2023-Fall160731.2023.10333867>

6.1 Introduction

IN THE landscape of modern technological advancements, the proliferation of Internet of Things (MTC) applications has been remarkable, propelling the need for supporting an array of computation-heavy and latency-critical applications. Among these are cutting-edge domains such as autonomous driving, Augmented Reality (AR), Virtual Reality (VR), and Unmanned Aerial Vehicles (UAVs), which are cited extensively in the literature [211, 212, 213]. These innovative applications are designed to facilitate instantaneous interactions, whether between humans and machines or among machines themselves, thereby underpinning the essence of Machine Type Communication (MTC). As we stand on the cusp of the 6th-Generation (6G) of wireless networks, there is a pressing need to ensure these networks are robustly equipped to support a myriad of IoT devices, enabling them to perform real-time computations, communications, and control tasks seamlessly.

However, IoT devices often face significant constraints due to their inherent design considerations, which prioritize cost-effectiveness and compactness. These constraints typically manifest as limited battery life and less powerful processors. Thus, one of the paramount challenges for forthcoming IoT networks is enhancing the computational capabilities of these devices. Such an enhancement is crucial for them to efficiently manage heavy computational loads while adhering to stringent latency requirements [214]. While cloud computing has been a traditional recourse, offering abundant computational resources, it often falls short by introducing considerable computational latency. This latency is primarily attributed to the physical remoteness of cloud servers from the devices [215].

In response to these challenges, especially for mission-critical and time-sensitive applications spanning healthcare, autonomous driving, and tactical internet operations, Ultra-Reliable and Low-Latency Communication (URLLC) has emerged as a beacon of hope. URLLC is a specialized MTC service category tailored to meet the demanding reliability and latency specifications envisioned for future 6G networks [216]. It is distinguished by its capacity to achieve exceptionally low decoding error rates, below 10^{-5} , and to meet latency requirements as stringent as 1 ms. Nevertheless, it is important to acknowledge that the conventional Shannon capacity formula may not be entirely adequate for characterizing the performance within the URLLC-driven, short packet communication regime of IoT systems [3]. This recognition paves the way for exploring novel approaches and solutions to fulfill the ambitious goals set forth for 6G networks, ensuring they are capable of supporting the dynamic landscape of IoT applications with their evolving demands.

To address the challenges posed by computational latency within the IoT networks, mobile edge computing (MEC) emerges as a robust solution, offering a strategic alternative to alleviate network congestion and significantly reduce latency when compared to traditional cloud computing solutions [217]. This innovative approach involves the strategic placement of computing servers at the network's edge, such as within cellular Base Stations (BSs), gNBs, or WiFi Access Points (APs). By doing so, MEC facilitates the immediate offloading of both data and computational tasks directly from IoT devices to the nearest MEC server. This not only facilitates the data processing workflow but also enhances the quality of experience for end-users by minimizing delays and optimizing data throughput [218].

In the realm of MEC, the nature of tasks to be offloaded can vary widely, necessitating a

classification system based on their specific dependencies and the possibility of partitioning. This differentiation gives rise to two primary offloading paradigms: *partial* offloading and *binary* offloading. Each paradigm requires a delicate approach to optimize the allocation of computational and communication resources. The ultimate goals of these optimizations include minimizing energy consumption [219], reducing computation latency [220, 221], maximizing computation throughput [222], and boosting the overall energy efficiency of the system [223]. However, the effectiveness of these offloading strategies [219, 220, 221, 222, 223] can be adversely affected by wireless channel attenuation that occurs during data transmission between APs and IoT devices, potentially hampering the efficiency of the offloading process [224].

To overcome this challenge and ensure optimal offloading efficacy, the deployment of massive Multiple-Input Multiple-Output (m-MIMO) technology stands out as an enhancement strategy within MEC systems. m-MIMO technology amplifies the signal strength and improves the reliability of wireless connections, thus ensuring that the data offloading from IoT devices to MEC servers is not only feasible but performed with the highest efficiency. By employing m-MIMO, the system can counteract the detrimental effects of wireless channel attenuation, ensuring that the computational tasks are offloaded and processed with minimal latency and maximum reliability [225]. This technological synergy between MEC and m-MIMO paves the way for a new era of IoT network operations, where computational latency is significantly reduced, and the network's capacity to handle increasingly complex and latency-sensitive applications is greatly enhanced.

While the integration of m-MIMO technology into MEC systems significantly enhances offloading efficiency, it is accompanied by notable challenges, including increased energy consumption and the financial burden associated with sophisticated hardware requirements. In response to these challenges, the concept of Intelligent Reflecting Surfaces (IRSs) emerges as a groundbreaking and cost-effective approach to amplify spectral and energy efficiency across forthcoming mobile network generations [226].

An IRS is characterized by its innovative digitally-controlled meta-surface, which is comprised of a controller and numerous low-cost passive reflecting elements. These elements are distinctive because they operate without the need for any Radio-Frequency (RF) chains, thereby sidestepping the energy and cost implications associated with traditional active elements. Through precise manipulation of the phase shifts of each element under the guidance of the IRS controller, it is possible to dynamically modify the wireless propagation environment. This capability enables a range of beneficial outcomes, such as enhanced signal strength or diminished interference, customized to specific network requirements [8, 10, 3].

Furthermore, IRS technology significantly upholds task offloading efficiency within MEC networks by leveraging substantial passive beamforming gains. By optimally positioning IRSs in proximity to IoT devices, it is feasible to counteract the severe signal attenuation that often occurs due to distance challenges or Non-Line-of-Sight (NLoS) conditions. Such strategic deployment not only extends the effective service range of MEC systems but also ensures a more reliable and efficient communication link between IoT devices and MEC servers. This enhancement is pivotal for realizing the ambitious goal of equipping future IoT networks with superior computational capabilities, enabling them to support a plethora of advanced, real-time applications [227].

Fundamentally, IRS technology represents a paradigm shift in how wireless networks can optimize environmental interactions to improve performance metrics. By harmonizing with MEC architectures, IRSs unlock new possibilities for achieving unparalleled efficiency and coverage, thus setting a new benchmark for the deployment of high-performance, energy-efficient IoT networks specialized for the demands of the next-generation wireless communication.

In this chapter, our focus is on the joint uplink resource allocation design for IRS-assisted URLLC MEC systems operating with finite block-lengths. The essence of our study is captured through several key contributions, which are elaborated as follows:

- Our first major contribution involves a thorough examination of the ‘joint’ radio resource allocation and edge offloading decision-making process within an IRS-enhanced MEC network. This network architecture features a multi-antenna AP tasked with receiving information symbols from a set of MTC-enabled IoT users with finite block-length transmission. Central to our approach is the development of a sophisticated resource allocation algorithm. This algorithm is crafted to minimize the overall system power consumption, adhering to strict constraints related to peak transmit power and QoS requirements while taking interference into account
- The complexity of the problem is further underscored by its non-convex and Mixed Integer Non-Linear Programming (MINLP) nature, presenting a challenge to conventional solution strategies. To navigate these complexities, we adopt a two-step methodological approach. Firstly, we utilize Successive Convex Approximation (SCA) techniques to iteratively approach the problem’s solution. Secondly, we integrate a penalty-based framework specifically designed to address the unique unit-modulus constraints associated with the passive reflecting elements of the IRS. This innovative combination of strategies enables us to obtain a suboptimal solution that closely aligns with the ideal objectives of the system.
- Our investigation concludes with a series of simulation studies designed to validate the theoretical models and proposed solution strategies. These simulations vividly illustrate the transformative potential of integrating IRS technology within MEC-assisted URLLC frameworks. The results unambiguously demonstrate that the strategic deployment of IRS, in tandem with multi-antenna APs, is a potent enabler of both low latency and high reliability. This is a significant milestone, heralding a new era of performance capabilities for URLLC systems in the context of edge computing.

This chapter is organized as follows: Section 6.2 introduces the system and channel models. Section 6.3 formulates the proposed resource allocation problem. The resource allocation algorithm design policy is presented in Section 6.4. Section 6.5 evaluates the performance of the proposed schemes using computer simulations, and conclusions are drawn in Section 6.6.

Notations: We use the following notations in this chapter. Matrices are represented by capital boldface letters, vectors and scalars by small boldface and small normal face letters, respectively. \mathbf{I} denotes an identity matrix, and $\mathbb{C}^{x \times y}$ represents a $x \times y$ complex-valued matrix. The superscript $(\cdot)^H$ denotes the conjugate transpose of a matrix, and $(\cdot)^T$ denotes the transpose of a matrix. The notations $\mathbb{E}[\cdot]$, $tr(\cdot)$, and $diag(\cdot)$ are used to denote statistical expectation, trace, and diagonalization operator, respectively. $\nabla f(\cdot)$ represents the gradient

of the function $f(\cdot)$, and $\frac{\partial f(a)}{\partial g}$ is the first derivative of f with respect to g evaluated at $g = a$. $|\mathbf{x}|$ denotes the 2-norm of vector \mathbf{x} . The notation $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ represents the distribution of a circularly symmetric complex Gaussian (CSCG) random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, where \sim indicates “with the distribution of.”

6.2 System Model for a MEC-enabled Uplink IRS Network

This section presents the system and channel models in the IRS-assisted Orthogonal Frequency Division Multiple Access (OFDMA) MEC system for IoT users with short packet lengths. In particular, we consider a single-cell multi-user uplink communication that comprises an AP associated with a MEC server, equipped with N_{AP} antennas Fig. 6.1. The primary function of this AP is to facilitate edge computing services by capturing uplink transmissions from K single-antenna URLLC users. These users are denoted by the index k and collectively form the set $\mathcal{K} = \{1, \dots, K\}$. The transmissions from these users to the AP can be direct or mediated through an IRS. The IRS itself is comprised of M passive reflecting elements, each characterized by specific phase shifts and amplitudes.

To quantitatively describe the operational dynamics of the IRS, we introduce $\boldsymbol{\Phi} = \text{diag}(\beta_1 e^{j\alpha_1}, \beta_2 e^{j\alpha_2}, \dots, \beta_M e^{j\alpha_M})$ as the matrix of reflection coefficients. In this representation, $\beta_m \in [0, 1]$ and $\alpha_m \in (0, 2\pi]$, for each $m \in \{1, \dots, M\}$, signify the reflection amplitude and phase shift of the m -th reflective element at the IRS, respectively¹.

The system's bandwidth is partitioned into N orthogonal sub-carriers, each identified by the index set $\mathcal{N} = \{1, \dots, N\}$. The bandwidth allocated to each sub-carrier is denoted by B_s , which in turn defines the symbol duration $T_s = \frac{1}{B_s}$. In terms of temporal structure, the uplink frame is segmented into L time slots, enumerated by the set $\mathcal{L} = \{1, 2, \dots, L\}$. For analytical rigor, we assume the availability of perfect channel state information (CSI) of the entire system at the AP, serving as a theoretical performance benchmark.

The system is designed with a keen awareness of the delay requirements for all users, with the AP being privy to this critical information. This ensures that only users whose delay requirements are potentially feasible within the current resource block are considered for system admission. Furthermore, each user is associated with a computational task, denoted by (B_k, D_k) , where B_k represents the task size in bits, and D_k specifies the computation time in time slots, also referred to as the service delay. This comprehensive modeling framework lays the foundation for optimizing the performance of IRS-assisted OFDMA MEC systems, ensuring the seamless provision of edge computing services to IoT users with strict latency requirements.

6.2.1 Signal and Channel Models

In an uplink scenario, each IoT user independently sends its own signal to the network's AP. The signal that arrives at the AP within time slot l on subcarrier n is characterized by the

¹Notably, for the sake of maximizing reflection efficiency, it is generally assumed that the amplitudes of all passive elements are set to one, i.e., $\beta_m = 1$, for all m , as proposed by Basar *et al.* [160].

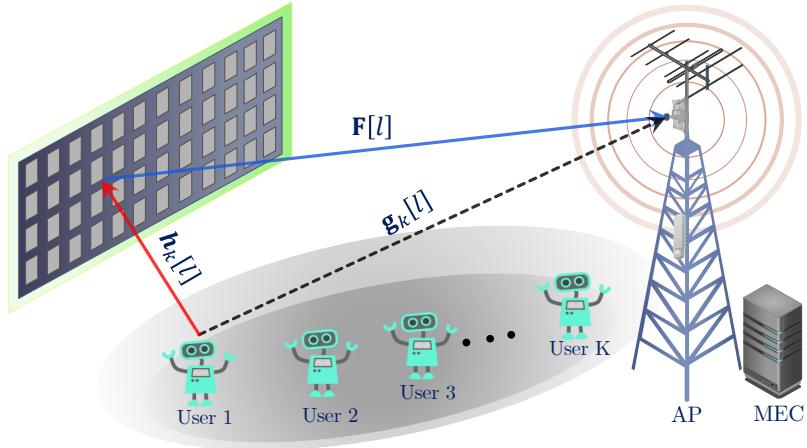


Figure 6.1: Multi-user IRS-assisted URLLC MEC system with one AP, and K IoT users with finite block-length transmission. The single-antenna MTC-enabled IoT users offload their tasks in the uplink to an MEC, directly or via IRS, using a single multi-antenna AP with an edge server.

equation:

$$\mathbf{y}[l, n] = \sum_{k=1}^K \sqrt{p_k[l, n]} (\mathbf{F}[n] \Phi \mathbf{h}_k[n] + \mathbf{g}_k[n]) u_k[l, n] + \mathbf{z}[l, n], \quad \forall l \in \mathcal{L}, \forall n \in \mathcal{N}, \quad (6.1)$$

where $\mathbf{h}_k[n] \in \mathbb{C}^{M \times 1}$ and $\mathbf{g}_k[n] \in \mathbb{C}^{N_{AP} \times 1}$ are the IRS-user and AP-user channel vectors of the k -th user. Also, $\mathbf{F}[n] \in \mathbb{C}^{N_{AP} \times M}$, $\Phi \in \mathbb{C}^{M \times M}$, $u_k[l, n] \in \mathbb{C}$, and $p_k[l, n]$ are AP-IRS channel matrix, phase shift matrix of the IRS, transmit symbol of user k on subcarrier n in time slot l , and the power of user k on subcarrier n and time slot l , respectively. Furthermore, $\mathbf{z}[l, n] \in \mathbb{C}^{N_{AP} \times 1}$ is the received noise vector at the AP with $\mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_{N_{AP}})$, and we assume $\mathbb{E}\{|u_k[l, n]|^2\} = 1, \forall k, l, n$. The received signal vector on n -th subcarrier via adopting receive beamforming is given by:

$$\tilde{\mathbf{u}}[l, n] = \mathbf{V}^H[l, n] \mathbf{y}[l, n], \quad \forall l \in \mathcal{L}, \forall n \in \mathcal{N}, \quad (6.2)$$

where $\mathbf{V}[l, n] \in \mathbb{C}^{N_{AP} \times K}$ is a matrix whose k -th columns are given by $[\mathbf{w}_k[n]] \in \mathbb{C}^{N_{AP} \times 1}, \forall k, n$. As a result, the signal-to-interference plus noise ratio (SINR) of user k on subcarrier l can be expressed as:

$$\gamma_k[l, n] = \frac{\|\mathbf{w}_k^H[n] \bar{\mathbf{h}}_k[n]\|^2 p_k[l, n]}{\sum_{j \neq k}^K \|\mathbf{w}_k^H[n] \bar{\mathbf{h}}_j[n]\|^2 p_j[l, n] + \tilde{\sigma}^2}, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \forall n \in \mathcal{N}, \quad (6.3)$$

where

$$\bar{\mathbf{h}}_k[n] = \mathbf{F}[n] \Phi \mathbf{h}_k[n] + \mathbf{g}_k[n], \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \forall n \in \mathcal{N}, \quad (6.4)$$

$$\tilde{\sigma}^2 = \sigma^2 \|\mathbf{w}_k^H[l, n]\|^2, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}. \quad (6.5)$$

6.2.2 Achievable Rate with Short Packet Transmission

Utilizing the SINR values calculated for each IoT user, we can now analyze the computation of achievable data rates, which is essential in the context of MTC-enabled systems where low-latency communication is of significant importance. The constraints of such systems inherently involve the necessity for finite and short blocklengths in transmissions, diverging from traditional models that assume infinite blocklengths where Shannon's capacity formula would apply directly. Given these conditions, the achievable rate for each user, considering the finite blocklength regime, can be expressed through an accurate approximation as follows:

$$R_k(\mathbf{p}_k, \Phi, \mathbf{x}_k) = F(\mathbf{p}_k, \Phi, \mathbf{x}_k) - G(\mathbf{p}_k, \Phi, \mathbf{x}_k), \quad \forall k \in \mathcal{K}, \quad (6.6)$$

where

$$F_k(\mathbf{p}_k, \Phi, \mathbf{x}_k) = \sum_{l=1}^L \sum_{n=1}^N \log_2 (1 + x_k[l, n] \gamma_k[l, n]), \quad \forall k \in \mathcal{K}, \quad (6.7)$$

$$G_k(\mathbf{p}_k, \Phi, \mathbf{x}_k) = Q^{-1}(\epsilon_k) \sqrt{\sum_{l=1}^L \sum_{n=1}^N x_k[l, n] V_k[l, n]}, \quad \forall k \in \mathcal{K}. \quad (6.8)$$

When formulating an optimization problem for resource allocation in IoT networks, particularly those utilizing MEC and IRS, the role of subcarrier assignment indicators becomes crucial. These indicators, denoted by $x_k[l, n]$ in (6.7) and (6.8), play a pivotal role in determining the allocation of subcarrier n in time slot l to user k . To elaborate, if subcarrier n during time slot l is assigned to user k , then $x_k[l, n] = 1$; otherwise, it is set to 0. This binary representation forms the foundation of the subcarrier allocation mechanism within the network, ensuring that each subcarrier's assignment is explicitly defined. For comprehensive optimization, the power allocation $p_k[l, n]$ optimization variables for each user k , across all time slots l and subcarriers n , are aggregated into a vector \mathbf{p}_k . Similarly, the subcarrier assignment indicators are collected into a vector \mathbf{x}_k , thereby facilitating a structured approach to resource management within the network. This vectorized representation not only simplifies the mathematical treatment of the optimization problem but also enhances the clarity of the computational model being employed.

Additionally, within the context of communication reliability, decoding error rates are symbolized by ϵ_k , reflecting the probability of erroneous interpretation of the transmitted data for user k . Moreover, the concept of channel dispersion, denoted as $V_k[l, n]$, quantifies the variability of the channel capacity around its mean, especially in scenarios involving short packet transmissions. The mathematical expression for channel dispersion incorporates the inverse of the Gaussian Q-function, $Q^{-1}(\cdot)$, a statistical tool used for mapping the relationship between the error probability and the channel's SINR. The channel dispersion $V_k[l, n]$ is calculated as:

$$V_k[l, n] = a^2 \left(1 - (1 + \gamma_k[l, n])^{-2} \right), \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \forall n \in \mathcal{N}, \quad (6.9)$$

where $a = \log_2(e)$. To meet the user's delay requirements, all symbols of user k are assigned to the first D_k time slots.

6.2.3 Offloading decision

In this section, we study the decision-making process regarding ‘task offloading’ in IoT networks, particularly those incorporating mobile edge computing MEC and IRS. A new aspect of our discussion revolves around the introduction and examination of the binary decision variable s_k , which plays an essential role in determining the offloading strategy for each IoT user. This variable serves as an indicator for the edge offloading decision for each IoT user. Specifically, when $s_k = 1$, it means that user k opts to offload its computational tasks to the edge server; conversely, a value of 0 indicates a preference for local processing.

The QoS for each IoT user k , is a concern, especially with regard to service delay D_k . To ensure adherence to QoS standards, D_k must not surpass an established acceptable maximum threshold, T_{\max} , within any given time slot. It is assumed for the purposes of this model that the data processing duration is short and that the response time delay is negligible, thereby facilitating a seamless user experience.

When the decision is made to process data locally, the computation’s energy demands are predominantly governed by the Central Processing Unit (CPU)’s power consumption. This consumption is complex, encompassing dynamic power, short circuit power, and leakage power, as detailed in the research findings presented in [228]. Further studies have demonstrated that the CPU’s power consumption at optimal operating frequency correlates directly with $(\frac{B_k}{D_k})^c$, with c representing the power scaling factor [229]. This relationship forms the basis of the proposed model to estimate local execution power consumption: Therefore, we adopt the following model to estimate the power consumption of local execution:

$$\mathcal{E}_k^{\text{loc}} = (1 - s_k) \hat{\beta} \frac{(B_k)^c}{(D_k)^c}, \quad \forall k \in \mathcal{K}, \quad (6.10)$$

where $\hat{\beta}$ is a constant value that depends on the application parameter. On the other hand, IoT users have the option and flexibility to offload their data processing tasks to the edge server via the uplink, should the need arise. Therefore, the power required for IoT user’s offloading transmission data to the edge server can be stated as follows:

$$\mathcal{E}_k^{\text{offl}} = \sum_{l \in \mathcal{L}} \sum_{n \in \mathcal{N}} s_k x_k[l, n] p_k[l, n] + s_k p_{cir}, \quad \forall k \in \mathcal{K}. \quad (6.11)$$

where p_{cir} is the constant circuit power consumption during transmission. Consequently, the total power consumption of the system in the uplink is represented by the sum of local and offloading power consumption, given by:

$$\mathcal{E}^{\text{total}} = \sum_{k \in \mathcal{K}} (\mathcal{E}_k^{\text{offl}} + \mathcal{E}_k^{\text{loc}}). \quad (6.12)$$

This comprehensive analysis not only describes the power consumption dynamics associated with edge offloading and local processing decisions but also emphasizes the importance of such decisions in optimizing the performance and sustainability of IoT networks. In the following section, we will focus on developing strategies to minimize this power consumption, thereby enhancing the efficiency of these networks.

6.3 Problem Formulation

In this section, we aim to construct a joint optimization framework that addresses both resource allocation and offloading decisions within IoT networks that benefit from the advancements of MEC and IRS. The main objective is to minimize the overall power consumption while simultaneously ensuring that the QoS requirements of MTC-enabled IoT users with short packet length transmission are met. This optimization ambitiously seeks to balance the uplink transmit power, adjust the IRS's phase shifts, and smartly allocate subcarriers, alongside making informed decisions about data offloading. The formulation of this optimization problem is precisely presented as follows:

$$\mathcal{P}_1 : \min_{\mathbf{p}, \Phi, \mathbf{s}, \mathbf{x}} \mathcal{E}^{\text{total}} \quad (6.13a)$$

$$\text{s.t. : } D_k \leq s_k T_{\max}, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \quad (6.13b)$$

$$\sum_{l=1}^L \sum_{n=1}^N x_k[l, n] p_k[l, n] \leq s_k p_{k,\max}, \quad \forall k \in \mathcal{K}, \quad (6.13c)$$

$$|\Phi_{m,m}| = 1, \quad \forall m \in \mathcal{M}, \quad (6.13d)$$

$$0 \leq \alpha_m < 2\pi, \quad \forall m \in \mathcal{M}, \quad (6.13e)$$

$$\sum_{k=1}^K x_k[l, n] \leq 1, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \quad (6.13f)$$

$$s_k \in \{0, 1\}, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \quad (6.13g)$$

$$x_k[l, n] \in \{0, 1\}, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \forall n \in \mathcal{N}. \quad (6.13h)$$

In this formulation, the sets of variables \mathbf{p} , \mathbf{x} , \mathbf{s} , and Φ embody a comprehensive collection of optimization parameters, crucial for fine-tuning system performance. These variables are utilized to optimize the system performance and make decisions related to power allocation, subcarrier assignment, offloading decisions, and phase shifts of the IRS elements, respectively.

In \mathcal{P}_1 , the constraint (6.13b) is imperative for ensuring that the service delay experienced by each user upon offloading computational tasks does not exceed a predetermined maximum threshold, denoted by T_{\max} . The condition (6.13c) effectively restricts each user's transmit power to a specified maximum limit, p_{\max} , ensuring energy-efficient operation. The constraint (6.13d) imposes the unit modulus constraint on the IRS elements, while constraint (6.13e) restricts that each reflecting element can be adjusted according to its phase. Additionally, (6.13f) mandates that each subcarrier is exclusively assigned to a single user to avoid interference. Lastly, the constraints (6.13g) and (6.13h) represent the binary nature of the subcarrier assignment and offloading decision variables.

The problem \mathcal{P}_1 is a challenging non-convex Mixed Integer Non-Linear Problem (MINLP) with interdependent optimization variables, non-convex phase shift constraints, and binary variables. Solving such non-convex optimization problems optimally is a complex task. However, we propose an efficient solution using the SCA method and a penalty-based approach

to tackle the unit-modulus constraints because of passive reflecting elements at the IRS, which offers a sub-optimal yet computationally efficient approach.

6.4 Solution of the Optimization Problem

In addressing the solution to the optimization problem outlined in \mathcal{P}_1 , our approach initiates with a strategic simplification of the delay threshold constraint, denoted by (6.13b). This initial step converts the constraint into a format that eases mathematical manipulation and subsequent analysis. Such a transformation is not only methodical but essential, as it lays the groundwork for a more linear and manageable exploration of the problem and solution space.

Given that the achievable data rate for each user k within the network is intricately linked to two pivotal factors — the size of the transmitted bitstream and the experienced delay — it becomes imperative to precisely define the relationship between these elements and the user's data rate. To this end, we derive an expression that correlates the aforementioned parameters with the achievable data rate, thereby facilitating a direct connection to the user's operational parameters as captured in the following equation derived from (6.6):

$$R_k(\mathbf{p}_k, \Phi, \mathbf{x}_k) = \frac{B_k}{D_k}, \quad \forall k \in \mathcal{K}. \quad (6.14)$$

To continue the solution process for the optimization problem initially presented as \mathcal{P}_1 , we propose an alternative formulation, denoted as \mathcal{P}_2 . This reformulation contains a revised constraint that is more amenable to analytical treatment, thus facilitating a more efficient resolution pathway. Accordingly, reformulated problem can be given as follows:

$$\mathcal{P}_2 : \min_{\mathbf{p}, \Phi, \mathbf{s}, \mathbf{x}} \mathcal{E}^{\text{total}} \quad (6.15a)$$

$$\text{s.t. : } R_k(\mathbf{p}_k, \Phi, \mathbf{x}_k) \geq s_k B_k, \quad \forall k \in \mathcal{K}, \quad (6.15b)$$

$$\sum_{l=1}^L \sum_{n=1}^N x_k[l, n] p_k[l, n] \leq s_k p_{k,\text{max}}, \quad \forall k \in \mathcal{K}, \quad (6.15c)$$

$$|\Phi_{m,m}| = 1, \quad \forall m \in \mathcal{M}, \quad (6.15d)$$

$$0 \leq \alpha_m < 2\pi, \quad \forall m \in \mathcal{M}, \quad (6.15e)$$

$$\sum_{k=1}^K x_k[l, n] \leq 1, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \quad (6.15f)$$

$$s_k \in \{0, 1\}, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \quad (6.15g)$$

$$x_k[l, n] \in \{0, 1\}, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \forall n \in \mathcal{N}. \quad (6.15h)$$

In this redefined context, the transformed constraint (6.15b) ensures that the offloading traffic for each user k satisfies a minimum data transmission requirement of B_k bits. This

adjustment is critical for maintaining the integrity of QoS standards across the network, particularly in scenarios where offloading is deemed beneficial or necessary.

Secondly, addressing the complexities associated with binary variable interactions, notably the multiplication of two binary variables in \mathcal{P}_2 , necessitates a strategic simplification. To this end, we impose a maximum value constraint on $x_k[l, n]$ by setting $x_k[l, n] \leq s_k$. This maneuver effectively reduces the complexity of terms involving the product of s_k and $x_k[l, n]$, thereby simplifying the complexity.

With these modifications, \mathcal{P}_2 emerges as a refined optimization problem that retains the core objectives of minimizing total power consumption and adhering to QoS prerequisites, yet is framed in a manner that enhances tractability and solution feasibility. This restructured approach paves the way for deploying sophisticated analytical techniques, such as SCA, to navigate the challenges inherent in solving non-convex MINLPs, thereby inching closer to identifying a viable and efficient solution. Thus, the optimization problem stated in equation \mathcal{P}_2 can be reformulated as follows:

$$\mathcal{P}_3 : \min_{\mathbf{p}, \Phi, \mathbf{s}, \mathbf{x}} \bar{\mathcal{E}}^{\text{total}} = \sum_{k \in \mathcal{K}} (\bar{\mathcal{E}}_k^{\text{offl}} + \mathcal{E}_k^{\text{loc}}) \quad (6.16a)$$

$$\text{s.t.} : \sum_{l=1}^L \sum_{n=1}^N p_k[l, n] \leq p_{k,\text{max}}, \quad \forall k \in \mathcal{K}, \quad (6.16b)$$

$$\sum_{l=1}^L p_k[l, n] \leq s_k p_{k,\text{max}}, \quad \forall k \in \mathcal{K}, \forall n \in \mathcal{N}, \quad (6.16c)$$

$$x_k[l, n] \leq s_k, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \forall n \in \mathcal{N}, \quad (6.16d)$$

$$(6.13c) - (6.13h), (6.15b),$$

where

$$\bar{\mathcal{E}}_k^{\text{offl}} = \sum_{l \in \mathcal{L}} \sum_{n \in \mathcal{N}} x_k[l, n] p_k[l, n] + s_k p_{cir}, \quad \forall k \in \mathcal{K}.$$

To effectively address the inherent non-convexity arising from the multiplication of $x_k[l, n]$ and $p_k[l, n]$, a key step in our optimization strategy involves introducing a new variable, $\tilde{p}_k[l, n] = x_k[l, n] p_k[l, n]$. This innovative approach is beneficial in handling of the product term, thereby simplifying the complexity of the optimization problem. Utilizing the well-established big-M method [129], which is a common technique for linearizing product terms involving binary variables in optimization problems, we incorporate additional constraints into our revised problem \mathcal{P}_3 . These new constraints effectively linearize the non-convex term, thereby rendering the optimization problem more tractable (getting even closer to the solution but not quite there). The modified optimization problem, inclusive of these new constraints, is presented as follows:

$$\mathcal{P}_4 : \min_{\mathbf{p}, \tilde{\mathbf{p}}, \Phi, \mathbf{s}, \mathbf{x}} \tilde{\mathcal{E}}^{\text{total}} = \sum_{k \in \mathcal{K}} (\tilde{\mathcal{E}}_k^{\text{offl}} + \mathcal{E}_k^{\text{loc}}) \quad (6.17a)$$

$$\text{s.t. : } \tilde{R}_k(\mathbf{p}_k, \tilde{\mathbf{p}}_k, \Phi, \mathbf{x}_k) \geq s_k B_k, \quad \forall k \in \mathcal{K}, \quad (6.17b)$$

$$\sum_{l=1}^L \sum_{n=1}^N \tilde{p}_k[l, n] \leq s_k p_{k, \text{max}}, \quad \forall k \in \mathcal{K}, \quad (6.17c)$$

$$\tilde{p}_k[l, n] \leq x_k[l, n] p_{k, \text{max}}, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \forall n \in \mathcal{N}, \quad (6.17d)$$

$$\tilde{p}_k[l, n] \leq p_k[l, n], \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \forall n \in \mathcal{N}, \quad (6.17e)$$

$$\tilde{p}_k[l, n] \geq p_k[l, n] - (1 - x_k[l, n]) p_{k, \text{max}}, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \forall n \in \mathcal{N}, \quad (6.17f)$$

$$\tilde{p}_k[l, n] \geq 0, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \forall n \in \mathcal{N}, \quad (6.17g)$$

$$(6.13d) - (6.13h), (6.16b) - (6.16d),$$

where

$$\tilde{\mathcal{E}}_k^{\text{offl}} = \sum_{l \in \mathcal{L}} \sum_{n \in \mathcal{N}} \tilde{p}_k[l, n] + s_k p_{k, \text{cir}}, \quad \forall k \in \mathcal{K}. \quad (6.18)$$

Moreover, we have:

$$\tilde{R}_k(\mathbf{p}_k, \tilde{\mathbf{p}}_k, \Phi, \mathbf{x}_k) = \tilde{F}_k(\mathbf{p}_k, \tilde{\mathbf{p}}_k, \Phi, \mathbf{x}_k) - \tilde{G}_k(\mathbf{p}_k, \tilde{\mathbf{p}}_k, \Phi, \mathbf{x}_k), \quad \forall k \in \mathcal{K}, \quad (6.19)$$

where

$$\tilde{\gamma}_k[l, n] = \frac{\|\mathbf{w}_k^H[n] \bar{\mathbf{h}}_k[n]\|^2 \tilde{p}_k[l, n]}{\sum_{j \neq k}^K \|\mathbf{w}_k^H[n] \bar{\mathbf{h}}_j[n]\|^2 \tilde{p}_j[l, n] + \tilde{\sigma}^2}, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \forall n \in \mathcal{N}, \quad (6.20)$$

$$\tilde{F}_k(\mathbf{p}_k, \tilde{\mathbf{p}}_k, \Phi, \mathbf{x}_k) = \sum_{l=1}^L \sum_{n=1}^N \log_2 (1 + \tilde{\gamma}_k[l, n]), \quad \forall k \in \mathcal{K}, \quad (6.21)$$

$$\tilde{G}_k(\mathbf{p}_k, \tilde{\mathbf{p}}_k, \Phi, \mathbf{x}_k) = a Q^{-1}(\epsilon_k) \sqrt{\sum_{l=1}^L \sum_{n=1}^N \left(1 - (1 + \tilde{\gamma}_k[l, n])^{-2}\right)}, \quad \forall k \in \mathcal{K}. \quad (6.22)$$

Furthermore, $\tilde{\mathbf{p}}_k$ represents the collection of optimization variables $\tilde{p}_k[l, n], \forall l, n$.

Next, we deal with the relaxation of integer variables (the l in MINLP) by considering their continuous counterparts. In our system model, this relaxation involves transforming the binary nature of certain variables — binary subcarrier allocation and offloading decision variables — into continuous variables that lie within the range from zero to one. Such a transformation makes employing conventional optimization techniques that are better suited for continuous rather than discrete variable spaces possible.

To outline the boundaries within which our optimization should operate, and to ensure the integrity and feasibility of solutions, we introduce additional constraints into the optimization problem. These constraints are crafted to define the feasible regions for the newly continuous variables, ensuring that the solutions remain realistic and applicable to the practical

scenarios envisaged in our IoT network optimization (getting very close to the solution). The imposition of these constraints is articulated as follows:

$$\mathcal{P}_5 : \min_{\mathbf{p}, \tilde{\mathbf{p}}, \Phi, \mathbf{s}, \mathbf{x}} \tilde{\mathcal{E}}^{\text{total}} + \lambda_1 \left(\sum_{k=1}^K s_k - s_k^2 \right) + \lambda_2 \left(\sum_{k=1}^K \sum_{l=1}^L \sum_{n=1}^N (x_k[l, n] - (x_k[l, n])^2) \right) \quad (6.23a)$$

$$\text{s.t. : } 0 \leq s_k \leq 1, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \quad (6.23b)$$

$$0 \leq x_k[l, n] \leq 1, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \forall n \in \mathcal{N}, \quad (6.23c)$$

$$\sum_{k=1}^K (s_k - (s_k)^2) \leq 0, \quad (6.23d)$$

$$\sum_{k=1}^K \sum_{l=1}^L \sum_{n=1}^N (x_k[l, n] - (x_k[l, n])^2) \leq 0, \quad (6.23e)$$

$$(6.13d) - (6.13f), (6.16b) - (6.16d), (6.17b) - (6.17g).$$

where λ_1 and λ_2 are penalty factors that need to be greater than one.

To facilitate the optimization of phase shifts in our IRS-assisted network, we convert the SINR function, (6.20), into a formulation that is more attractive to mathematical manipulation. The application of Semi-Definite Programming (SDP) emerges as a strategic approach to achieve this objective. SDP is a powerful optimization framework that allows for the optimization of a linear objective over the cone of positive semidefinite matrices, making it particularly suitable for handling problems involving quadratic forms and linear matrix inequalities. By leveraging SDP, we can transform the SINR function into a form that not only retains the essence of the original problem but also simplifies the process of optimizing the phase shifts of the IRS elements. The transformed SINR function under the SDP framework can be represented as follows:

$$\|\mathbf{w}_k^H[n]\bar{\mathbf{h}}_k[n]\|^2 = \text{Tr}(\mathbf{U}_k[n]\boldsymbol{\Upsilon}\mathbf{U}_k^H[n]\mathbf{W}_k[n]), \quad \forall k \in \mathcal{K}, \forall n \in \mathcal{N}, \quad (6.24)$$

where

$$\mathbf{U}_k[n] = \left[\left(\mathbf{F}^H[n] \text{diag}(\mathbf{h}_k^H[n]) \right)^T \mathbf{g}_k^*[n] \right]^T, \quad \forall k \in \mathcal{K}, \forall n \in \mathcal{N}, \quad (6.25)$$

$$\mathbf{W}_k[n] = \mathbf{w}_k[n]\mathbf{w}_k^H[n], \quad \forall k \in \mathcal{K}, \forall n \in \mathcal{N}, \quad (6.26)$$

$$\boldsymbol{\Upsilon} = \boldsymbol{\varpi}\boldsymbol{\varpi}^H \in \mathbb{C}^{(M+1) \times (M+1)}, \quad (6.27)$$

$$\boldsymbol{\varpi} = [\boldsymbol{\varrho}^T \boldsymbol{\varkappa}]^T \in \mathbb{C}^{(M+1) \times 1}. \quad (6.28)$$

In (6.28), $\boldsymbol{\varkappa} \in \mathbb{C}$ is a dummy variable with $|\boldsymbol{\varkappa}|^2 = 1$, and $\boldsymbol{\varrho} = [e^{j\alpha_1}, e^{j\alpha_2}, \dots, e^{j\alpha_M}]^H \in \mathbb{C}^{M \times 1}$. To simplify the complexity and facilitate the solution of \mathcal{P}_5 , we introduce a set of auxiliary variables $\chi_k[l, n], \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \forall n \in \mathcal{N}$. These auxiliary variables are designed to set a lower bound on the SINR equation (6.20). This approach not only simplifies the mathematical

representation of the SINR constraint but also ensures that it becomes more amenable to the optimization techniques we intend to apply. Therefore, we have:

$$0 \leq \chi_k[l, n] \leq \tilde{\gamma}_k[l, n] \triangleq \frac{\mathcal{C}_k[l, n]}{\mathcal{D}_k[l, n]}, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \forall n \in \mathcal{N}, \quad (6.29)$$

where

$$\mathcal{C}_k[l, n] = \text{Tr}(\mathbf{Z}_k[n] \mathbf{W}_k[n]) \tilde{p}_k[l, n], \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \forall n \in \mathcal{N}, \quad (6.30)$$

$$\mathcal{D}_k[l, n] = \sum_{j \neq k}^K \text{Tr}(\mathbf{Z}_j[n] \mathbf{W}_k[n]) \tilde{p}_j[l, n] + \tilde{\sigma}^2, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \forall n \in \mathcal{N}, \quad (6.31)$$

with

$$\mathbf{Z}_q[n] = \mathbf{U}_q[n] \boldsymbol{\Upsilon} \mathbf{U}_q^H[n], \quad \forall q = \{k, j\}, \forall k \in \mathcal{K}. \quad (6.32)$$

Consequently, the achievable rate in (6.17b) can be restated as:

$$\tilde{R}_k(\chi_k) = \tilde{F}_k(\chi_k) - \tilde{G}_k(\chi_k), \quad \forall k \in \mathcal{K}, \quad (6.33)$$

where χ_k is the collection of optimization variables $\chi_k[l, n], \forall l, n$.

To further refine the approach towards solving \mathcal{P}_5 , we introduce an additional layer of optimization variables, denoted by $\mathcal{I}_k[l, n], \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \forall n \in \mathcal{N}$. These slack optimization variables are utilized to set an upper bound on the denominator of the SINR approximation as expressed in equation (6.29). This deliberate move simplifies the handling of the SINR constraint by decoupling the interactions within its denominator and numerator, thereby rendering the optimization problem more tractable². The incorporation of $\mathcal{I}_k[l, n]$ into our optimization framework leads to the following reformulation:

$$\chi_k[l, n] \mathcal{I}_k[l, n] \leq \mathcal{C}_k[l, n], \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \forall n \in \mathcal{N}, \quad (6.34)$$

$$\mathcal{I}_k[l, n] \geq \mathcal{D}_k[l, n], \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \forall n \in \mathcal{N}, \quad (6.35)$$

where $\mathcal{I}_k[l, n]$ can be thought of the k 's user interference on time slot l and subcarrier n . By referring to the objective function of \mathcal{P}_5 as $\tilde{\mathcal{E}}^{\text{total}}$, the revised optimization problem (which is a touch away from the solution) is formulated as follows:

²Transforming fractional problems, like those involving SINR, into non-fractional forms simplifies solving them by enabling the use of more straightforward optimization techniques. This approach improves solution efficiency and accuracy, making complex problems more tractable. By introducing slack variables and employing techniques such as the big-M method, we essentially decouple the numerator and denominator of this ratio, thereby simplifying the SINR constraint into a more tractable form. This approach aligns with the goal of making the problem more amenable to optimization algorithms, enabling us to apply advanced mathematical tools and techniques.

$$\mathcal{P}_6 : \min_{\mathbf{p}, \tilde{\mathbf{p}}, \Phi, \mathbf{s}, \mathbf{x}, \boldsymbol{\chi}, \boldsymbol{\Upsilon}} \tilde{\mathcal{E}}^{\text{total}} \quad (6.36a)$$

$$\text{s. t. : } \tilde{F}_k(\boldsymbol{\chi}_k) - \tilde{G}_k(\boldsymbol{\chi}_k) \geq s_k B_k, \quad \forall k \in \mathcal{K}, \quad (6.36b)$$

$$\text{diag}(\boldsymbol{\Upsilon}) = \mathbf{1}_{M+1}, \quad (6.36c)$$

$$\boldsymbol{\Upsilon} \succeq 0, \quad (6.36d)$$

$$\text{rank}(\boldsymbol{\Upsilon}) \leq 1, \quad (6.36e)$$

$$(6.13d) - (6.13f), (6.16b) - (6.16d), (6.17b) - (6.17g),$$

$$(6.23b) - (6.23e), (6.34), (6.35).$$

The optimization problem \mathcal{P}_6 remains non-convex due to the non-convex nature of (6.34), (6.35), and the rank constraint (6.36e). This non-convexity presents significant challenges in the pursuit of an efficient solution. Notably, constraints (6.34) and (6.35) are characteristic of Difference of Convex (DC) problems, a category well-documented in literature [3, 8, 129, 9]. This identification is crucial as it opens avenues for applying DC programming strategies to navigate the problem's complexity.

Moreover, the bilinear term $\chi_k[l, n]\mathcal{I}_k[l, n]$ on the left-hand side of (6.34) further complicates the optimization by introducing an additional layer of non-convexity. Despite this complexity, there's a silver lining: this product term can be reformulated as the difference of two convex functions. Such a reformulation effectively translates both (6.34) and (6.35) into a DC problem framework. This transformation not only demystifies the path to addressing the non-convexity but also significantly enhances the feasibility of devising a potent and efficient resource allocation algorithm. The DC representation thus serves as a linchpin in our strategy, enabling a more nuanced and tractable approach to solving the optimization problem presented by \mathcal{P}_6 . This, the DC form of (6.34) and (6.35) can be expressed as follows:

$$\begin{aligned} \varsigma_1(\chi_k[l, n], \mathcal{I}_k[l, n]) - \varsigma_2(\chi_k[l, n], \mathcal{I}_k[l, n]) \leq \\ (\varsigma_3(\boldsymbol{\Upsilon}, \tilde{p}_k[l, n]) - \varsigma_4(\boldsymbol{\Upsilon}, \tilde{p}_k[l, n])), \\ \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \forall n \in \mathcal{N}, \quad (6.37) \end{aligned}$$

$$\sum_{j \neq k}^K (\varsigma_5(\boldsymbol{\Upsilon}, \tilde{p}_j[l, n]) - \varsigma_6(\boldsymbol{\Upsilon}, \tilde{p}_j[l, n])) + \tilde{\sigma}^2 \leq I_k[l, n], \\ \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \forall n \in \mathcal{N}, \quad (6.38)$$

where

$$\varsigma_1(\chi_k[l, n], \mathcal{I}_k[l, n]) = 0.5(\chi_k[l, n] + \mathcal{I}_k[l, n])^2, \quad (6.39)$$

$\forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \forall n \in \mathcal{N}$,

$$\varsigma_2(\chi_k[l, n], \mathcal{I}_k[l, n]) = 0.5(\chi_k[l, n])^2 + 0.5(\mathcal{I}_k[l, n])^2, \quad (6.40)$$

$\forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \forall n \in \mathcal{N}$,

$$\varsigma_{\zeta}(\mathbf{\Upsilon}, \tilde{p}_{\vartheta}[l, n]) = 0.5(\tilde{p}_{\vartheta}[l, n] + \text{Tr}(\mathbf{Z}_j[n]\mathbf{W}_k[n]))^2, \quad (6.41)$$

$\forall \{\zeta, \vartheta\} = \{\{3, k\}, \{5, j\}\}, \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \forall n \in \mathcal{N}$,

$$\varsigma_{\zeta}(\mathbf{\Upsilon}, \tilde{p}_{\vartheta}[l, n]) = 0.5(\tilde{p}_{\vartheta}[l, n])^2 + 0.5(\text{Tr}(\mathbf{Z}_j[n]\mathbf{W}_k[n]))^2, \quad (6.42)$$

$\forall \{\zeta, \vartheta\} = \{\{4, k\}, \{6, j\}\}, \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \forall n \in \mathcal{N}$.

Both sides of (6.37) and the left-hand side of (6.38) are not convex. To address the non-convexity of the left-hand side of (6.37), we apply the SCA technique, which involves using a first-order Taylor expansion to obtain a convex approximation of the non-convex terms as follows:

$$\begin{aligned} \bar{\varsigma}_2(\chi_k[l, n], \chi_k^{(i)}[l, n], \mathcal{I}_k[l, n], \mathcal{I}_k^{(i)}[l, n]) &= 0.5(\chi_k^{(i)}[l, n])^2 \\ &+ \chi_k^{(i)}[l, n](\chi_k[l, n] - \chi_k^{(i)}[l, n]) + 0.5(\mathcal{I}_k^{(i)}[l, n])^2 \\ &+ \mathcal{I}_k^{(i)}[l, n](\mathcal{I}_k[l, n] - \mathcal{I}_k^{(i)}[l, n]), \\ &\forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \forall n \in \mathcal{N}. \end{aligned} \quad (6.43)$$

Similarly, to address the non-convexity of the right-hand side of (6.37) and the left-hand side of (6.38), we also employ the SCA technique. Thus, we can approximate these non-convex terms as follows:

$$\begin{aligned} \tilde{\varsigma}_{\zeta}(\mathbf{\Upsilon}, \tilde{p}_{\vartheta}[l, n], \mathbf{\Upsilon}^{(i)}, \tilde{p}_{\vartheta}^{(i)}[l, n]) &= \varsigma_{\zeta}(\mathbf{\Upsilon}^{(i)}, \tilde{p}_{\vartheta}^{(i)}[l, n]) \\ &+ \text{Tr}(\nabla_{\mathbf{\Upsilon}}(\varsigma_{\zeta}(\mathbf{\Upsilon}^{(i)}, \tilde{p}_{\vartheta}^{(i)}[l, n]))^H(\mathbf{\Upsilon} - \mathbf{\Upsilon}^{(i)})) \\ &+ \text{Tr}(\nabla_{\tilde{p}_{\vartheta}}(\varsigma_{\zeta}(\mathbf{\Upsilon}^{(i)}, \tilde{p}_{\vartheta}^{(i)}[l, n]))^H(\tilde{p}_{\vartheta}[l, n] - \tilde{p}_{\vartheta}^{(i)}[l, n])), \\ &\forall \{\zeta, \vartheta\} = \{\{4, k\}, \{6, j\}\}, \forall l \in \mathcal{L}, \forall n \in \mathcal{N}, \end{aligned} \quad (6.44)$$

where $\nabla_{\mathbf{\Upsilon}}(\varsigma_{\zeta}(\mathbf{\Upsilon}, \tilde{p}_{\vartheta}[l, n]))$ and $\nabla_{\tilde{p}_{\vartheta}}(\varsigma_{\zeta}(\mathbf{\Upsilon}, \tilde{p}_{\vartheta}[l, n]))$ are the gradients of $\varsigma_{\zeta}(\mathbf{\Upsilon}, \tilde{p}_{\vartheta}[l, n])$, (6.42), with respect to $\mathbf{\Upsilon}$ and \tilde{p}_{ϑ} , respectively. Therefore, (6.37) and (6.38) can be approximated as follows:

$$\begin{aligned} \varsigma_1(\chi_k[l, n], \mathcal{I}_k[l, n]) - \bar{\varsigma}_2(\chi_k[l, n], \chi_k^{(i)}[l, n], \mathcal{I}_k[l, n], \mathcal{I}_k^{(i)}[l, n]) \\ \leq (\varsigma_3(\mathbf{\Upsilon}, \tilde{p}_k[l, n]) - \tilde{\varsigma}_4(\mathbf{\Upsilon}, \tilde{p}_k[l, n], \mathbf{\Upsilon}^{(i)}, \tilde{p}_k^{(i)}[l, n])), \end{aligned} \quad (6.45)$$

$$\begin{aligned} \sum_{j \neq k}^K (\varsigma_5(\mathbf{\Upsilon}, \tilde{p}_j[l, n]) - \tilde{\varsigma}_6(\mathbf{\Upsilon}, \tilde{p}_j[l, n], \mathbf{\Upsilon}^{(i)}, \tilde{p}_j^{(i)}[l, n])) + \tilde{\sigma}^2 \\ \leq I_k[l, n], \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \forall n \in \mathcal{N}. \end{aligned} \quad (6.46)$$

Unfortunately, \mathcal{P}_6 is still not convex. However, the convexity of \mathcal{P}_6 hinges on the rank of $\mathbf{\Upsilon}$. Typically, \mathcal{P}_6 yields solutions with a rank higher than one. To overcome this last challenge, we reformulate constraint (6.36e) utilizing the DC method, resulting in the following expression:

$$\|\mathbf{\Upsilon}\|^* - \|\mathbf{\Upsilon}\|^2 \leq 0. \quad (6.47)$$

Algorithm 7 Proposed Iterative SCA Algorithm for Offloading and Resource Allocation in an Uplink IRS-assisted MEC Network

Input: Set iteration index $i = 1$, and maximum number of iteration \mathcal{T}_{\max} , randomly initialize $\mathbf{p}^0, \tilde{\mathbf{p}}^0, \Phi^0, \mathbf{s}^0, \mathbf{x}^0, \boldsymbol{\chi}^0, \boldsymbol{\Upsilon}^0$, and penalty factors $[\lambda_1, \lambda_2, \delta]^T \succ \mathbf{1}_3$

- 1: **repeat**
- 2: Calculate (6.43) and (6.44)
- 3: Solve \mathcal{P}_7 for given $\mathbf{p}^{(i)}, \tilde{\mathbf{p}}^{(i)}, \Phi^{(i)}, \mathbf{s}^{(i)}, \mathbf{x}^{(i)}, \boldsymbol{\chi}^{(i)}$, and $\boldsymbol{\Upsilon}^{(i)}$, and retain the intermediate solution
- 4: Set $i = i + 1$ and $\mathbf{p}^{(i)} = \mathbf{p}^*$, $\tilde{\mathbf{p}}^{(i)} = \tilde{\mathbf{p}}^*$, $\Phi^{(i)} = \Phi^*$, $\mathbf{s}^{(i)} = \mathbf{s}^*$, $\mathbf{x}^{(i)} = \mathbf{x}^*$, $\boldsymbol{\chi}^{(i)} = \boldsymbol{\chi}^*$, and $\boldsymbol{\Upsilon}^{(i)} = \boldsymbol{\Upsilon}^*$
- 5: **until** $i = \mathcal{T}_{\max}$
- 6: **return** $\mathbf{p}^*, \tilde{\mathbf{p}}^*, \Phi^*, \mathbf{s}^*, \mathbf{x}^*, \boldsymbol{\chi}^*, \boldsymbol{\Upsilon}^*$

Note that $\|\boldsymbol{\Upsilon}\|^* = \sum_i \tau_i \geq \|\boldsymbol{\Upsilon}\|^2 = \max_i \{\tau_i\}$ holds for any given $\boldsymbol{\Upsilon}$, where τ_i is the i -th singular value of $\boldsymbol{\Upsilon}$. The equality holds if and only if $\boldsymbol{\Upsilon}$ achieves rank one i.e., $\text{rank}(\boldsymbol{\Upsilon}) = 1$ [3]. Now, we take the first-order Taylor approximation of $\|\boldsymbol{\Upsilon}\|^2$ as:

$$\|\boldsymbol{\Upsilon}\|^2 \geq \overbrace{\|\boldsymbol{\Upsilon}^{(t)}\|^2 + \text{Tr} \left(\lambda_{\max} \left(\boldsymbol{\Upsilon}^{(t)} \right) \lambda_{\max}^H \left(\boldsymbol{\Upsilon}^{(t)} \right) (\boldsymbol{\Upsilon} - \boldsymbol{\Upsilon}^{(t)}) \right)}^{=\kappa(\boldsymbol{\Upsilon})}. \quad (6.48)$$

By utilizing (6.48), we can obtain a convex approximation for (6.47), expressed as:

$$\tilde{\kappa}^t(\boldsymbol{\Upsilon}) \triangleq \|\boldsymbol{\Upsilon}\|_* - \kappa(\boldsymbol{\Upsilon}) \leq 0. \quad (6.49)$$

Finally, the optimization problem is formulated by adding $\tilde{\kappa}^t(\boldsymbol{\Upsilon})$ to the objective function of \mathcal{P}_6 with a penalty factor $\delta \gg 1$ to penalize non-rank-one matrices. The modified optimization problem with convex objective function and constraints (we arrived to a solution) can be written as follows:

$$\mathcal{P}_7 : \min_{\mathbf{p}, \tilde{\mathbf{p}}, \Phi, \mathbf{s}, \mathbf{x}, \boldsymbol{\chi}, \boldsymbol{\Upsilon}} \tilde{\mathcal{E}}^{\text{total}} - \delta(\tilde{\kappa}^t(\boldsymbol{\Upsilon})) \quad (6.50a)$$

(6.13d) – (6.13f), (6.16b) – (6.16d), (6.17b) – (6.17g),
(6.23b) – (6.23e), (6.36b) – (6.36d), (6.45), (6.46).

The optimization problem \mathcal{P}_7 can be effectively solved by utilizing well-established convex optimization packages like CVX [9, 8, 129, 3]. Finally, we outline our proposed algorithm in **Algorithm 7**.

6.5 Performance Evaluation

In this section, we present the simulation results that verify the performance of our proposed IRS-assisted OFDMA uplink URLLC MEC system, adhering to the simulation parameters

provided in *Table 6.1*, unless specified otherwise. The simulation environment is designed, positioning the network's center at the coordinate origin $(0, 0)$ meters, with the AP strategically placed at $(0, -100)$ meters and the IRS at $(50, 0)$ meters. Furthermore, the network hosts five MTC-enabled IoT users ($K = 5$), whose locations are randomly distributed within a circle of 4 meters radius centered at $(25, 0)$ meters.

To accurately model the signal propagation within this environment, we adopt a path-loss model expressed as $L(d) = a_0(d/d_0)^{-\xi}$, where $a_0 = -0.001$ represents the signal attenuation at a reference distance $d_0 = 1$ meter, d denotes the link distance, and ξ is the path-loss exponent. Specific to our setup, the path-loss exponents for the AP-IRS, IRS-user, and AP-user links are respectively configured to 2.2, 2.2, and 3.4. This setup assumes Rician fading across all communication links, characterized by a Rician factor of 3 dB, to closely mimic realistic signal propagation scenarios and evaluate the performance of our system under practical conditions.

In our simulation framework, we address the small-scale fading by adopting distinct models for different channel links. Specifically, we model the fading between the AP and the users using Rayleigh fading, which is typical for environments where the LoS component is absent or negligible. Conversely, for the channels connecting the AP to the IRS and the IRS to the users, we employ a Rician fading model. The Rician model is characterized by a Rician factor of 10, indicating a significant LoS component in these links, which is common in scenarios where the IRS is strategically placed to enhance the communication link's quality.

The choice of fading models is crucial for accurately representing the physical environment's impact on the transmitted signals. Rayleigh fading is ideal for urban environments where obstacles frequently obstruct the direct path, while Rician fading models scenarios with a clear dominant path but accompanied by scattered multipath components.

The comprehensive simulation results we present are the culmination of averaging over multiple realizations of both path loss and multi-path fading effects. This averaging process ensures that the outcomes reflect a robust understanding of the system's performance across various propagation conditions, thereby providing insights that are both reliable and indicative of real-world operational scenarios. The consideration of these fading models enhances the fidelity of our simulations, ensuring they capture the essential characteristics of wireless transmission in diverse environments.

6.5.1 Performance Bound and Benchmark Schemes

To rigorously evaluate the effectiveness of the resource allocation algorithm we propose, we undertake a comparative analysis against a suite of benchmark schemes, each designed to highlight different aspects of system performance under varied conditions:

- **Lower bound:** To obtain a lower bound on the system performance, Shannon's capacity formula is adopted in problem \mathcal{P}_1 , i.e., $V_k[l, n], \forall k$, (6.8), is set to zero. The resulting optimization problem is solved using a modified version of the proposed algorithm.
- **Method A:** This is the proposed **Algorithm 7**.

Table 6.1: Simulation Parameters for Offloading and Resource Allocation in an Uplink IRS-assisted MEC Network

Parameter	Value
Total number of reflecting elements, M	50
Total number of uplink time slots, L	4
Total number of subcarriers, N	32
Bandwidth of each sub-carrier	30 kHz
Noise power density	-174 dBm/Hz
Maximum transmit power of each user, $p_{k,\max}$	23 dBm
Circuit power consumption of user k , p_{cir}	50 mW
Packet decoding error probability, ϵ_k	10^{-6}
Number of bits per packet (bitstream size), B_k	160 bits

- **Method B:** In this scheme, we adopt random phase shifts for the IRS elements and optimize the users' power allocation and offloading decisions.
- **Method C:** In this approach, we maintain a fixed sub-carrier allocation for offloading while optimizing other variables using SCA.
- **Method D:** We remove the IRS from the system in this scheme. We consider the uplink power allocation and passive beamforming based solely on the direct link between the AP and the user.

6.5.2 Simulation Results

Fig. 6.2 illustrates the relationship between system power consumption and packet error probability, highlighting the impact of acceptable error rate and joint resource optimization on power consumption. Notably, power consumption demonstrates a monotonically decreasing behavior as packet error probability increases. This trend is primarily due to the characteristics of the complementary error function employed in the normal approximation for the data rate function, i.e., (6.6), which naturally declines as packet error probability escalates. Therefore, embracing a higher packet error probability threshold can significantly diminish the required transmit power to adhere to the latency demands of MTC-enabled IoT applications. Consequently, this means as the packet error probability increases, the influence of the dissipation part in the normal approximation fades away.

This depiction also reveals that, in scenarios adhering to the lower bound on performance (lower bound), power consumption remains unaffected by packet error probability variations. This is because of the foundational assumption of zero packet error probability inherent in Shannon's capacity formula, where channel dispersion $V_k[l, n]$, $\forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \forall n \in \mathcal{N}$, (6.8), is nullified. The discernible gap between the lower bound and the outcomes of our proposed algorithm (Method A) represents the necessary compromise to satisfy the rigorous delay and reliability requisites of ultra-reliable low-latency communication with short packet transmissions.

Incorporating IRSs into the system architecture not only paves the way for remarkable power efficiency enhancements but also significantly boosts energy conservation. Nonetheless, it

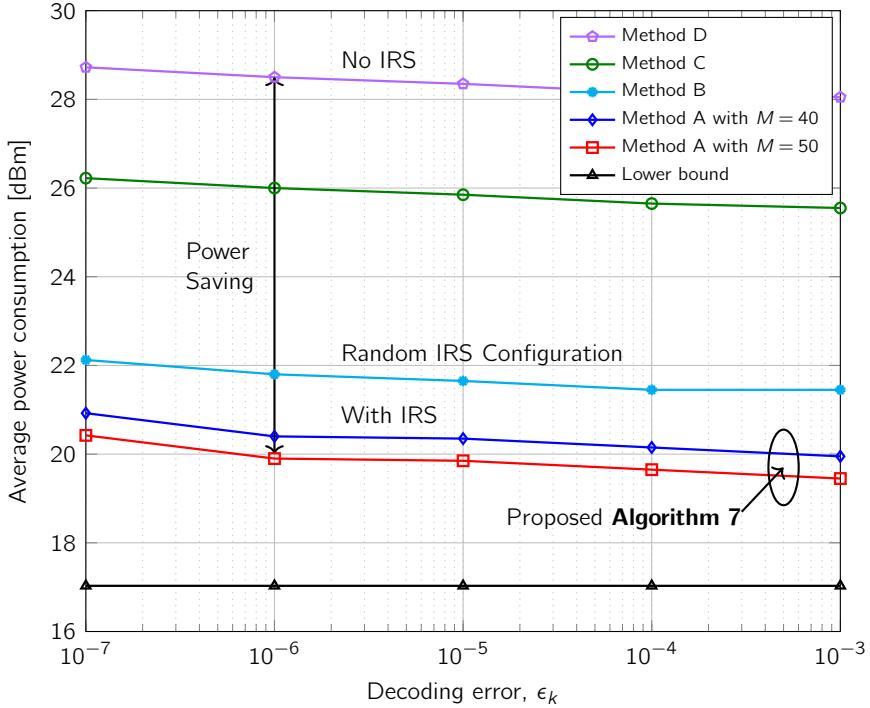


Figure 6.2: Average power consumption [dBm] vs. packet decoding error probability in an IRS-aided network.

is important to weigh the computational demands and the initial outlay required for IRS implementation. Increasing the number of reflecting elements (M) enhances the passive beamforming gain, thereby reducing the transmit power of IoT devices and facilitating efficient offloading.

The performance of the scheme utilizing random IRS beamforming (Method B) is inferior to that of a system utilizing an optimal beamforming vector (Method A), but it still surpasses the performance benchmarks set by strategies reliant on fixed sub-carrier allocation (Method C). It is important to note that deploying IRSs plays a crucial role in maximizing the capabilities of MEC servers. By dynamically modifying the wireless propagation environment in real-time, IRSs help ensure that users are not forced to allocate more power due to poor channel conditions. This allows for efficient offloading of user tasks to edge servers rather than having to compute the tasks locally. Furthermore, deploying IRSs helps guarantee that transmissions are completed within the desired delay, meeting the latency requirements of the system. This comprehensive analysis accentuates the pivotal role of IRSs in enhancing system performance, facilitating effective task offloading in MEC landscapes, and ensuring that the network's energy consumption is optimized.

Fig. 6.3 presents an in-depth analysis of how varying task sizes (commonly known as bit-stream sizes), influence the overall power consumption within the system. It is observed that with the increase in the task sizes, there is a corresponding increment in power consumption across all schemes. This trend is primarily attributed to the increased demand

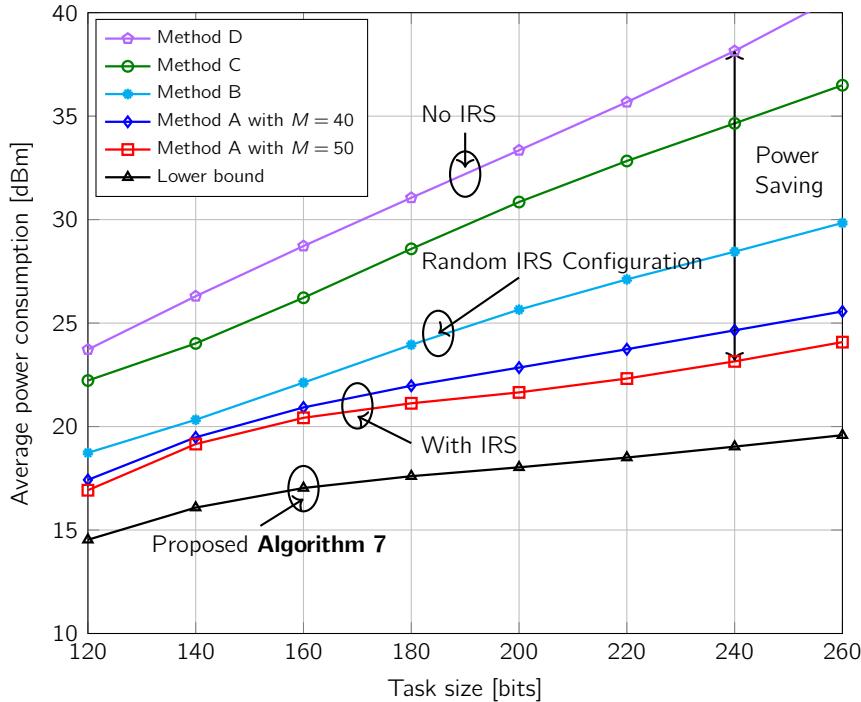


Figure 6.3: Average power consumption [dBm] vs. the task size [bits] in the uplink of an IRS-assisted OFDMA URLLC MEC network.

for higher SINRs and the subsequent necessity for elevated transmit power levels to sustain reliable communication. Significantly, the incorporation of IRSs into the system architecture markedly enhances the SINR values by introducing additional LoS connections. This enhancement allows the system equipped with IRS technology (method A) to efficiently manage and support larger task sizes with reduced power consumption compared to configurations without IRS capabilities (method D).

The effectiveness of the proposed algorithm is further demonstrated through its superior performance relative to the alternative strategies that employ non-optimized subcarrier allocation and random IRS phase shifts (methods B and C). By strategically leveraging the SINR improvements offered by the IRS, the proposed algorithm substantially minimizes power consumption. This is achieved through the optimization of data offloading decisions, subchannel allocation strategies, and transmission power settings, demonstrating the algorithm's capability to adaptively balance the system's power requirements.

Furthermore, these results highlight the critical decision-making process involved in selecting between offloading computational tasks to the edge or processing them locally, especially in scenarios characterized by the transmission of larger data sets. The proposed algorithm's approach to resource allocation demonstrates its potential to significantly influence power efficiency, making it a vital tool for managing the energy demands of IRS-assisted OFDMA URLLC MEC systems. This detailed examination determines the transformative impact of IRS technology on enhancing network performance, particularly in facilitating energy-efficient

management of increasing computational demands, especially when processing huge data sets.

6.6 Conclusions

This chapter studied the process of designing a resource allocation algorithm for an uplink multiuser IRS-aided MEC system. Centered on addressing the strict requirements for end-to-end transmission delay and reliability, crucial for MTC-enabled IoT users, our exploration continued toward a comprehensive joint resource allocation and offloading decision scheme. This scheme, distinctively characterized by its focus on short packet transmission, leverages the IRS's capability to significantly enhance the communication channel, thereby ensuring improved levels of reliability through the provision of virtual LoS links. The IRS was deployed to enhance the communication channel and to increase reliability by providing virtual LoSs.

To navigate the complexities of this study, we formulated an optimization problem aimed at minimizing the average system power consumption, all while adhering to URLLC MTC-enabled user's QoS constraints. This formulation revealed the problem's inherent non-convex MINLP nature, underscoring the substantial challenges encountered in the quest for an optimal solution. In response to these challenges, we engineered an efficient, low-complexity algorithm. This algorithm, which integrates SCA techniques alongside an iterative rank minimization method, demonstrated remarkable proficiency in converging to a local optimum.

The simulation results presented serve not only as a testament to the proposed algorithm's effectiveness but also illuminate the significant, practical benefits of integrating IRS technology within MEC systems. Particularly noteworthy is the IRS's invaluable role in extending coverage and facilitating task offloading for multiple energy-constrained URLLC devices, thereby heralding a new era of enhanced network performance and efficiency.

As we move to the next chapter, we aim to investigate the application of IRS technology within the millimeter-wave (mmWave) spectrum. This forthcoming discussion promises to unravel the potential of the IRS to revolutionize communication in the high-frequency mmWave bands, offering insights into novel strategies for overcoming the challenges posed by this spectrum while maximizing the benefits of IRS-enhanced communication networks.

Active IRS for mmWave Wireless Networks

INTELLIGENT REFLECTING SURFACES (IRSs) have emerged as a revolutionary technology capable of substantially enhancing the efficiency and performance of various applications by intelligently manipulating signal propagation paths. By acting as low-cost passive devices, IRSs can adjust the scattering, refraction, and reflection characteristics of radio waves, thereby significantly reducing interference at one or more selected receivers. In this chapter, we explore the concept of an *active* IRS-assisted Multiple-Input Single-Output (MISO) system within the context of a millimeter-Wave (mmWave) wireless network. This setup allows for signals transmitted from the Access Point (AP) to be reflected by the IRS, facilitating their reception by the user. To fully leverage the advantages of IRS-assisted wireless networks, it is crucial to optimize not just the phase shift at the IRS but also to take advantage of amplitude variation (now the elements are ‘active’) through the use of cost-effective hardware.

Until now, in previous chapters, our discussion has centered on the lower frequency range, specifically Frequency Range 1 (FR1), which encompasses frequencies from 450 MegaHertz (MHz) to 6 GHz. This range, utilized by current cellular networks, supports traditional cellular bands and strikes a balance between coverage and capacity. We have established a fundamental comprehension of how IRSs operate within these spectrums. Shifting our focus to a higher frequency domain, the significance of Frequency Range 2 (FR2) becomes apparent in appreciating the utility of IRSs. FR2, which extends from 24 GHz to 52 GHz, delves into the mmWave spectrum. This range is distinguished by its potential to deliver substantially higher data rates and bandwidth, courtesy of larger available frequency blocks. However, this comes with the trade-offs of reduced transmission distances and an increased vulnerability to physical obstructions. In this chapter, we want to answer the question of whether Active IRS is useful for FR2.

Thus, in this chapter, we formulate the system sum rate maximization problem to optimize both the active and passive beamformer at the AP and the IRS, considering amplitude control at the IRS from the limited available power with low-cost hardware. To tackle this complex problem, we introduce two low-complexity algorithms. In particular, the first sub-problem utilizes the Weighted Minimum Mean Square Error (WMMSE) methodology to enhance beamforming at the AP, while the second is approached through Successive Convex

Approximation (SCA). Our numerical results show the advantages of active IRS configurations, demonstrating their superior performance in comparison to passive IRS setups.

This chapter is based on:

J.Jalali, A. Khalili, A. Rezaei and J. Famaey, "Is Active IRS Useful for mmWave Wireless Networks or Not?", in *2023 International Conference on Computing, Networking and Communications (ICNC)*, Honolulu, HI, USA, Feb. 2023, pp. 377–382. <https://doi.org/10.1109/ICNC57223.2023.10074428>

J.Jalali, M. Bustamante, F. Lemic, H. Tabassum, J. Struye, J. Famaey, and X. Costa Pérez, "Location Optimization and Resource Allocation of IRS in a Multi-User Indoor mmWave VR Network", in *2024 IEEE Wireless Communications and Networking Conference (WCNC)*, Dubai, United Arab Emirates, Apr. 2024, pp. 1–7. (Accepted)

7.1 Introduction to IRS-assisted mmWave Networks

To meet the growing demand for high-speed multimedia access, it's crucial to significantly enhance the capacity of existing wireless networks through the integration of diverse wireless technologies and network architectures. Recent investigations have highlighted the potential of ultra-dense networking and massive Multiple-Input Multiple-Output (m-MIMO) systems. Furthermore, the adoption of innovative techniques, such as exploiting the millimeter-wave (mmWave) spectrum, has been identified as a promising approach to satisfy the burgeoning capacity requirements [230]. Notably, the use of mmWave bands and the incorporation of additional antenna elements, coupled with the deployment of Radio Frequency (RF) chains operating at exceedingly high frequencies, pose challenges related to increased hardware costs and elevated energy consumption in practical mmWave systems. Consequently, there is a pressing need for novel strategies that address both spectral and energy efficiency to ensure the sustainable development and deployment of future wireless networks [18, 231].

In response to these challenges, intelligent reflecting surfaces (IRS) have emerged as a critical technological innovation, facilitating the creation of smarter radio environments. This cutting-edge technology employs an array of artificial reflecting elements, such as low-cost printed dipoles, to reflect incident RF waves in specific directions while minimizing power consumption. These reflecting elements are controlled by an intelligent mechanism, enabling the manipulation of signal propagation without introducing additional thermal noise. This attribute of IRS, leveraging passive reflection beamforming, ensures significantly lower power consumption compared to traditional Amplify-and-Forward (AF) relays, thereby offering an efficient and sustainable solution for the evolution of wireless networks [91, 232].

Recent research on IRS-aided communication systems has predominantly concentrated on refining IRS parameters to enhance the performance of outdoor communication networks [64, 60, 233]. Notably, [64] delves into an IRS-assisted single-cell multi-user Multiple-Input Single-Output (MISO) setup, focusing on optimizing the induced phases for both passive and active beamformers. This optimization aims to amplify the overall received signal power for users equipped with single antennas. The studies presented in [60, 233] extend this approach by striving to boost both energy and spectral efficiency within an IRS-aided multi-user MISO framework. They achieve this through the development of a transmit power allocation strategy at the Access Point (AP) and the adjustment of IRS phase-shifts, specifically tailored for scenarios where zero-forcing precoding is applied in the digital domain.

While a substantial body of literature has addressed IRS-based networks [136, 174], only a select few have explored the manipulation of reflection amplitudes within IRS systems [231, 234]. The investigations in [231, 234] are particularly groundbreaking, evaluating the impact of amplitude variation in networks characterized by imperfect Channel State Information (CSI). By employing advanced optimization techniques, such as penalized Dinkelbach and block successive upper-bound minimization algorithms, these studies succeed in fine-tuning reflection coefficients to maximize data transmission rates. This meticulous control over reflection amplitudes facilitates notable performance enhancements when compared to traditional full reflection/phase-shift control methods.

Answering whether to go with *passive* or *active* IRS deployment in the next generation of

wireless communication networks led to significant recent attraction in the field [235, 236, 237, 238, 239]. Research in [235] prioritizes the strategic placement of IRS elements to optimize data rates in a system comprising a single-user, Single-Input Single-Output (SISO) configuration alongside an active or passive IRS and a single-antenna AP. In the context of the upcoming sixth-Generation (6G) wireless communications, [236] advocates for active IRS as a pivotal 6G technology, capable of mitigating multiplicative fading. This study provides a comprehensive comparison between active and conventional passive IRS models to identify performance limitations. Additionally, [237] explores the optimization of user data rates in a multi-user MISO downlink network, incorporating different types of IRSs as integral components of the 6G ecosystem. A practical design for active IRS, aimed at minimizing outage probabilities within a SISO framework with full CSI knowledge, is examined in [238]. Lastly, [239] focuses on a resource allocation problem, aiming to reduce the transmit power requirements of an AP within an active IRS-enhanced communication network. Here, active IRSs are leveraged to amplify the reflected signal by harnessing an additional power source, demonstrating the potential of active IRS deployment in future wireless networks.

The advent of mmWave technology presents a compelling avenue to significantly enhance throughput in wireless networks, leveraging its capability to support higher frequency bands. Despite its promise, the mmWave technology is not without its challenges, primarily due to its short wavelengths, which lead to diminished signal propagation and increased absorption by physical obstructions. In this context, IRS emerges as a potential solution to mitigate some of the limitations associated with mmWave communication networks, potentially enhancing performance by intelligently managing signal propagation [84]. However, the integration of IRS with mmWave technology is not straightforward and can introduce complexities, such as resource inefficiencies [86]. While IRS technology can amplify signals for certain users, it may inadvertently escalate destructive interference for others, underlining the dualistic nature of its impact on network performance.

Addressing these challenges necessitates a nuanced approach, particularly focusing on the control of reflection amplitudes by IRS. Our objective centers on evaluating the benefits of amplitude control within IRS-equipped networks, especially against traditional methods that predominantly utilize full reflection capabilities. To achieve this, we introduce a novel optimization algorithm aimed at maximizing the data rate of the network, tailored to the unique characteristics of the mmWave communication channel. This approach employs a Weighted Minimum Mean Square Error (WMMSE) methodology for optimizing transmit beamforming with a fixed IRS configuration. Subsequently, we apply Successive Convex Approximation (SCA) techniques to adjust both the amplitude and phase shift of the IRS elements, assuming a predetermined transmit beamforming setup.

The core of our analysis is demonstrated through comprehensive simulation results, which underline the efficacy of amplitude control in maximizing the utility of IRS-assisted wireless systems. Specifically, our findings highlight the significant potential of employing amplitude control strategies within mmWave networks. By finely tuning the reflective properties of IRS elements, we can not only overcome some of the inherent drawbacks of mmWave communication but also unlock new dimensions of network performance optimization. This approach marks a pivotal step towards realizing the full potential of IRS technology in enhancing mmWave wireless networks, offering a promising pathway to overcome the complex challenges posed by high-frequency wireless communication.

We summarize the main contribution of this chapter as follows:

- **Introduction of IRS in mmWave Networks:** The chapter discusses the integration of IRS with mmWave communication systems, highlighting IRS as a novel solution to address the inherent challenges of mmWave technology, such as limited signal propagation and absorption by obstacles.
- **Optimization of IRS Parameters:** It introduces a new optimization algorithm designed to maximize the network data rate specifically for mmWave channels. This algorithm uniquely focuses on the amplitude control of the IRS elements, aside from the traditional phase shift control, to enhance network performance.
- **Utilization of Advanced Optimization Techniques:** The chapter details the application of WMMSE for transmit beamforming optimization when IRS settings are fixed. Additionally, it employs SCA for the fine-tuning of both amplitude and phase shifts of IRS elements, providing a dual approach to optimization.
- **Simulation Results and Performance Gains:** Through comprehensive simulation, the chapter demonstrates the efficacy of controlling the amplitude of IRS reflections. This is particularly shown to significantly enhance the performance of IRS-assisted mmWave wireless networks, suggesting that amplitude control is a promising avenue to fully exploit the potential of IRS technology.
- **Application of IRS in mmWave VR Systems:** We then expand the scope of IRS applications by exploring its integration into mmWave-based Virtual Reality (VR) systems. This includes optimizing the placement and radiation patterns of IRSs to enhance user experience and system efficiency in VR applications.
- **IRS Location and Radiation Pattern Optimization:** A significant contribution is the detailed optimization of the location and radiation patterns of IRSs, improving the performance and immersive quality of mmWave VR systems, marking a novel approach in the field.

These contributions collectively advance the understanding of how IRS technology can be effectively deployed in mmWave networks to overcome their limitations, providing valuable insights into the optimization of network performance through intelligent reflective surfaces.

This chapter is structured as follows: Section 7.2 introduces the system model for an active IRS-aided mmWave network. Section 7.3 formulates the sum data rate maximization problem and resolves it using an AO algorithm. In Section 7.4, preliminary results are presented for the active IRS network to explore scenarios where an active IRS is superior to a passive one. Section 7.5 provides insights and practical applications, with a special focus on a VR use case. We study location optimization and resource allocation for the IRS-assisted VR network in Section 7.6. Section 7.7 formulates the sum data rate problem for the IRS-assisted VR use case, with the solution provided in Section 7.8. The complexity of the resource allocation problem for this use case is analyzed in Section 7.9. An evaluation setup is deployed, and results for the IRS-assisted VR network are analyzed in Section 7.10. Finally, Section 7.11 concludes the chapter.

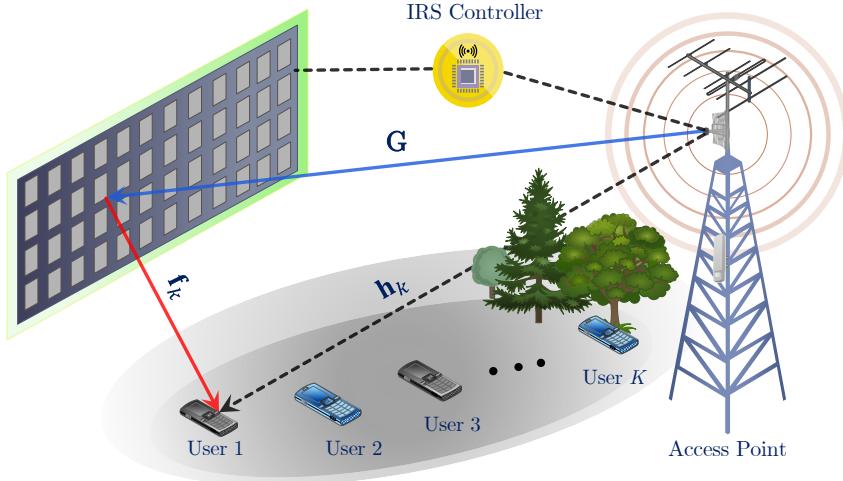


Figure 7.1: An IRS-assisted multi-user MISO mmWave wireless communication system. The AP is equipped with N_t -antennae and serves K single-antenna mobile users in the downlink. The active IRS has M reflecting elements in the AP's LoS signals.

Notations: Throughout this chapter, the following notations are used. The capital bold face letters are used to denote matrices while using the small bold and small normal face to denote vectors and scalars, respectively. \mathbf{I} represents an identity matrix, $\mathbb{C}^{x \times y}$ is a $x \times y$ complex-valued matrix. The superscript $(\cdot)^H$ is the conjugate transpose of a matrix, and transpose of a matrix is expressed as $(\cdot)^T$. The notations $\mathbb{E}[\cdot]$, $tr(\cdot)$, and $\text{diag}(\cdot)$ are used to denote the statistical expectation, trace and diagonalization operator, respectively. $\nabla f(\cdot)$ reads as the gradient of the function $f(\cdot)$ and $\frac{\partial f(a)}{\partial g}$ is the first derivative of f with respect to g at $g = a$. $|\mathbf{x}|$ is the 2-norm of vector \mathbf{x} . $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ gives the distribution of a Circularly Symmetric Complex Gaussian (CSCG) random vector with the mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, where \sim means “with the distribution of.” $\ln(\cdot)$ represents the natural logarithm of its argument, and $\Re\{\cdot\}$ signifies the real part of the argument.

7.2 System Model of an IRS-assisted mmWave Network

In this section, we focus on the design of a communication network that benefits from the inclusion of an IRS to aid a MISO multi-user setup. As shown in Fig. 7.1, an AP, equipped with N_t antennas, is tasked with managing downlink communication to K users in this network. These users are collectively represented by the set $\mathcal{K} = \{1, \dots, K\}$, which enumerates them in such a way that each user is distinctly identified.

Central to our discussion is the deployment of an active IRS, which comprises M elements capable of modifying the phase and amplitude of incoming signals; this ability essentially makes these elements act as phase shifters and amplifiers. The location of the IRS is

strategically chosen to be in the Line-of-Sight (LoS) of the AP — such as on a building's facade — to facilitate optimal signal reflection towards the users. The set of these reflective elements is denoted by $\mathcal{M} = \{1, \dots, M\}$.

A distinguishing feature of the IRS in this scenario is its ability to dynamically switch between modes that allow reflection and modes that do not, catering to various operational phases like downlink transmission and channel estimation. This adaptability is crucial for optimizing the network's performance across different communication phases. Additionally, the model presupposes the availability of perfect Channel State Information (CSI) both at the AP and the IRS and operates under a flat-fading channel model assumption. This model simplifies the analysis by assuming uniform channel characteristics across the transmission band.

We use the term '*active*' to describe the IRS's capability in adjusting both the phase and amplitude of reflected signals, enhancing its utility in the network by providing a more subtle control over signal propagation. The IRS is not active in the sense that it has its own data. This ability is encapsulated in the channel matrix $\mathbf{G} \in \mathbb{C}^{M \times N_t}$, which represents the complex interactions between the AP's antennas and the IRS's reflective elements. Understanding this matrix is key to harnessing the full potential of IRS-assisted communication, enabling signal enhancement and interference mitigation strategies that are responsive to the unique requirements of the IRS-aided network's users.

In our system model, the communication between the IRS and each user k is characterized by a reflecting channel vector $\mathbf{f}_k \in \mathbb{C}^{M \times 1}$, representing the complex path from the IRS's reflecting elements to the mobile user. Concurrently, the direct channel vector linking the AP directly to user k is denoted as $\mathbf{g}_k \in \mathbb{C}^{N_t \times 1}$, capturing the LoS and any multi-path effects between the AP's antennas and the user. To manipulate the signals reflected by the IRS, we employ a phase shift matrix $\Theta = \text{diag}(\theta_1, \theta_2, \dots, \theta_M)$, where each $\theta_m = e^{j\varphi_m}$ corresponds to the phase shift introduced by the m -th element of the IRS. Here, φ_m represents the phase shift angle for each element, ranging within $[0, 2\pi]$, and j is the square root of -1 , denoting the imaginary unit. This phase shift matrix is crucial for steering the reflected signals in desirable directions to enhance communication quality.

Additionally, we introduce an amplitude control matrix $\mathbf{A} = \text{diag}(\alpha_1, \alpha_2, \dots, \alpha_M)$, which governs the reflection amplitude of each IRS element. The amplitude values α_m are adjustable within the range of $[0, \alpha_{\max}]$, i.e., $\alpha_m \in [0, \alpha_{\max}]$, allowing for fine-tuned control over the signal's strength as it is reflected towards the users. This amplitude control is pivotal for optimizing the IRS's impact on the network, enabling a balance between signal enhancement and interference management.

Contrary to the assumptions made in prior works [86] that considered IRS devices to operate under full reflection (i.e., $\alpha_m = 1$ for all m), our study explores the joint optimization of both amplitude and phase shift controls at the IRS. This dual-parameter optimization approach is aimed at unleashing the full potential of IRS technology, facilitating a more nuanced and effective enhancement of signal propagation and reception in the wireless network. By adopting this comprehensive control strategy, we endeavor to maximize the benefits of IRS-assisted communication, making the reflected signals meet the specific needs of the network's configuration and the users it serves.

7.2.1 A Simplified mmWave Channel Model

Full CSI knowledge helps disclose the upper bound of the performance gain [136]. To accurately characterize the mmWave communication channel, we employ the Saleh-Valenzuela model, a well-established representation that captures the essence of mmWave propagation [240]. This model is a statistical channel model primarily used to represent the propagation characteristics of radio waves, encapsulating the complex interactions between transmitted signals and the environment, including the effects of path loss, multipath propagation, and the specific spatial characteristics unique to mmWave frequencies. The simplified Saleh-Valenzuela mmWave channel matrix \mathbf{G} is given by:

$$\mathbf{G} = \sqrt{N_t \times M} \sum_{i=0}^{\Delta} \gamma_i \zeta_t \zeta_r \boldsymbol{\alpha}_M(\varphi_i) \boldsymbol{\alpha}_{N_t}(\theta_i), \quad (7.1)$$

where, Δ , represents the total number of paths, including the LoS and NLoS paths. ζ_t and ζ_r are the transmit and receive antenna gains, respectively. Additionally, γ_i stands for complex gain of i -th path, φ_i , and θ_i are the angles of arrival and the angle of departure for the i -th path, respectively [240]. Also, the array response function of the IRS can be expressed as:

$$\boldsymbol{\alpha}_M(\varphi) = \frac{1}{\sqrt{M}} \left[e^{j2\pi(d/\lambda)\sin(\varphi)} \right], \quad (7.2)$$

$$\boldsymbol{\alpha}_{N_t}(\theta) = \frac{1}{\sqrt{N_t}} \left[e^{j2\pi(d/\lambda)\sin(\theta)} \right], \quad (7.3)$$

where λ and d denote the mmWave wavelength and the antenna spacing, respectively. Therefore, the channel gain from the IRS to the user k is:

$$\mathbf{f}_k = \sqrt{M} \zeta_t \zeta_r \gamma_k \boldsymbol{\alpha}_M(\varphi), \quad \forall k \in \mathcal{K}. \quad (7.4)$$

Since mmWave links are notably prone to obstructions, the direct channel gain between the AP and each user in the network can be significantly weakened. This susceptibility to blockages is a critical factor in designing and optimizing communication systems in the mmWave spectrum. Given this context, the formulation of the transmitted signal at the AP is essential for understanding system performance and is expressed as follows:

$$\mathbf{x} = \sum_{k=1}^K \mathbf{w}_k s_k, \quad \forall k \in \mathcal{K}, \quad (7.5)$$

where s_k denotes the transmit data symbol intended for the k -th user and $\mathbf{w}_k \in \mathbb{C}^{N_t \times 1}$ is the transmit beamforming vector associated with that user. The beamforming vector directs the transmitted signal's energy toward the intended user, thereby optimizing the signal's reception and mitigating interference with other users. Furthermore, the aggregate transmit data vector for all K users in the system is defined as $\mathbf{s} = [s_1, \dots, s_K]$, containing the data symbols for each user. It is important to note the assumption of normalized power for the transmit data symbols, which is mathematically represented as $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = \mathbf{I}$. This normalization implies that the expected value of the outer product of \mathbf{s} and its conjugate transpose \mathbf{s}^H equals the identity matrix \mathbf{I} , indicating that the transmit symbols are uncorrelated and have unit power. This assumption simplifies the analysis and design of the communication

system, allowing for clearer insights into the effects of beamforming and channel conditions on system performance. Consequently, the signal received by mobile user k is a result of both direct and reflected transmission paths. The formulation of the received signal can be expressed as follows:

$$y_k = (\mathbf{g}_k^H + \mathbf{f}_k^H \mathbf{A} \boldsymbol{\Theta} \mathbf{G}) \mathbf{x} + \mathbf{f}_k^H \mathbf{A} \boldsymbol{\Theta} n_d + n_k, \quad \forall k \in \mathcal{K}. \quad (7.6)$$

Here, the received signal y_k at user k is a sum of several components:

- The first component, $\mathbf{g}_k^H \mathbf{x}$, represents the direct signal from the AP to user k , considering the direct channel vector \mathbf{g}_k^H .
- The second component, $\mathbf{f}_k^H \mathbf{A} \boldsymbol{\Theta} \mathbf{G} \mathbf{x}$, models the signal that is reflected by the IRS before reaching user k . This involves the reflecting channel vector \mathbf{f}_k^H , the amplitude control matrix \mathbf{A} , the phase shift matrix $\boldsymbol{\Theta}$, and the AP to IRS channel matrix \mathbf{G} .
- The third component, $\mathbf{f}_k^H \mathbf{A} \boldsymbol{\Theta} n_d$, accounts for the noise introduced by the IRS, characterized as dynamic noise n_d , which includes the noise arising from both the input signal and the inherent noise of the IRS's electronic components due to the amplification noise power [241].
- The last term, n_k , represents the ambient noise at the receiver of user k , also known as static noise.

Both noise terms n_d and n_k are modeled as Additive White Gaussian Noise (AWGN) with a circularly symmetric Gaussian distribution, denoted by $n_d \sim \mathcal{CN}(\mathbf{0}, \sigma_d^2 \mathbf{I}_{N_t})$ and $n_k \sim \mathcal{CN}(0, \sigma_k^2)$, respectively. Here, σ_d^2 and σ_k^2 represent the variance of the dynamic and static noise components, illustrating the stochastic nature of these noise sources within the communication environment. Now, we quantify the quality of a received signal in relation to the background noise and interference from other signals. For user k in an IRS-aided mmWave network, the received Signal-to-Interference-plus-Noise Ratio (SINR) can be comprehensively described by considering both direct paths and paths reflected by the IRS. Accordingly, the SINR at the receiver k can be expressed as:

$$\gamma_k = \frac{|(\mathbf{g}_k^H + \mathbf{f}_k^H \mathbf{A} \boldsymbol{\Theta} \mathbf{G}) \mathbf{w}_k|^2}{\sum_{i=1, i \neq k}^K |(\mathbf{g}_k^H + \mathbf{f}_k^H \mathbf{A} \boldsymbol{\Theta} \mathbf{G}) \mathbf{w}_i|^2 + \sigma_d^2 |\mathbf{f}_k^H \mathbf{A} \boldsymbol{\Theta}|^2 + \sigma_k^2}, \quad \forall k \in \mathcal{K}. \quad (7.7)$$

This SINR formulation comes in handy in understanding how the IRS's reflective capabilities can significantly influence the received signal quality by not only enhancing the desired signal but also by potentially increasing the interference and noise levels. We develop our problem formulation next based on this SINR formulation.

7.3 Sum Data Rate Problem Formulation for the IRS-assisted mmWave Network

In this section, we maximize the total sum data rate by optimizing the corresponding transmit beamforming matrix $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_K] \in \mathbb{C}^{N_t \times K}$ at the AP and reflection coefficients at the IRS. The associated optimization problem is formulated as:

$$P_1 : \max_{\mathbf{W}, \boldsymbol{\theta}_m, \alpha_m} \sum_{k=1}^K R_k \quad (7.8a)$$

$$\text{s.t. : } \sum_{k=1}^K |\mathbf{w}_k|^2 \leq P_{\max}, \quad (7.8b)$$

$$0 \leq \theta_m < 2\pi, \quad \forall m \in \mathcal{M}, \quad (7.8c)$$

$$0 \leq \alpha_m \leq \alpha_{\max}, \quad \forall m \in \mathcal{M}, \quad (7.8d)$$

$$\sum_{k=1}^K |\mathbf{A}\boldsymbol{\Theta}\mathbf{G}\mathbf{w}_k|^2 + \sigma_d^2 |\mathbf{A}\boldsymbol{\Theta}|^2 \leq P_A, \quad (7.8e)$$

where

$$R_k = \ln(1 + \gamma_k) \\ = \left(1 + \frac{|\mathbf{(g}_k^H + \mathbf{f}_k^H \mathbf{A} \boldsymbol{\Theta} \mathbf{G}) \mathbf{w}_k|^2}{\sum_{i=1, i \neq k}^K |\mathbf{(g}_k^H + \mathbf{f}_k^H \mathbf{A} \boldsymbol{\Theta} \mathbf{G}) \mathbf{w}_i|^2 + \sigma_d^2 |\mathbf{f}_k^H \mathbf{A} \boldsymbol{\Theta}|^2 + \sigma_k^2} \right), \quad \forall k \in \mathcal{K}, \quad (7.9)$$

Within the realm of wireless communication system optimization, the challenge often lies in addressing non-convex optimization problems, where the objective function and constraints do not adhere to convexity principles. Such is the case with the optimization problem denoted as P_1 , characterized by its non-convex nature. This complexity arises from the inherent structure of P_1 's objective function and its constraints, which defy straightforward solutions. In P_1 , the constraint labeled as (7.8b) sets a cap on the total transmit power emanating from the AP, establishing a threshold at P_{\max} for the maximum permissible transmit power. This limitation is crucial for adhering to regulatory standards and managing interference within the network. Further, the constraints specified in (7.8c) and (7.8d) define the flexibility afforded by each reflecting element on the IRS. Specifically, these elements can be individually adjusted in terms of their phase ($\boldsymbol{\Theta}$) and reflection amplitude (\mathbf{A}) coefficients. This adaptability introduces a significant degree of freedom into the optimization problem, thereby facilitating enhanced performance gains across the wireless network. Finally, the constraint (7.8e) implies that the power amplified by an active IRS is constrained not to surpass a specified maximum allowance, denoted by P_A . This limitation ensures that the IRS operates within safe and efficient power levels, optimizing the system's overall energy consumption and performance.

Given the non-convexity of P_1 due to the non-convex functions in the objection and constraints, finding an optimal solution presents a formidable challenge. To navigate this complexity, we advocate for an iterative strategy that alternates between optimizing power allocation and phase shifts. Initially, with a given reflection coefficient $\boldsymbol{\Theta}$ and amplitude control \mathbf{A} , the algorithm seeks the optimal transmit beamforming vector \mathbf{W} . Subsequently, it recalibrates to find the best values for \mathbf{W} given fixed $\boldsymbol{\Theta}$ and \mathbf{A} . This iterative loop continues until the solution stabilizes at an optimal value for the objective function. The ensuing segments of this section will delve deeper into the mechanics of this iterative algorithm, elucidating

its procedural steps and underlying logic.

7.3.1 Optimization with respect to the Transmit Beamforming

Given the reflection coefficient matrix Θ and the amplitude control matrix \mathbf{A} , the optimization problem P_1 can be simplified. This adjustment focuses the problem on optimizing the transmit beamforming vectors, represented as \mathbf{W} , at AP within the constraints of the pre-determined Θ and \mathbf{A} . This scenario allows for a targeted optimization strategy, aiming to enhance the network's performance by adjusting the AP's transmission parameters to harmonize with the static configuration of the IRS. Thus, for given Θ and \mathbf{A} , the optimization problem P_1 is reduced to:

$$P_2 : \max_{\mathbf{W}} \sum_{k=1}^K R_k \quad (7.10a)$$

$$\text{s.t.} : \sum_{k=1}^K \|\mathbf{w}_k\|^2 \leq P_{\max}, \quad (7.10b)$$

$$\sum_{k=1}^K |\mathbf{A}\Theta\mathbf{G}\mathbf{w}_k|^2 + \sigma_d^2 |\mathbf{A}\Theta|^2 \leq P_A, \quad (7.10c)$$

where $R_k = \ln(1 + \gamma_k)$ with the SINR, γ_k , given by:

$$\gamma_k = \frac{|\mathbf{h}_k^H \mathbf{w}_k|^2}{\sum_{i=1, i \neq k}^K |\mathbf{h}_k^H \mathbf{w}_i|^2 + \sigma_d^2 |\mathbf{f}_k^H \mathbf{A}\Theta|^2 + \sigma_k^2}, \forall k \in \mathcal{K}, \quad (7.11)$$

in which

$$\mathbf{h}_k^H = \mathbf{g}_k^H + \mathbf{f}_k^H \mathbf{A}^H \Theta^H \mathbf{G}, \quad \forall k \in \mathcal{K} \quad (7.12)$$

denotes the combined channel from the AP and the IRS to the k -th user.

To address the complexity of optimizing the original problem P_1 with the refined focus as outlined in P_2 , we employ the Weighted Minimum Mean Square Error (WMMSE) algorithm. The main idea of the WMMSE algorithm is to “transform the objective of maximizing the weighted sum rate into an equivalent problem of minimizing the weighted sum mean square error.” This transformation facilitates the use of Alternating Optimization (AO) techniques, enabling an iterative refinement of the solution by optimizing over one set of variables at a time while keeping others fixed.

The application of the WMMSE algorithm begins with the assumption that the signal intended for user k , s_k , is decoded with the help of an equalizer u_k . Consequently, the estimated signal at user k is derived through this decoding process. The iterative nature of the WMMSE algorithm allows for sequential adjustments to the transmit beamforming vectors, equalizers, and weight matrices, progressively converging to an optimal or near-optimal

solution that enhances the network's overall performance. Therefore, by assuming the signal s_k is decoded by using the equalizer u_k , the estimated signal at user k becomes:

$$\hat{s}_k = u_k y_k, \quad \forall k \in \mathcal{K}. \quad (7.13)$$

Given the assumption that the signals s_k and the noise n_k are independent across all users k in the set \mathcal{K} , the Mean Square Errors (MSEs) for estimating the transmitted signals can be quantified. The MSE for user k , denoted as e_k , measures the expected value of the squared difference between the estimated signal \hat{s}_k and the actual transmitted signal s_k . Mathematically, this is expressed as:

$$e_k = \mathbb{E} \left[|\hat{s}_k - s_k|^2 \right], \quad \forall k \in \mathcal{K}. \quad (7.14)$$

Based on this definition, the MSE for each user k can be computed using the following formula:

$$e_k = |u_k|^2 B_k - 2\Re \{ u_k \mathbf{h}_k^H \mathbf{w}_k \} + 1, \quad \forall k \in \mathcal{K}, \quad (7.15)$$

where B_k is defined as:

$$B_k = |\mathbf{h}_k^H \mathbf{w}_k|^2 + \sum_{i=1}^K |\mathbf{h}_k^H \mathbf{w}_i|^2 + \sigma_d^2 |\mathbf{f}_k^H \mathbf{A} \boldsymbol{\Theta}|^2 + \sigma_k^2, \quad \forall k \in \mathcal{K}. \quad (7.16)$$

The parameter B_k accumulates the power of the signal intended for user k , the interference from other users, and noise impacts, including the noise amplified through the IRS (σ_d^2) and ambient noise at the user (σ_k^2).

To achieve the optimal reception quality, the Minimum Mean Square Error (MMSE) equalizer for user k , denoted as y_k^{MMSE} , is derived by minimizing the MSE with respect to the equalizer setting. The optimal MMSE equalizer is given by:

$$y_k^{\text{MMSE}} = \mathbf{w}_k^H \mathbf{h}_k B_k^{-1}, \quad \forall k \in \mathcal{K}, \quad (7.17)$$

This expression is obtained by setting the derivative of e_k with respect to y_k to zero, i.e., $\frac{\partial e_k}{\partial y_k} = 0$, assuming that all transmit beamforming vectors \mathbf{W} are held fixed. This optimization step ensures that the equalizer is tuned to minimize the impact of interference and noise, thereby enhancing the accuracy of the signal estimation at each receiver.

Incorporating the optimal MMSE equalizer obtained earlier, (7.17), into the MSE formula, (7.15), refines our understanding of the system's performance. This insertion allows for the derivation of the MMSE for user k , which is represented as follows:

$$e_k^{\text{MMSE}} = \min_{u_k} e_k = B_k^{-1} \left(B_k - |\mathbf{h}_k^H \mathbf{w}_k|^2 \right), \quad \forall k \in \mathcal{K}. \quad (7.18)$$

This equation highlights the minimized error achievable through the application of the MMSE equalizer, thereby optimizing the reception quality at each user's receiver. Furthermore, this minimized error directly relates to the SINR for user k , which is expressed as:

$$\gamma_k = (e_k^{\text{MMSE}})^{-1} - 1, \quad \forall k \in \mathcal{K}. \quad (7.19)$$

Consequently, the rate of communication for the k -th user, denoted R_k , can be recalculated as [242]:

$$R_k = -\ln (e_k^{\text{MMSE}}), \quad \forall k \in \mathcal{K}, \quad (7.20)$$

which illustrates the relationship between MMSE, SINR, and the achievable rate within this communication system. Beyond this, the concept of an augmented Weighted Mean Square Error (WMMSE) is introduced to further refine the system's optimization process as follows:

$$E_k = \rho_k e_k - \ln(\rho_k), \forall k \in \mathcal{K}, \quad \forall k \in \mathcal{K}. \quad (7.21)$$

where ρ_k represents the weight assigned to the MSE of the k -th user. This weighted formulation allows for an additional layer of optimization by adjusting ρ_k to balance the emphasis on error minimization and rate maximization across users, ultimately enhancing the overall system performance through a more nuanced optimization approach.

To find the optimal equalizer, which aligns with the MMSE equalizer, we begin by setting the derivative of the augmented WMMSE with respect to the equalizer u_k to zero, that is $\frac{\partial E_k}{\partial u_k} = 0$. This process yields the optimal augmented WMMSE value for the MMSE equalizer as:

$$E_k (y_k^{\text{MMSE}}) = \rho_k e_k^{\text{MMSE}} - \ln(\rho_k), \quad (7.22)$$

Following this, to determine the optimal weighting factor for the MMSE, ρ_k^* , we now take the derivative of the augmented WMMSE with respect to the weight ρ_k and set it to zero. The optimal weight of the MMSE is achieved as follows:

$$\frac{\partial E_k (y_k^{\text{MMSE}})}{\partial \rho_k} = 0 \rightarrow \rho_k^* = (\ln 2. e_k^{\text{MMSE}})^{-1}, \forall k \in \mathcal{K}, \quad (7.23)$$

Finally, motivated by the data rate WMMSE relationship in (7.23), the optimization problem P_2 is transformed into:

$$P_3 : \min_{\mathbf{W}, \boldsymbol{\rho}, \mathbf{u}} \sum_{k=1}^K [\rho_k e_k^{\text{MMSE}} - \ln(\rho_k)] \quad (7.24a)$$

$$\text{s.t.} : \sum_{k=1}^K \|\mathbf{w}_k\|^2 \leq P_{\max}, \quad (7.24b)$$

$$\sum_{k=1}^K |\mathbf{A}\boldsymbol{\Theta}\mathbf{G}\mathbf{w}_k|^2 + \sigma_d^2 |\mathbf{A}\boldsymbol{\Theta}|^2 \leq P_A. \quad (7.24c)$$

The optimization problem involves determining the optimal transmit beamforming weights and the set of equalizers and weights to minimize the WMMSE. Specifically, $\boldsymbol{\rho} = [\rho_1, \dots, \rho_K]$ represents the MSE weights for each user, and $\mathbf{u} = [u_1, \dots, u_K]$ is the equalizer coefficients.

It can be easily shown that when we minimize P_3 with respect to $\boldsymbol{\rho}$ and \mathbf{u} , respectively, the MMSE solutions $(\boldsymbol{\rho}^{\text{MMSE}}, \mathbf{u}^{\text{MMSE}})$ including the corresponding MMSE weights and equalizers can be achieved. While fixing $\{\boldsymbol{\rho}, \mathbf{u}\}$, the optimization problem P_3 is now changed into a Quadratic Constrained Quadratic Programming (QCQP) problem at the point \mathbf{W} . Thus, a standard convex optimization package like CVX can be employed to solve the optimization problem efficiently [174, 243, 129, 151, 1].

7.3.2 Optimization with respect to the Active IRS Parameters

In this subsection, we focus on the optimization of the active IRS parameters, specifically targeting the IRS's reflection coefficients, i.e., the phase and the amplitude, with a given transmit beamforming matrix \mathbf{W} . When the transmit beamforming matrix is found from the previous subproblem, P_1 , we adapt the initial optimization problem of P_1 , to address this new focus of finding Θ and \mathbf{A} with fixed \mathbf{W} . The objective now shifts towards fine-tuning the IRS's reflection capabilities to further enhance the system's performance. Thus, we redefine P_1 as:

$$P_4 : \max_{\theta_m, \alpha_m} \sum_{k=1}^K R_k \quad (7.25a)$$

$$\text{s.t.} : \sum_{k=1}^K |\mathbf{A}\Theta\mathbf{G}\mathbf{w}_k|^2 + \sigma_d^2 |\mathbf{A}\Theta|^2 \leq P_A \quad (7.25b)$$

$$0 \leq \theta_m < 2\pi, \quad \forall m \in \mathcal{M}, \quad (7.25c)$$

$$0 \leq \alpha_m \leq \alpha_{\max}, \quad \forall m \in \mathcal{M}. \quad (7.25d)$$

To enhance clarity in mathematical expressions, let's introduce a matrix Ξ that directly represents the combined effects of amplitude adjustments and phase shifts applied by the active IRS. Consequently, this matrix which is the product of \mathbf{A} and Θ can be expressed as:

$$\Xi \triangleq \text{diag}(a_1 e^{j\theta_1}, \dots, a_M e^{j\theta_M}). \quad (7.26)$$

Furthermore, for each reflecting element m , we define $\chi_m = \alpha_m e^{j\theta_m}, \forall m \in \mathcal{M}$, and aggregate these individual element settings into a vector $\chi = [\chi_1, \dots, \chi_M]^T$. This vector χ represents the configured state of the IRS in terms of both amplitude and phase adjustments for each of its elements. With these definitions in place, we can now represent the interaction between the transmitted signal, the IRS, and the receiving user k more compactly. Specifically, the effective channel through the IRS for user k can be represented as:

$$\mathbf{f}_k^H \Xi \mathbf{G} \mathbf{w}_k \triangleq \psi_k^H \chi, \quad \forall k \in \mathcal{K}, \quad (7.27)$$

$$\mathbf{g}_k^H \mathbf{w}_k \triangleq \tilde{g}_k, \quad \forall k \in \mathcal{K}, \quad (7.28)$$

where

$$\psi_k = (\text{diag}(\mathbf{f}_k^H) \mathbf{G} \mathbf{w}_k)^H, \quad \forall k \in \mathcal{K}. \quad (7.29)$$

This is, (7.29), the effective channel from the AP through the IRS to user k , after being reflected and phase-shifted. The optimization problem P_4 incorporates these simplified representations. By introducing a slack variable ζ_k , can be formulated as follows:

$$P_5 : \max_{\boldsymbol{\chi}, \boldsymbol{\zeta}} \sum_{k=1}^K \log_2(1 + \zeta_k) \quad (7.30a)$$

$$\text{s.t.} : \frac{|\boldsymbol{\psi}_k^H \boldsymbol{\chi} + \tilde{g}_k|^2}{\sum_{i \neq k} |\boldsymbol{\psi}_k^H \boldsymbol{\chi} + \tilde{g}_i|^2 + \sigma_d^2 |\boldsymbol{\chi}|^2 + \sigma_k^2} \geq \zeta_k, \forall k \in \mathcal{K}, \quad (7.30b)$$

$$\sum_{k=1}^K |\boldsymbol{\Xi} \mathbf{w}_k|^2 + \sigma_d^2 |\boldsymbol{\Xi}|^2 \leq P_A, \quad (7.30c)$$

$$|\chi_m| \leq \alpha_m, \quad \forall m \in \mathcal{M}, \quad (7.30d)$$

$$0 \leq \theta_m < 2\pi, \quad \forall m \in \mathcal{M}, \quad (7.30e)$$

$$0 \leq \alpha_m \leq \alpha_{\max}, \quad \forall m \in \mathcal{M}, \quad (7.30f)$$

where $\boldsymbol{\zeta} = [\zeta_1, \dots, \zeta_K]^T$. To tackle the non-convexity issue presented by constraint (7.30b), we introduce a slack variable, denoted as Υ_k , into our optimization framework. This allows us to reformulate (7.30b) into two separate inequalities, enhancing the tractability of the problem. The redefined constraints leverage Υ_k and ζ_k to express the original non-convex condition in a more manageable form. Specifically, we can represent the relationship between Υ_k and ζ_k as follows:

$$\sum_{i \neq k} |\boldsymbol{\psi}_i^H \boldsymbol{\chi} + \tilde{g}_i|^2 + \sigma_d^2 |\boldsymbol{\chi}|^2 + \sigma_k^2 \leq \Upsilon_k, \forall k \in \mathcal{K}, \quad (7.31)$$

$$|\boldsymbol{\psi}_k^H \boldsymbol{\chi} + \tilde{g}_k|^2 \geq \Upsilon_k \zeta_k, \quad \forall k \in \mathcal{K}, \quad (7.32)$$

where

$$\Upsilon_k \zeta_k = \frac{1}{2}(\Upsilon_k + \zeta_k)^2 - \frac{1}{2}(\Upsilon_k^2 + \zeta_k^2), \quad \forall k \in \mathcal{K}. \quad (7.33)$$

This transformation results in two new constraints for the optimization problem, effectively replacing the original non-convex constraint (7.30b). Among these two new constraints, the first one, (7.31), is convex and straightforward to handle. However, the second constraint, (7.32), maintains a non-convex nature, posing a challenge for direct optimization. To convert (7.32) into a form amenable to convex optimization techniques, we employ the Successive Convex Approximation (SCA) algorithm. The SCA method based on the difference of the two concave function approaches as follows [174, 129, 151]:

$$\begin{aligned} 2\Re \left\{ (\boldsymbol{\psi}_k^H \boldsymbol{\chi}^{[t']})^H \boldsymbol{\psi}_k^H \boldsymbol{\chi} \right\} - |\boldsymbol{\psi}_k^H \boldsymbol{\chi}^{[t']} + \tilde{g}_k|^2 &\geq \\ \frac{1}{2}(\Upsilon_k + \zeta_k)^2 - \frac{1}{2}((\Upsilon_k^{[t']})^2 + (\zeta_k^{[t']})^2) \\ - (\Upsilon_k^{[t']})(\Upsilon_k - (\Upsilon_k^{[t']})) - (\zeta_k^{[t']})(\zeta_k - (\zeta_k^{[t']})), \forall k \in \mathcal{K}. \end{aligned} \quad (7.34)$$

This method effectively addresses the non-convexity by approximating (7.32) as a difference of two concave functions, a strategy that simplifies the optimization process. The adaptation involves iterative updates, where $\boldsymbol{\chi}^{[t']}$, $\Upsilon_k^{[t']}$, and $\zeta_k^{[t']}$ denote the solutions obtained in the

$[t']$ -th iteration of the algorithm. Through this iterative procedure, we are equipped to solve the ensuing convex problem, progressively refining our estimates to converge towards an optimal solution for P_5 .

This step-by-step approach, articulated within the framework of SCA and facilitated by the introduction of slack variables, enables the effective solution of the original non-convex optimization problem, paving the way for achieving optimal system performance through strategic IRS parameter adjustment. Now, we solve the following convex problem:

$$P_6 : \max_{\boldsymbol{\chi}, \boldsymbol{\zeta}, \boldsymbol{\gamma}} \sum_{k=1}^K \log_2(1 + \zeta_k) \quad (7.35)$$

$$\text{s.t.} : \sum_{k=1}^K |\boldsymbol{\Xi} \mathbf{w}_k|^2 + \sigma_d^2 |\boldsymbol{\Xi}|^2 \leq P_A, \quad (7.36)$$

$$|\boldsymbol{\chi}_m| \leq \alpha_m, \quad \forall m \in \mathcal{M}, \quad (7.37)$$

$$0 \leq \theta_m < 2\pi, \quad \forall m \in \mathcal{M}, \quad (7.38)$$

$$0 \leq \alpha_m \leq \alpha_{\max}, \quad \forall m \in \mathcal{M}, \quad (7.39)$$

$$\sum_{i \neq k} |\boldsymbol{\psi}_i^H \boldsymbol{\chi} + \tilde{g}_i|^2 + \sigma_d^2 |\boldsymbol{\chi}|^2 + \sigma_k^2 \leq \gamma_k, \quad \forall k \in \mathcal{K}, \quad (7.40)$$

$$2\Re \left\{ (\boldsymbol{\psi}_k^H \boldsymbol{\chi}^{[t']} + \tilde{g}_k)^H \boldsymbol{\psi}_k^H \boldsymbol{\chi} \right\} - \left| \boldsymbol{\psi}_k^H \boldsymbol{\chi}^{[t']} + \tilde{g}_k \right|^2 \geq \\ \frac{1}{2} (\gamma_k + \zeta_k)^2 - \frac{1}{2} ((\gamma_k^2)^{[t']} + (\zeta_k^2)^{[t']}) \\ - (\gamma_k)^{[t']} (\gamma_k - (\gamma_k)^{[t']}) - (\zeta_k)^{[t']} (\zeta_k - (\zeta_k)^{[t']}), \quad \forall k \in \mathcal{K}. \quad (7.41)$$

It is important to highlight that the solution quality of the original optimization problem is expected to improve or, at the very least, remain consistent (monotonically non-decreasing) following the application of this iterative algorithm [2]. Consequently, this makes the problem well-suited for standard convex optimization software packages, such as CVX [175]. This approach ensures efficient computation and facilitates the practical implementation of the optimization solution.

7.4 Simulation Results for the IRS-assisted mmWave Network

In this section, we present the numerical results obtained from simulating an IRS-assisted mmWave communication network. The simulation setup considers an AP located within a rectangular area of dimensions 50×50 meters. The AP is positioned at the origin $(0, 0)$ meters, while the IRS is strategically placed at $(30, 0)$ meters. All mobile users are assumed to be randomly distributed within this rectangular zone. The IRS under consideration is

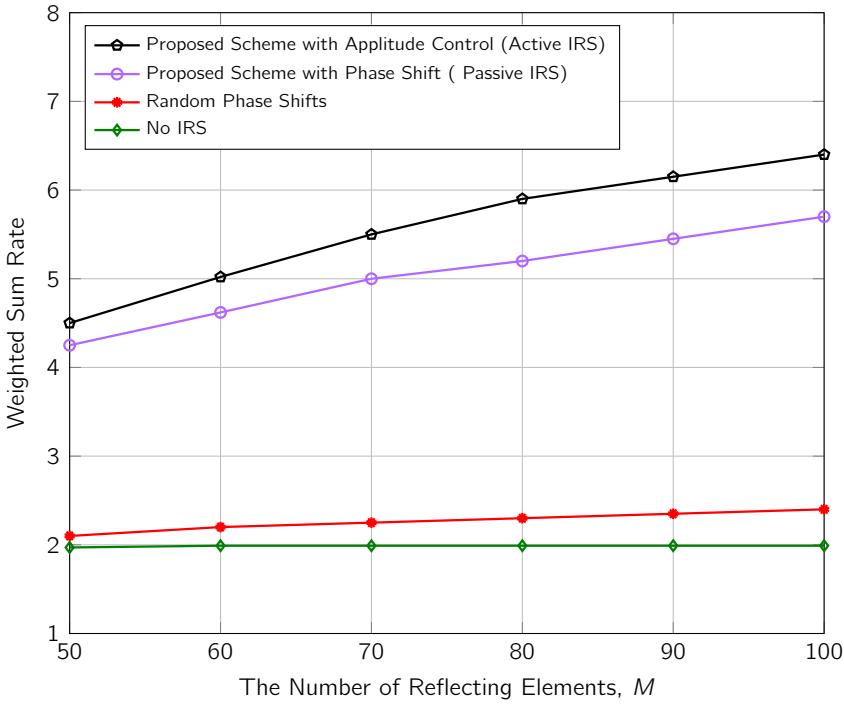


Figure 7.2: Average sum data rate versus the number of reflecting elements.

equipped with 60 reflecting elements to modulate the incident signal effectively. The simulation scenario includes $K = 6$ users to evaluate the network's performance comprehensively. The maximum allowable transmit power from the AP is restricted to $P_{\max} = 30$ dBm, whereas the active IRS is constrained by a maximum power allowance of $P_A = 13$ dBm.

The path loss model employed for the simulations follows the equation $35.3 + 37.6 \log_{10}(d_k)$ dB, where d_k denotes the distance between the AP and user k in meters. The channel's bandwidth is set to 500 MHz, with a system parameter for the number of paths $\Delta = 5$. The transmit power antenna gain (ζ_t) is 9.82 deciBel-isotropic (dBi), and the receive power antenna gain (ζ_r) is 0 dBi. The interference power γ_i adheres to configurations detailed in [86]. To ensure precise results, the convergence criterion for the iterative AO processes is established at 10^{-2} . Additionally, the static noise power observed at each user is set to $\sigma_k^2 = -114$ dBm, and the dynamic noise variance, accounting for the noise contributions from the IRS's electronics and processing, is $\sigma_d^2 = -110$ dBm, as described in references [240, 239]. This setup aims to provide a realistic and comprehensive understanding of the potential benefits and challenges associated with deploying IRS technology in mmWave networks. Moreover, all statistical results were derived by aggregating data from a comprehensive series of simulation trials, each producing numerous random realizations of the mmWave channel gains.

Fig. 7.2 illustrates the relationship between the average weighted sum data rate and the number of reflecting elements in the system. It is observed that the average sum rate exhibits a positive correlation with the number of reflecting elements. This trend can be attributed

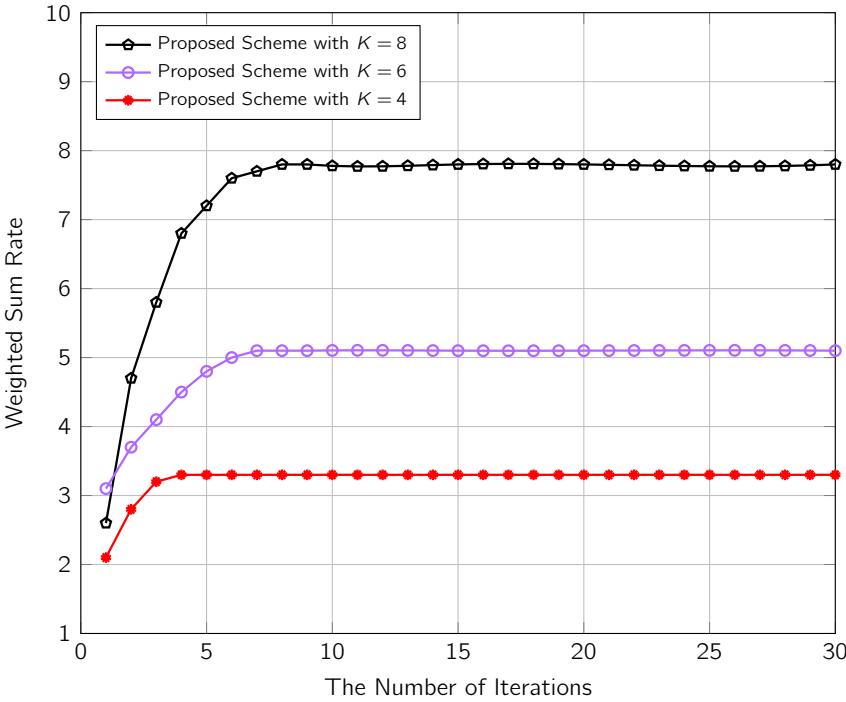


Figure 7.3: Average sum data rate versus the number of iterations.

to the enhanced communication paths facilitated by a larger array of reflecting elements, which, in turn, bolsters the system's overall performance gain. However, a noteworthy observation is the relatively marginal performance improvement when utilizing random phase shifts, especially in comparison to configurations with no IRS. This phenomenon underscores the significance of optimizing the phase shift parameters at the IRS to harness its full potential in communication systems. Moreover, the comparison between active and passive IRS implementations reveals the superior efficiency of active IRS in augmenting communication performance. This advantage becomes particularly pronounced with an increase in the number of reflecting elements. Our proposed scheme distinguishes itself by not only optimizing the phase shifts at the IRS but also by fine-tuning the amplitude reflections. The beneficial impact of amplitude control becomes increasingly apparent with a higher number of users. This is attributed to the exacerbation of multiuser interference within mmWave channels, necessitating amplitude control as a means to mitigate such adverse effects. Furthermore, the results indicate that an active IRS is capable of mitigating the "multiplicative fading" phenomenon, thereby achieving significant gains in the sum data rate. This analysis proves the critical role of both phase and amplitude optimization in leveraging IRS technology to enhance mmWave communication networks.

Figure 7.3 presents the convergence behavior of our proposed algorithm across various user counts. The plot reveals that the algorithm typically converges within approximately 10 iterations. Furthermore, an interesting trend is observed where the sum data rate escalates with an increase in the number of users. This finding suggests that our proposed scheme efficiently scales with user count, enhancing network capacity and demonstrating

the effectiveness of the algorithm in managing multiuser environments.

7.5 Insights and Practical Applications

So far, we introduced an optimization framework aimed at enhancing the performance of mmWave wireless communication systems by strategically designing the reflection coefficients of an IRS in this chapter. Specifically, we considered a scenario involving an active IRS-assisted MISO mmWave system, with the goal of maximizing the network's data rate through the dual approach of optimizing the AP's beamforming strategy and the IRS's reflection coefficients. The optimization problem was tackled through an iterative process, where the first sub-problem pertaining to beamforming was addressed using the Weighted Minimum Mean Square Error (WMMSE) technique. Subsequently, the second sub-problem related to the IRS's coefficients was approached with the Successive Convex Approximation (SCA) method. The simulation results provided compelling evidence of the superiority of active IRS over its passive counterpart, particularly in its ability to mitigate performance degradations attributed to multiplicative fading phenomena. These results showed the potential of active IRS technology to significantly improve the efficiency and reliability of mmWave communication networks. Looking ahead, our future research will explore the implications of active and passive IRS deployment in mmWave channels under conditions of imperfect or full Channel State Information (CSI) knowledge. This endeavor could include the development and analysis of codebook designs for AP beamforming, aiming to further refine the performance enhancements achievable through sophisticated IRS configurations.

In the next section, we proceed by exploiting the insights gained from our study so far towards a practical use case within IRS-supported networks, specifically focusing on IRS-aided mmWave networks for Virtual Reality (VR).

7.5.1 Practical Use Case: IRS-aided mmWave Network for Virtual Reality

Next-generation VR technology promises unprecedented levels of user immersion and support for intricate multiuser Virtual Experiences (VEs). Given the cost-effective and passive nature of IRSs, we study the optimal design of a multi-user IRS-assisted VR network in the remainder of this chapter. Specifically, we explore the strategic deployment of an IRS within a confined space as a function of the dynamic trajectories of fully immersed VR users. The core objective will be to maximize the sum data rate across all VR users, enhancing the overall quality and responsiveness of the VR environment.

To achieve this, we focus on optimizing several key components within the network:

- The active beamforming strategies at the AP efficiently direct the signal toward the IRS and subsequently to the users.
- The precise placement of the IRS within the indoor environment to ensure optimal signal reflection and coverage.

- The adjustment of IRS phase shifts and radiation patterns to maximize signal reception quality for VR users, accommodating their movement and interaction within the space.

Operating within mmWave frequencies, known for their high bandwidth and susceptibility to blockages, presents unique challenges and opportunities for indoor VR applications. To navigate the complexities of this multi-faceted optimization problem, we introduce an AO algorithm. This algorithm decomposes the problem into manageable sub-problems, each solvable in an optimal manner:

1. **Active Beamforming at the AP:** Utilizing Maximum Ratio Transmission (MRT) techniques, we derive optimal beamforming vectors that maximize the signal power directed toward each user through the IRS.
2. **IRS Phase Shifts (Passive Beamforming):** A quadratic transformation approach is employed to calculate optimal closed-form solutions for IRS phase shifts, ensuring efficient signal reflection towards users.
3. **IRS Placement:** We identify the most effective locations, globally optimal, for IRS elements within the indoor space to enhance signal path and coverage.
4. **Radiation Pattern Optimization:** An analysis of monotonic optimal radiation patterns is conducted to further refine the IRS's ability to support dynamic user positions and orientations.

We conclude this use case by providing simulation results to demonstrate the significant impact of strategic IRS resource allocation and placement on enhancing signal stability and maximizing throughput for each VR user. This comprehensive approach not only addresses the inherent challenges of mmWave communication within indoor environments but also unlocks new potentials for immersive and interactive VR experiences. Let's dive into the VR world!

7.6 Location Optimization and Resource Allocation of IRS in a Multi-User Indoor mmWave VR Network

VR is anticipated to transform our digital interactions in various domains such as health-care, tourism, education, entertainment, and occupational safety [244]. VR systems are poised to accommodate multiple fully immersed users who can freely navigate their VE in an indoor environment. To enable cost-effective indoor VE deployment, the deployment of IRS on the walls as a function of users' trajectory is a potential solution [245]. An IRS comprises large arrays of passive reflecting elements on a reconfigurable planar surface, capable of independently modifying the phase of an incoming signal before reflecting it towards its intended receiver. This capability of the IRS can significantly benefit VR users facing considerable path loss or blockages in the direct link, a common challenge in environments operating within the mmWave frequency band. The IRS introduces additional propagation pathways — namely, reflected channels [10], enhancing signal reach and reliability for better VE. Furthermore, the IRS offers added degrees of freedom through the phase shifts of

its reflective elements, which can be strategically manipulated to minimize interference and optimize signal quality [3]. Notably, IRSs are envisioned to be passive, cost-effective, and flexible solutions, making them ideally suited for indoor VR streaming setups. They can serve as ‘soft’ environmental boundaries, seamlessly integrating into the indoor architecture to support immersive VR experiences without intrusive hardware installations [246].

Prior research has illuminated the significant benefits that IRSs bring to multi-user wireless communication systems, notably enhancing data transfer rates and reliability — a critical aspect for bandwidth-intensive and latency-sensitive applications like VR applications [3, 247, 248, 249]. Chaccour *et al.* demonstrated how IRS technology could significantly boost the sum data rate and ensure the reliability of data transfers within VR environments. This advancement is particularly crucial in VR contexts where immersive experiences demand high data throughput and low latency to maintain user immersion and interaction fidelity [247]. Jalali *et al.* explored the optimization of IRS design focusing on energy efficiency and admission control maximization. Their work is particularly relevant for Internet of Things (IoT) applications, which often involve transmitting short packets — a scenario where IRS can play a vital role in ensuring efficient and reliable communication [3]. Besser *et al.* introduced an innovative phase hopping algorithm for IRS-supported systems aimed at enhancing data transfer reliability. Their approach is notable for its operation without the need for CSI, simplifying the implementation and reducing the overhead typically associated with adapting to channel variations [248]. Zhou *et al.*’s investigation into a latency minimization problem in a multi-user secure IRS-aided VR delivery network addresses the challenge of imperfect CSI. Their work shows the importance of optimizing communication networks to support latency-sensitive applications like VR, ensuring that immersive experiences are not degraded by delays or security concerns [249].

These studies collectively emphasize the revolutionary role of IRS in boosting the capabilities of wireless networks for cutting-edge applications like VR. The subsequent sections of this chapter aim to demonstrate how an IRS-aided mmWave network could be used for a VR environment. The mmWave frequency band, known for its vast bandwidth, stands as an excellent candidate for VR technologies, yet it faces challenges like severe path loss and sensitivity to blockages. IRS technology emerges as a solution to these challenges, offering a means to dynamically control the propagation environment and improve signal coverage and fidelity. Through exploring this synergy, our goal is to navigate the complex challenges inherent in mmWave transmissions, thereby laying the groundwork for future VR experiences that are deeply engaging, immersive, interactive, fluid, and seamless. To the best of our knowledge, this work represents a pioneering effort to have the design of an IRS-assisted indoor VR network optimized; This optimization specifically considers the deployment of the IRS within a constrained three-dimensional space directly influenced by the trajectories of VR users.

The contributions of the subsequent sections (the VR use case) in this chapter, focusing on an IRS-enabled multi-user mmWave VR environment, are summarized as follows:

- Introduction of an IRS-enabled multi-user mmWave VR environment setup, where the IRS is strategically placed on a wall to enhance data transmission from a multi-antenna Access Point (AP) to single-antenna HMDs.
- Aim to maximize the sum data rate of all HMDs by:

- Optimizing the IRS's location to facilitate communication in a confined indoor environment.
- Optimizing active beamforming at the AP for efficient data transfer.
- Adjusting phase shifts and radiation patterns in response to VR users' trajectories, modeled using redirected walking techniques.
- Design of a resource allocation algorithm to maximize the system's sum data rate, adhering to peak transmit power feasibility and Quality of Service (QoS) constraints.
- Application of the AO algorithm to segment the non-convex main optimization problem into four solvable sub-problems:
 - Identification of MRT as the optimal beamforming strategy at the AP.
 - Derivation of a closed-form optimal solution for IRS phase shifts using quadratic transformation.
 - Global optimization for the IRS's placement based on the first-order derivative of the objective function.
 - Determination of the optimal radiation pattern in a closed-form format, using the monotonicity of the transformed objective function.
- Simulation results underscore the effectiveness of combining passive beamforming at the IRS with location-based IRS placement and optimal active beamforming at the AP. This approach significantly enhances data rates over various baseline schemes, highlighting the impact of IRS technology in improving mmWave VR networks.

The following sections delineate the focus on leveraging IRS technology to overcome mmWave communication challenges for a VR use case through optimized IRS deployment and functionality. Let's begin.

7.7 VR IRS-assisted System Model and Problem Formulation

As depicted in Fig. 7.4, we consider an IRS-assisted MISO communication system in which IRS relays data to a single antenna HMD VR user. The direct LoS link between the transmitter and receiver is considered to be blocked. Doppler effect caused by the HMD's mobility is presumed to be fully compensated. In this network, an AP with L antennas serves a set of HMD represented as $\mathcal{K} = \{1, \dots, K\}$ using an IRS whose elements are denoted by $\mathcal{M} = \{1, \dots, M\}$. Our goal is to fine-tune the IRS's resource allocation based on a placement optimization problem to achieve maximum SNR over a fixed time span $T > 0$. The time duration T is partitioned into N uniformly spaced time intervals, given by $T = N\xi_t$. Specifically, ξ_t denotes the length of each individual time slot, and \mathcal{N} is defined as the set of all these time slots, represented by $\mathcal{N} = \{1, \dots, N\}$.

To accurately model the spatial dynamics of this system, we employ a 3D Cartesian coordinate framework. The AP is stationed at a fixed point denoted by $\mathbf{a} = [a_x, a_y, a_z]^T \in \mathbb{R}^{3 \times 1}$.

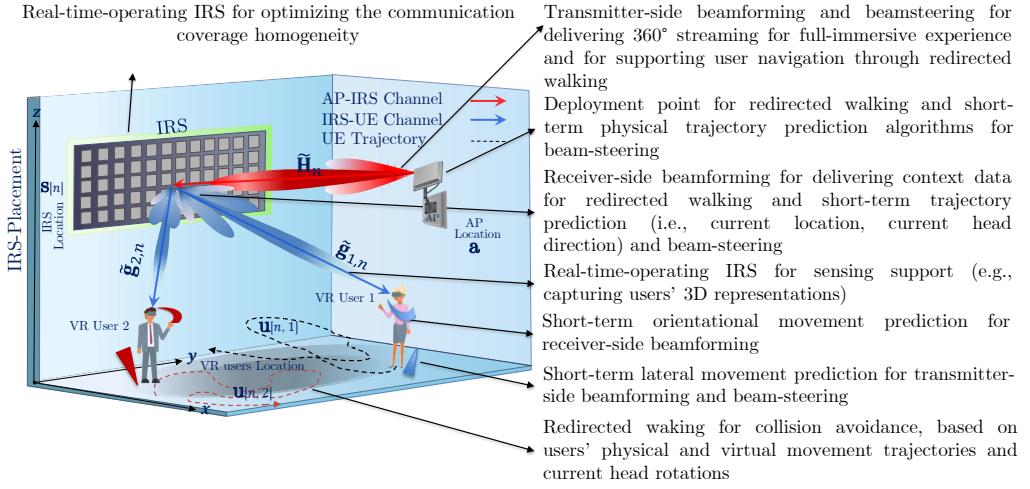


Figure 7.4: A multi-user IRS-assisted full-immersive VR scenario. The IRS is deployed on one of the walls, and a multi-antenna AP transmits data indirectly to a set of single-antenna HMD via the IRS.

In contrast, the HMD users traverse a predefined path on the ground, their movements captured by the trajectory function $\mathbf{u}[n, k] = [u_x[n, k], u_y[n, k], u_z[n, k]]^T \in \mathbb{R}^{3 \times 1}$, which is guided by redirected walking principles [250]. The potential locations for deploying IRS resources, when visualized on a vertical plane, are encapsulated by $\mathbf{s}[n] = [s_x[n], s_y[n], s_z[n]]^T \in \mathbb{R}^{3 \times 1}$. Furthermore, we limit the spatial domain to four distinct vertical planes, labelled \mathcal{H}_1 through \mathcal{H}_4 , each representing a possible region for IRS resource allocation, with \mathcal{H}_1 to \mathcal{H}_4 being specified as:

$$\mathcal{H}_1 : y_{min} < s_y[n] < y_{max}, z_{min} < s_z[n] < z_{max}, s_x[n] = x_{mix}, \quad \forall n \in \mathcal{N}, \quad (7.42)$$

$$\mathcal{H}_2 : y_{min} < s_y[n] < y_{max}, z_{min} < s_z[n] < z_{max}, s_x[n] = x_{max}, \quad \forall n \in \mathcal{N}, \quad (7.43)$$

$$\mathcal{H}_3 : x_{min} < s_x[n] < x_{max}, z_{min} < s_z[n] < z_{max}, s_y[n] = y_{min}, \quad \forall n \in \mathcal{N}, \quad (7.44)$$

$$\mathcal{H}_4 : x_{min} < s_x[n] < x_{max}, z_{min} < s_z[n] < z_{max}, s_y[n] = y_{max}, \quad \forall n \in \mathcal{N}. \quad (7.45)$$

To provide a comprehensive understanding of the deployment strategy for IRS within an indoor VR environment, we focus on positioning the IRS on the corner walls of a room. This strategic placement is illustrated in Fig. 7.4, where we explore the optimal locations for IRS installation to enhance communication between the AP and HMD. The chosen regions allow the IRS to effectively manage and redirect signals across the entire room, thereby mitigating potential signal obstructions and optimizing coverage. Moreover, we consider the radiation pattern of the IRS in our analysis. The radiation pattern is critical as it determines the directional distribution of the reflected signal strength, which in turn affects the overall effectiveness of the IRS in enhancing communication. By analyzing and optimizing the radiation pattern, we can ensure that the IRS not only boosts the signal strength but also directs the signal in a manner that maximizes the SINR at the receiver's end, particularly for VR users who require high-bandwidth and low-latency connections for immersive experiences. This consideration is crucial for realizing the full potential of IRS-assisted mmWave networks.

in providing robust and efficient communication pathways in complex indoor environments. The radiation pattern of the IRS is given by:

$$F(\psi_k, \varphi) = \begin{cases} \cos^3(\psi_k), & \psi_k \in [0, \pi/2], \varphi \in [0, 2\pi], \\ 0, & \psi_k \in (\pi/2, \pi], \varphi \in [0, 2\pi], \end{cases} \quad (7.46)$$

where $\psi_k, \forall k \in \mathcal{K} \cup \{0\}$, and φ represent the elevation and azimuth angles, respectively, from the IRS to the AP/HMD link [251]. The radiation pattern, as defined by $F(\psi_k, \varphi)$, primarily focuses on the elevation angle ψ_k , indicating a strong directional preference of the reflected signals when ψ_k lies within the range of $[0, \pi/2]$. Notably, this pattern is maintained consistently across various azimuth angles φ , as evidenced by the function's independence from φ in the specified range of ψ_k [251]. For the sake of simplicity and to maintain focus on the significant impact of the elevation angle on the IRS's radiation pattern, we will henceforth use $F(\psi_k)$ in place of $F(\psi_k, \varphi)$ in subsequent discussions and equations. This simplification allows for a clearer analysis, emphasizing the critical role of elevation angle in optimizing the IRS's contribution to the communication link's quality and reliability. Given these conditions, the dynamic channel between AP and IRS, and between IRS and the k -th HMD adheres to the free-space path loss model, which can be detailed as [252, 8]:

$$\tilde{\mathbf{H}}_n = \mathbf{H} \sqrt{\beta_{0,n} F(\psi_0)}, \quad \forall n \in \mathcal{N}, \quad (7.47)$$

$$\tilde{\mathbf{g}}_{k,n} = \mathbf{g}_{k,n} \sqrt{\beta_{k,n} F(\psi_k)}, \quad \forall k \in \mathcal{K}, \forall n \in \mathcal{N}, \quad (7.48)$$

$$\beta_{k',n} = \sqrt{c_0} \|\mathbf{d}_{k',n}\|^{-\frac{\alpha_{k'}}{2}}, \quad \forall k' \in \mathcal{K} \cup \{0\}, \forall n \in \mathcal{N}, \quad (7.49)$$

where $\beta_{0,n}$ and $\beta_{k,n}$ symbolize the path loss with c_0 being the reference channel power at a distance of 1 meter, while α_0 and $\alpha_k, \forall k \in \mathcal{K}$ are the path loss exponents of AP–IRS and IRS– k -th HMD links, respectively. Moreover, the small-scale fading of the links between AP and IRS, (7.47), and between IRS and the k -th HMD, (7.48), are denoted by $\mathbf{H} \in \mathbb{C}^{M \times L}$ and $\mathbf{g}_{k,n} \in \mathbb{C}^{M \times 1}$, respectively¹. Besides, the distance vectors from the IRS to the AP and k -th HMD, coming from (7.49), are respectively given by:

$$\begin{aligned} \mathbf{d}_{0,n} &= \mathbf{s}[n] - \mathbf{a} = \\ &= [s_x[n] - a_x, s_y[n] - a_y, s_z[n] - a_z]^T, \\ &\quad \forall n \in \mathcal{N}, \end{aligned} \quad (7.50)$$

$$\begin{aligned} \mathbf{d}_{k,n} &= \mathbf{s}[n] - \mathbf{u}[n, k] = \\ &= [s_x[n] - u_x[n, k], s_y[n] - u_y[n, k], s_z[n] - u_z[n, k]]^T, \\ &\quad \forall k \in \mathcal{K}, \forall n \in \mathcal{N}. \end{aligned} \quad (7.51)$$

Therefore, the received signal of k -th HMD becomes:

$$y_{k,n} = \tilde{\mathbf{g}}_{k,n}^H \Theta \tilde{\mathbf{H}} \mathbf{w}_k b_{k,n} + n_k, \quad \forall k \in \mathcal{K}, \forall n \in \mathcal{N}, \quad (7.52)$$

where $b_{n,k}$ is the bearing-information transmitted symbol for the k -th HMD with normalized power at n -th time slot, $\mathbf{w}_k \in \mathbb{C}^{L \times 1}$ is the transmit beamforming vector, and n_k is the AWGN noise, which follows a complex normal distribution with zero mean and variance σ_k^2 . The IRS

¹By incorporating the IRS's radiation pattern, $F(\psi_k)$, into the path loss model, we refine our understanding of how the IRS can manipulate signal paths to enhance link quality and system performance.

phase shifts matrix is represented by Θ and is defined as $\Theta = \text{diag}(\theta_1, \theta_2, \dots, \theta_M) \in \mathbb{C}^{M \times M}$, where $\theta_m = \varrho_m e^{j\vartheta_m} \in \mathbb{C}$ characterizes the reflection coefficient of the m -th IRS element, in which $\varrho_m \in [0, 1]$ is the reflection amplitude², and $\vartheta_m \in [0, 2\pi]$ is the phase shifts. Ultimately, by assuming there is no multi-user interference, we represent the SNR at k -th HMD in time slot n as:

$$\gamma_{k,n}(\mathbf{W}, \Theta, \beta, \Psi) = \frac{\beta_{0,n} F(\psi_0) \beta_{k,n} F(\psi_k) |\mathbf{g}_{k,n}^H \Theta \mathbf{H} \mathbf{w}_k|^2}{\sigma_k^2}, \quad \forall k \in \mathcal{K}, \forall n \in \mathcal{N}, \quad (7.53)$$

where \mathbf{W} , β , and Ψ are the collection of \mathbf{w} 's, β 's and ψ 's according to $\mathbf{W} \triangleq [\mathbf{w}_1, \dots, \mathbf{w}_K]$, $\beta \triangleq [\beta_{0,0}, \dots, \beta_{0,n}, \dots, \beta_{K,N}]$, and $\Psi \triangleq [\psi_0, \dots, \psi_K]$. Consequently, the achievable data rate of the k -th HMD during the n -th time slot, measured in [bit/s/Hz], can be expressed as:

$$R(\mathbf{W}, \Theta, \beta, \Psi) = \log_2(1 + \gamma_{k,n}(\mathbf{W}, \Theta, \beta, \Psi)), \quad \forall k \in \mathcal{K}, \forall n \in \mathcal{N}. \quad (7.54)$$

Finally, the sum data rate for all HMD can be written as:

$$R_{\text{tot}}(\mathbf{W}, \Theta, \beta, \Psi) = B \sum_{\forall k \in \mathcal{K}} \sum_{\forall n \in \mathcal{N}} R(\mathbf{W}, \Theta, \beta, \Psi), \quad (7.55)$$

where B represents the bandwidth of the network.

In our pursuit, the primary objective is to elevate the system-level data rate within an IRS-assisted single-cell multi-user indoor VR network. Achieving this goal necessitates a strategic approach that leverages the unique capabilities of the IRS. This includes the judicious allocation of IRS resources based on the spatial dynamics of the network environment, which we refer to as location-based IRS resource allocation. Fine-tuning these parameters is the aim of our system-level data rate optimization. By meticulously configuring the transmit beamforming strategies employed at the AP, as well as carefully adjusting the IRS's phase shifts and radiation patterns, we can achieve this goal. Consequently, we can frame the sum data rate optimization problem as:

$$P_7 : \max_{\mathbf{W}, \Theta, \beta, \Psi} R_{\text{tot}}(\mathbf{W}, \Theta, \beta, \Psi), \quad (7.56a)$$

$$\text{s.t. : } \text{tr}(\mathbf{W} \mathbf{W}^H) \leq P_{\text{AP}}^{\max}, \quad (7.56b)$$

$$|\theta_m| \leq 1, \quad \forall m \in \mathcal{M}, \quad (7.56c)$$

$$\mathbf{s}[n] \in \mathcal{H}_q, \quad \forall q \in \{1, \dots, 4\}, \quad (7.56d)$$

$$\mathbf{u}[n, k] \in \mathcal{U}_k, \quad \forall k \in \mathcal{K}, \quad (7.56e)$$

$$0 \leq \psi_k \leq \pi/2, \quad \forall k \in \mathcal{K} \cup \{0\}. \quad (7.56f)$$

Constraint (7.56b) guarantees that the total transmission power from the AP does not exceed predefined maximum levels. This limitation is crucial for adhering to regulatory power

²To maximize reflection efficiency, we assume the amplitudes of all passive elements to be one [3, 12], i.e., $\varrho_m = 1, \forall m \in \mathcal{M}$.

standards and for minimizing potential interference with other wireless systems. Constraint (7.56c) specifies the bounds within which the reflection coefficient for every IRS element must operate. By setting these bounds, the constraint ensures that the IRS operates within its optimal reflection capabilities, thus maximizing the enhancement of signal strength and coverage. Constraint (7.56d) ensures the IRS is positioned in one of the corner walls of the room. Next, constraint (7.56f) confines the radiation pattern. This constraint is vital for ensuring that the IRS's signal reflection does not inadvertently increase interference or degrade the network's overall performance. Finally, constraint (7.56e) confirms that each VR user follows a predefined redirected walking path denoted as \mathcal{U}_k [253].

Given the presence of a non-concave objective function and the non-convex nature of constraint (7.56c), the optimization problem laid out in (7.56) is distinctly nonconvex [9]. This inherent complexity makes it challenging to derive a straightforward solution for the problem. As a result, AO methods or approximations are needed to address non-convexity effectively for effectively tackling the non-convex aspects of the problem, paving the way towards identifying viable solutions for enhancing the VR network's data rate and user experience..

7.8 Location Optimization and Resource Allocation of IRS in a VR Network

Optimization problem (7.56) exhibits non-convexity primarily due to the highly interdependent nature of the optimization variables involved. Such non-convex problems typically challenge conventional solution strategies, often lacking a straightforward well-organized method for resolution due to their complexity and the intricacies of the variable relationships.

Nonetheless, to navigate through these challenges, we introduce an AO strategy characterized by lower computational complexity, where a new objective function was proposed to avoid the feasibility problem. This approach is designed to iteratively converge towards a sub-optimal solution by decomposing the original problem into more manageable sub-problems, each tailored to address specific facets of the optimization problem:

- Initially, we redefine the objective function within the AO framework to circumvent potential issues related to solution feasibility. This redefinition ensures that the subsequent optimization steps are grounded in a solvable context, enhancing the overall strategy's effectiveness.
- For the first and second sub-problems, we derive closed-form solutions for active beamforming at the AP and passive beamforming at the IRS, respectively. These solutions provide precise configurations for both active and passive beamformers, optimizing signal transmission and reflection to improve network performance.
- The third sub-problem focuses on global optimization concerning the IRS's resource allocation. This step critically evaluates and adjusts the IRS's resource distribution to ensure optimal network operation, particularly in terms of enhancing signal coverage and quality.

- In the final sub-problem, we examine the radiation pattern exhibited by the IRS. This investigation aims to understand and optimize the way the IRS manipulates incoming signals to maximize the system's data rate and user experience.

By employing this AO approach, we effectively address the inherent non-convexity of the original problem, facilitating the attainment of sub-optimal solutions that significantly improve the IRS-aided mmWave VR network's performance.

7.8.1 Step 1: AP Active (transmitter-side) Beamforming

In the first step of our optimization process, we concentrate on the aspect of active beamforming at the AP side. To approach this, we initially hold the IRS's phase shift matrix Θ , path loss coefficients β , and the set of elevation angles for IRS radiation pattern Ψ constant. This simplification allows us to isolate the active beamforming component of the system for targeted optimization. With these parameters fixed, we can then precisely focus on the optimization of the AP's active beamforming matrix. This is essential for efficiently directing the transmitted signal towards the IRS, thereby ensuring that the reflected signals are optimally relayed to the intended receivers, in this case, the HMDs of the VR users.

Given these considerations, the optimization problem dedicated to active beamforming is redefined with a transformed objective function. This new formulation is designed to capture the essence of maximizing the received signal power, subject to power constraints. By transforming the objective function, we adapt the problem to a more tractable form, facilitating the derivation of an optimal or near-optimal solution for the AP's active beamforming strategy. This step is crucial for laying the groundwork for subsequent optimization stages. Thus, the corresponding optimization problem with a transformed objective function can then be formulated:

$$P_8 : \max_{\mathbf{W}} \sum_{\forall k \in \mathcal{K}} \sum_{\forall n \in \mathcal{N}} A_{n,k}^0 |\mathbf{g}_{k,n}^H \Theta \mathbf{H} \mathbf{w}_k|^2, \quad (7.57a)$$

$$\text{s.t. : } \text{tr}(\mathbf{W} \mathbf{W}^H) \leq P_{\text{AP}}^{\max}, \quad (7.57b)$$

where

$$A_{n,k}^0 = \frac{\beta_{0,n} F(\psi_0) \beta_{k,n} F(\psi_k)}{\sigma_k^2}, \quad \forall k \in \mathcal{K}. \quad (7.58)$$

One can readily prove that the optimization problem (7.57) is affine, thus convex. This intrinsic characteristic of convexity paves the way for employing the well-established theoretical and practical tools of convex optimization to discern a closed-form solution that is globally optimal.

When considering the specific task of optimizing AP active beamforming in the context of our IRS-assisted mmWave VR network, it emerges that the Maximum Ratio Transmission (MRT) strategy stands out as the optimal approach [11]. The MRT, renowned for its

efficiency in maximizing the power of the signal received at the target, aligns perfectly with our objective to enhance the system's overall performance. The essence of MRT lies in its focus on aligning the transmit beamforming vector with the direction of the channel, thereby amplifying the signal strength at the receiver's end. This method is succinctly captured in the formula: Therefore, the optimal transmit beamforming can be given by [11]:

$$\mathbf{w}_k^* = \sqrt{P_{\text{AP}}^{\max}} (\mathbf{g}_{k,n}^H \boldsymbol{\Theta} \mathbf{H})^H / \|\mathbf{g}_{k,n}^H \boldsymbol{\Theta} \mathbf{H}\|, \forall k \in \mathcal{K}. \quad (7.59)$$

7.8.2 Step 2: IRS Passive (receiver-side) Beamforming

In the next step of our solution methodology, after establishing the optimal active beamforming strategy for AP, attention will be given to the passive beamforming executed by IRS. This component is pivotal in sculpting the signal path from the AP to the end receivers, in this case, the HMDs utilized in the VR network. With the active beamforming parameters now set, along with fixed β and Ψ , we optimize the IRS's passive beamforming. Given these fixed parameters, the optimization sub-problem focusing on IRS passive beamforming can be reformulated as:

$$P_9 : \max_{\boldsymbol{\Theta}} \sum_{\forall k \in \mathcal{K}} \sum_{\forall n \in \mathcal{N}} A_{n,k}^0 |\text{vec}(\boldsymbol{\Theta})^H \boldsymbol{\Upsilon}_n|^2, \quad (7.60a)$$

$$\text{s.t.} : |\theta_m| = 1, \forall m \in \mathcal{M}, \quad (7.60b)$$

where we used the following change of variables:

$$\mathbf{g}_{k,n}^H \boldsymbol{\Theta} \mathbf{H} \mathbf{w}_k = \text{vec}(\boldsymbol{\Theta})^H \boldsymbol{\Upsilon}_{k,n}, \quad \forall k \in \mathcal{K}, \forall n \in \mathcal{N}, \quad (7.61)$$

in which

$$\boldsymbol{\Upsilon}_n = \text{diag}(\mathbf{g}_{k,n}) \mathbf{H} \mathbf{w}_k, \quad \forall n \in \mathcal{N}. \quad (7.62)$$

Despite the non-convex nature of the problem (7.60) due to the unit modulus constraints, a closed-form solution can be derived based on the quadratic transform method. To do so, we rewrite the problem into its equivalent form as:

$$P_{10} : \max_{\boldsymbol{\Theta}} \sum_{\forall k \in \mathcal{K}} \sum_{\forall n \in \mathcal{N}} A_{n,k}^0 \left(-\text{vec}(\boldsymbol{\Theta})^H \mathbf{U} \text{vec}(\boldsymbol{\Theta}) + 2\Re\{\text{vec}(\boldsymbol{\Theta})^H \boldsymbol{\Upsilon}_n\} \right), \quad (7.63a)$$

$$\text{s.t.} : |\theta_m| = 1, \forall m \in \mathcal{M}, \quad (7.63b)$$

where $\mathbf{U} = \text{vec}(\boldsymbol{\Theta}) \boldsymbol{\Upsilon}_n^H$. Now, we obtain the following simpler upper bound to the quadratic

term $\text{vec}(\Theta)^H \mathbf{U} \text{vec}(\Theta)$:

$$\begin{aligned} \text{vec}(\Theta)^H \mathbf{U} \text{vec}(\Theta) &\leq \text{vec}(\Theta)^H \mathbf{Q} \text{vec}(\Theta) \\ &\quad - 2\Re \left\{ \text{vec}(\Theta)^H (\mathbf{Q} - \mathbf{U}) \text{vec}(\Theta^t)^H \right\} \\ &\quad + \text{vec}(\Theta^t)^H (\mathbf{Q} - \mathbf{U}) \text{vec}(\Theta^t), \end{aligned} \quad (7.64)$$

where $\mathbf{Q} = \lambda_{\max}(\mathbf{U}) \mathbf{I}_M$ and $\lambda_{\max}(\mathbf{U})$ corresponds to the maximum eigenvalue of the semi-positive definite matrix \mathbf{U} . Additionally, the superscript t indicates the feasible solution achieved during the t -th iteration. Thus, the objective function of (7.63) can be lower bounded by:

$$\begin{aligned} \left(-\text{vec}(\Theta)^H \mathbf{U} \text{vec}(\Theta) + 2\Re \{ \text{vec}(\Theta)^H \boldsymbol{\Upsilon}_n \} \right) &\geq -\lambda_{\max}(\mathbf{U}) \|\text{vec}(\Theta)\|^2 \\ &\quad + 2\Re \{ \text{vec}(\Theta)^H ((\mathbf{Q} - \mathbf{U}) \text{vec}(\Theta^t) + \boldsymbol{\Upsilon}_n) \} \\ &\quad - \text{vec}(\Theta^t)^H (\mathbf{Q} - \mathbf{U}) \text{vec}(\Theta^t), \quad \forall n \in \mathcal{N}. \end{aligned} \quad (7.65)$$

Then, we can reformulate the IRS passive beamforming sub-problem in the following manner:

$$P_{11} : \max_{\Theta} \sum_{\forall k \in \mathcal{K}} \sum_{\forall n \in \mathcal{N}} A_{n,k}^0 \Re \{ \text{vec}(\Theta)^H \boldsymbol{\Gamma}_n \}, \quad (7.66a)$$

$$\text{s.t. : } |\theta_m| = 1, \forall m \in \mathcal{M}, \quad (7.66b)$$

where

$$\boldsymbol{\Gamma}_n = (\mathbf{Q} - \mathbf{U}) \text{vec}(\Theta^t) + \boldsymbol{\Upsilon}_n, \quad \forall n \in \mathcal{N}. \quad (7.67)$$

Ultimately, it can be verified that the optimal solution to (7.66) is expressible in closed-form, as presented:

$$\theta_m = e^{j\arg(\boldsymbol{\Gamma}_m)}, \quad \forall m \in \mathcal{M}. \quad (7.68)$$

7.8.3 Step 3: IRS Placement at Optimal Locations

In this subsection, we formulate the subproblem wherein the IRS's placement is optimized with fixed active beamforming and fixed IRS's phase shifts and radiation pattern, i.e., \mathbf{W} , Ψ , Θ are known. The goal here is to determine the most effective positions for the IRS to maximize the network's overall SNR efficiency. This involves a careful consideration of how the IRS's location influences the path loss and signal quality between the AP and the end users, as well as between the IRS and the end users. Therefore, the optimization problem for the IRS's location-based resource allocation can be written as:

$$P_{12} : \max_{\mathbf{s}[n]} \sum_{\forall k \in \mathcal{K}} \frac{c_0^2 F(\psi_0) F(\psi_k) |\mathbf{g}_{k,n}^H \Theta \mathbf{H}|^2}{\sigma_k^2 \|\mathbf{s}[n] - \mathbf{a}\|^{\alpha_0} \|\mathbf{s}[n] - \mathbf{u}[n, k]\|^{\alpha_k}}, \quad (7.69a)$$

$$\text{s.t. : } \mathbf{s}[n] \in \mathcal{H}_q, \quad \forall q \in \{1, \dots, 4\}, \quad (7.69b)$$

$$\mathbf{u}[n, k] \in \mathcal{U}_k, \quad \forall k \in \mathcal{K}, \forall n \in \mathcal{N}, \quad (7.69c)$$

where β 's are replaced by the IRS' location decision variables, $\mathbf{s}[n]$. It can be seen that (7.69) is convex. Thus, an optimal solution for the IRS' placement can be found. Assuming $\alpha_k = 2$ and by setting the first-order derivative of the objective function with respect to $\mathbf{s}[n]$ to zero, we obtain the following equalities:

$$\begin{aligned} & \frac{(a_x - s_x[n])}{(a_x - s_x[n])^2 + (a_y - s_y[n])^2 + (a_z - s_z[n])^2} \\ &= \frac{(s_x[n] - u_y[n, k])}{(s_x[n] - u_y[n, k])^2 + (s_y[n] - u_y[n, k])^2 + (s_z[n] - u_z[n, k])^2}, \forall k \in \mathcal{K}, \forall n \in \mathcal{N}, \end{aligned} \quad (7.70)$$

$$\begin{aligned} & \frac{(a_y - s_y[n])}{(a_x - s_x[n])^2 + (a_y - s_y[n])^2 + (a_z - s_z[n])^2} \\ &= \frac{(s_y[n] - u_y[n, k])}{(s_x[n] - u_y[n, k])^2 + (s_y[n] - u_y[n, k])^2 + (s_z[n] - u_z[n, k])^2}, \forall k \in \mathcal{K}, \forall n \in \mathcal{N}, \end{aligned} \quad (7.71)$$

$$\begin{aligned} & \frac{(a_z - s_z[n])}{(a_x - s_x[n])^2 + (a_y - s_y[n])^2 + (a_z - s_z[n])^2} \\ &= \frac{(s_z[n] - u_z[n, k])}{(s_x[n] - u_y[n, k])^2 + (s_y[n] - u_y[n, k])^2 + (s_z[n] - u_z[n, k])^2}, \forall k \in \mathcal{K}, \forall n \in \mathcal{N}, \end{aligned} \quad (7.72)$$

where an iterative approach could be employed to determine the optimal locations for IRS. With the knowledge of the AP location and the HMDs' trajectory [254], we initialize with predefined values for $s_x[n]$, $s_y[n]$, and $s_z[n]$. From these, we deduce the optimal IRS coordinates iteratively, based on equations (7.70) – (7.72), while simultaneously satisfying the constraint (7.56d) and (7.56e).

7.8.4 Step 4: IRS Radiation Pattern Optimization

In the final phase of our optimization process, we now consider the last subproblem of optimizing the IRS the radiation pattern. This step ensures that the reflective capabilities of the IRS are maximally utilized, directing the reflected signals in a manner that optimally supports the network's operational requirements and enhances the end-user experience. This can

significantly impact the network's overall performance, particularly in environments where obstacles or physical layout may impede direct signal paths. By fine-tuning the radiation pattern, we aim to maximize the effective use of the IRS in steering signals toward desired directions, thereby optimizing the network's data rate and reliability. Thus, the IRS radiation pattern Ψ optimization sub-problem with fixed \mathbf{W} , Θ , and β can be recast as:

$$P_{13} : \max_{\Psi} \sum_{\forall k \in \mathcal{K}} \sum_{\forall n \in \mathcal{N}} A_{n,k}^1 \cos^3(\psi_0) \cos^3(\psi_k), \quad (7.73a)$$

$$\text{s.t. : } 0 \leq \psi_k \leq \pi/2, \forall k \in \mathcal{K} \cup \{0\}, \quad (7.73b)$$

where

$$A_{n,k}^1 = \frac{\beta_{0,n} \beta_{k,n} |\mathbf{g}_{k,n}^H \Theta \mathbf{H} \mathbf{w}_k|^2}{\sigma_k^2}, \quad \forall k \in \mathcal{K}, \forall n \in \mathcal{N}. \quad (7.74)$$

Unlike the preceding sub-problems, where we had to work out how to get to the closed-form solutions algebraically, obtaining a closed-form solution is markedly more straightforward. This simplification arises from the inherent monotonic properties of the cosine function that dominate the objective function in this particular scenario.

The key to this simplification lies in recognizing that the function $\cos^3(\psi_k)$, for all k in the combined set of users and the IRS itself $\mathcal{K} \cup \{0\}$, exhibits a monotonically decreasing behavior within the interval $0 \leq \psi_k \leq \frac{\pi}{2}$. This monotonicity facilitates the identification of the global maximum of the objective function directly by examining the endpoints of the specified interval in (7.56f). This implies that an optimal elevation angle can be found that results in the most favorable IRS radiation patterns. The final iterative-based AO approach is provided in **Algorithm 8**. This algorithm incorporates the solution from each sub-problem's optimization, iteratively refining the system's configuration to achieve the best possible network performance within the constraints of the given IRS-assisted communication scenario.

7.9 Analyzing Resource Allocation Complexity in IRS-aided VR Networks

In this section, we conduct an analysis of the computational complexity of our proposed algorithm. The AO algorithm iteratively tackles the four subproblems related to \mathbf{W} , Θ , β , and Ψ until convergence is reached. We obtain efficient closed-form solutions for the first two subproblems, as in (7.59) and (7.68), respectively. The last two subproblems have been convexified and can be efficiently solved in polynomial time using CVX [3]. The computational complexities associated with \mathbf{W} , Θ , β , and Ψ are as follows: $O_1 = \mathcal{O}(KL^3)$, $O_2 = \mathcal{O}(K^2 L NM + K^2 M^2 + M^3)$, $O_3 = \mathcal{O}((3N)(12N + 3NK)^3)$, and $O_4 = \mathcal{O}((K + 1)^3)$. Hence, the proposed AO algorithm's computational complexity can be approximated as $\mathcal{O}(KL^3 + K^2 L NM + K^2 M^2 + M^3 + N^4 K^3)$.

Algorithm 8 Iterative AO algorithm for a Multi-User Indoor mmWave VR Network

Input: Set the iteration index $i = 0$, set maximum number of iteration I_{\max} , and initialize $\mathbf{W} = \mathbf{W}^0$, $\Theta = \Theta^0$, $\beta = \beta^0$, and $\Psi = \Psi^0$.

- 1: **Repeat**
- 2: Solve problem P_8 (7.57) for given $\{\Theta^{i-1}, \beta^{i-1}, \Psi^{i-1}\}$ and use (7.59) to obtain the optimal solution \mathbf{W}^i .
- 3: Solve problem P_{11} (7.66) for given $\{\mathbf{W}^{i-1}, \beta^{i-1}, \Psi^{i-1}\}$ and use (7.68) to obtain the optimal solution Θ^i .
- 4: Solve problem P_{12} (7.69) for given $\{\mathbf{W}^{i-1}, \Theta^{i-1}, \Psi^{i-1}\}$ and use (7.70)–(7.72) to obtain the optimal solution β^i .
- 5: Solve feasibility problem P_{13} (7.73) for given $\{\mathbf{W}^{i-1}, \Theta^{i-1}, \beta^{i-1}\}$ to obtain the optimal solution Ψ^i .
- 6: **until** $i = I_{\max}$
- 7: **Return** $\{\mathbf{W}^*, \Theta^*, \beta^*, \Psi^*\} = \{\mathbf{W}^i, \Theta^i, \beta^i, \Psi^i\}$.

7.10 Evaluation Setup and Simulations Results for the IRS-assisted VR Use Case

We deploy a comprehensive simulation framework to evaluate the AO algorithm's efficacy in enhancing network performance for an IRS-assisted, full-immersive VR environment operating in the mmWave spectrum. This evaluation considers the spatial arrangement of HMD used by VR users, the Wireless Fidelity (Wi-Fi) AP, and the IRS within a 3D space. The strategic deployment of IRS resources plays a pivotal role in our simulation, focusing on the utilization of the environment's peripheral boundaries — namely, the four outer walls — while intentionally excluding the floor and ceiling from the IRS's operational domain. The AP is centrally placed on the ceiling at a height of 3 meters, ensuring a dominant vantage point for broadcasting signals to the HMD users below. The virtual environments navigated by the HMDs are varied in size, encompassing dimensions of 10×10 , 15×15 , and 20×20 squared meters, to evaluate the system's adaptability and performance across different spatial scales [254]. This varied environmental setup allows us to assess the AO algorithm's effectiveness in optimizing the IRS's influence on the network, ensuring the delivery of high-quality, immersive VR experiences under diverse spatial configurations.

The proposed AO algorithm is designed to be flexible, catering to a generic number of IRS elements. This flexibility allows for the allocation of IRS resources to be dynamically adjusted based on the evolving data rate requirements of future VR systems. Initially, our simulations consider the deployment of 200 IRS elements, with each element being sized at $\lambda/5$, a dimension that is optimal for enhancing the system's performance [255, 65]. To rigorously evaluate the performance of our AO algorithm within an IRS-assisted mmWave network, we utilize discrete-event Network Simulator version-3 (ns-3), specifically its 60 GHz Wi-Fi (WiGig) module. This module is used to facilitate analyzing the performance of the IEEE 802.11ad/ay protocols, providing a robust framework for assessing high-frequency mmWave communications [256]. Furthermore, we enhance the ns-3 simulator by integrating a mmWave propagation model that accounts for the influence of IRS (with its location dependability) on signal propagation. This model is critical for understanding how the IRS can modify the signal environment to meet the stringent requirements of high-fidelity VR

Table 7.1: Overview of Baseline Simulation Parameters for the IRS-assisted mmWave VR Network.

Parameter Name	Parameter Value
Application Type	OnOffApplication
Data Rate	150 Mbps
Flow Direction	DL
Payload Size	1448 Bytes
Transport Protocol	User Datagram Protocol (UDP)
MAC Queue Size	4000 Packets
Aggregation Type	Aggregated MAC Service Data Unit (A-MSDU) and Aggregated MAC Protocol Data Unit (A-MPDU)
MAC/PHY	CSMA/CA/SC DMG MCS-10
Transmit Power / Sectors	10 mW / 8
Receiver Noise Figure	10 dB
Operating Frequency	60.48 GHz

applications [252, 8]. A comprehensive summary of the simulation parameters used in our study, including the specifics of the IRS elements, the ns-3 configurations, and the mmWave operational frequency, is detailed in *Table 7.1*. This table serves as a reference for the simulation settings of our analysis, ensuring transparency and reproducibility in our findings³.

The “Optimal” approach follows **Algorithm 8** for dynamically adjusting the IRS configuration and resource allocation, taking into account the locations of the HMD, AP, and IRS, as well as the IRS radiation patterns. This method optimizes the IRS’s influence on the com-

³The terms “CSMA/CA/SC DMG MCS” refer to a combination of protocols and modulation/coding schemes used in wireless network technologies, particularly in the context of IEEE 802.11 standards, which include Wi-Fi technologies. Let’s break down each part:

1. CSMA/CA: This stands for Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA). It is a MAC protocol used in many wireless networking standards, including Wi-Fi. Its primary function is to minimize the risk of collisions by ensuring that a transmitting station listens to the medium before starting a transmission and by using various mechanisms to avoid simultaneous transmissions by multiple stations.
2. SC: This typically stands for Single Carrier (SC). In the context of wireless communication, especially in standards like IEEE 802.11ad (which uses mmWave frequencies), SC refers to a type of transmission where the data is sent using a single carrier frequency as opposed to multiple carriers (OFDM). It is often used for simpler, potentially more robust transmissions over certain types of channels.
3. DMG: Directional Multi-Gigabit (DMG). This is a term often associated with the IEEE 802.11ad amendment, which is part of the broader IEEE 802.11 standard focused on very high-throughput wireless networks operating in the 60 GHz spectrum. The “directional” part of DMG refers to the use of directional antennas that focus the energy in specified directions to enhance signal strength and range, crucial in high-frequency bands like 60 GHz where signal attenuation is significant.
4. MCS-10: Modulation and Coding Scheme (MCS), level 10. MCS levels define the specific modulation type and coding rate used to transmit data. Each level is a combination that provides a specific data rate and robustness to signal interference and noise. In the context of IEEE 802.11 standards, MCS-10 would specify a particular modulation type (e.g., Quadrature Amplitude Modulation (QAM)) and a coding rate, which together determine the transmission speed and reliability. MCS levels can vary across different amendments of the 802.11 standards (e.g., 802.11n, 802.11ac, 802.11ad).

These technologies collectively describe a communication setup that is capable of supporting very high-speed data transmission, particularly suited for environments where rapid data transfer is required over relatively short distances, such as in indoor scenarios or for applications like VR, where latency and throughput are critical.

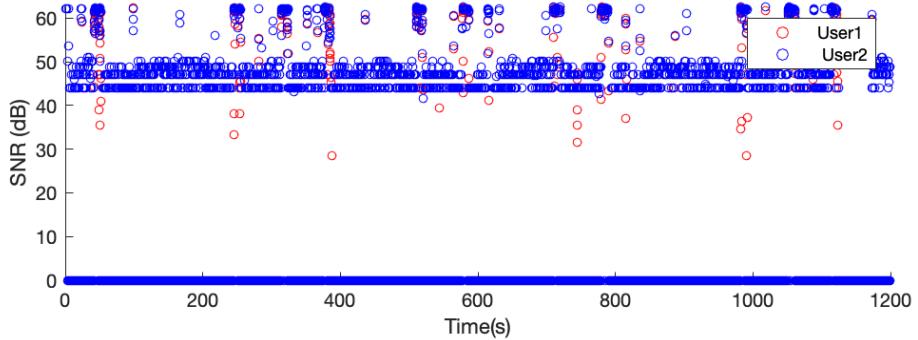
Table 7.2: Summary of Achieved Results: Comparison of Different Approaches Based on Room Sizes, Average (Avg) Throughput, and Standard Deviation (SD).

Approach	Room size [m^2]	Avg [Mbps]	SD [Mbps]
No IRS	10 × 10	124, 08	68, 4051
	15 × 15	112, 25	74, 7407
	20 × 20	98, 063	79, 2939
Random	10 × 10	131, 97	60, 4667
	15 × 15	117, 58	62, 8617
	20 × 20	109, 40	67, 1633
Optimal	10 × 10	144, 97	50, 5537
	15 × 15	125, 10	68, 5524
	20 × 20	115, 87	72, 2035
Oracle	10 × 10	147, 89	40, 4191
	15 × 15	129, 34	47, 5146
	20 × 20	118, 74	49, 0862
Best Path	10 × 10	148, 90	39, 1131
	15 × 15	131, 34	46, 5221
	20 × 20	120, 73	48, 9770

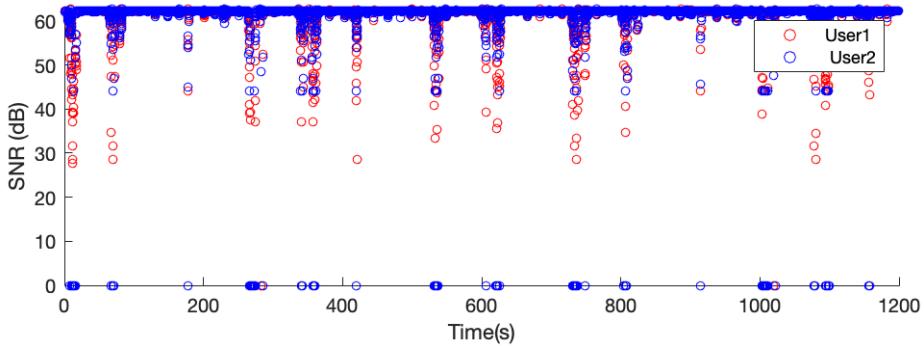
munication network, aiming to meet the high data-rate demands of future VR systems. In contrast, we explore the IRS's placement in a “*Random*” location to understand the impact of non-strategic IRS positioning. Additionally, an “*Oracle*” scenario is investigated, which entails determining the IRS's placement across all potential locations along the room walls on a 0.1 meter-sized grid. This exhaustive search is conducted for every conceivable HMD location, providing insights into the optimal positioning of the IRS without the constraints of real-time adaptability. Furthermore, the “*Best path*” metric evaluates the combined performance of direct AP-HMD and AP-IRS-HMD links. Here, the IRS's placement is determined using the “*Optimal*” approach, showcasing the effectiveness of strategic IRS configuration in enhancing the overall network performance. This comprehensive analysis enables a detailed understanding of the various factors that influence the efficiency of IRS-assisted networks, particularly in the context of supporting immersive VR applications.

To evaluate the effectiveness of various IRS configuration strategies, we conduct a detailed analysis, contrasting average throughput and its Standard Deviation (SD) against a benchmark throughput of 150 Mbps per HMD. This evaluation spans scenarios featuring both single and multiple HMD users navigating through a variety of environments, as outlined in *Table 7.2* and shown in *Fig. 7.5*. The focal point of this analysis is to discover the impact of IRS placement and resource allocation on network performance.

Our analysis reveals that the throughput reaches its peak in scenarios where the IRS is optimally positioned for each HMD, significantly surpassing setups that either lack an IRS or feature IRS resource allocation at random locations. Interestingly, while an IRS placed at an oracle-determined location occasionally achieves higher throughput, the AO algorithm demonstrates a remarkable ability to closely match this performance across the majority of



(a) 2 HMD users with no IRS in a mmWave VR network.



(b) 2 HMD users with IRS at an optimal location in a mmWave VR network.

Figure 7.5: SNR variability enhancements due to the utilization of IRS resources at locations optimized by the proposed AO approach.

HMD locations, despite operating in real-time. Furthermore, the combined average throughput from the optimal IRS path and the direct AP-HMD link exhibits a performance closely comparable to that achieved by the oracle, highlighting the effectiveness of strategic IRS placement and resource allocation (cf., Table 7.2). When analyzing SD, we observe that the performance variability of the network without IRS support is more pronounced across different environments, in contrast to scenarios with IRS assistance, where even randomly allocated IRS resources lead to more consistent throughput levels. The "Optimal" location-based IRS resource allocation determined through the AO approach, especially when combined with LoS communications, consistently offers stable throughput and minimizes SNR variability. This robust performance is maintained even in multi-user scenarios, as illustrated in Fig. 7.5. These findings underscore the critical importance of meticulously strategized IRS deployment and resource allocation in enhancing network throughput and stability, thereby enriching the immersive quality of VR experiences in diverse spatial configurations.

7.11 Conclusion

This chapter embarked on an exploratory journey beginning with a question of whether active Intelligent Reflecting Surfaces (IRS) are useful within millimeter-Wave (mmWave) frequencies, a frequency spectrum domain critical for high data-rate applications. Answering affirmatively, we established that IRS technology, irrespective of being passive or active, improves mmWave communication frameworks, with active IRS configurations offering superior performance enhancements. Following this initial finding, we ventured into a detailed examination of a use case involving passive IRS deployment within mmWave frequencies, specifically in the Virtual Reality (VR) domain. This shift in focus was motivated by the intent to fully understand the IRS applicability in overcoming the inherent challenges of mmWave technology, such as its sensitivity to environmental obstructions and the need for high data transmission capabilities. Our study conclusively found that IRS technology, in its versatile forms, holds significant promise for advancing the capabilities and reach of mmWave networks, particularly in enriching VR experiences.

To navigate the complexities of integrating IRS technology into mmWave-supported VR environments, we developed and implemented an Alternative Optimization (AO) algorithm. This algorithm was designed to dynamically allocate IRS resources, factoring in the spatial relationships between Head-Mounted Devices (HMDs), the Access Point (AP), and the IRS. The primary goal of our algorithm was to augment the immersive quality of VR applications by capitalizing on the IRS to improve communication coverage and enhance the fidelity of Virtual Experiences (VEs). The empirical evaluation of our approach demonstrated the significant impact of IRS-aided networks on expanding signal coverage and enriching the network's quality of service and user's quality of experiences, particularly in VR scenarios. These findings affirm the proposed algorithm's effectiveness and highlight the transformative role of IRS technology in advancing the capabilities of mmWave networks.

As we concluded our exploration of IRS applications in mmWave frequencies in this chapter, we also set the stage for future investigations into the realm of even higher frequencies, that is the TeraHertz (THz) frequencies. The next chapter is poised to study this advanced frequency band, exploring its potential to further elevate the performance and adaptability of IRS-assisted wireless networks. By venturing into THz frequencies, we aim to uncover innovative solutions for overcoming the limitations of current technologies and unlocking new possibilities for new types of applications other than immersive VR experiences and beyond. This forward-looking perspective emphasizes our commitment to pushing the boundaries of what's possible in the intersection of IRS technology, mmWave and THz frequencies, and immersive digital world.

Chapter 8

Energy Efficient THz Miniature UAV Networks

TERAHERTZ (THz) band communication, a key technology for sixth-Generation (6G) and beyond mobile networks, has the potential to enhance a wide range of promising applications. This chapter delves into the optimization of Energy Efficiency (EE) in miniature Unmanned Aerial Vehicle (UAV)-assisted Non-Orthogonal Multiple Access (NOMA) networks operating within the THz band. Specifically, it explores the deployment of UAVs in a cooperative THz NOMA network where the UAV, apart from receiving information, also harvests energy through Simultaneous Wireless Information and Power Transfer (SWIPT) mechanisms. This dual capability allows the UAV to relay data to a targeted destination node efficiently using the energy accumulated from the harvesting process. The inherent uncertainty of the THz channel necessitates a novel framework for UAV deployment and network resource allocation policy design. Addressing the inherent uncertainty of the THz channel, the study introduces an innovative framework tailored for the strategic deployment of miniature UAVs and the formulation of effective network resource allocation policies. A novel optimization challenge is presented, focusing on maximizing EE by fine-tuning the NOMA power allocation coefficients, the SWIPT Power-Splitting (PS) ratio, and the trajectory of the UAV. This complex problem is deconstructed into three manageable subproblems, and each is addressed through an alternating optimization strategy to find a solution. This study highlights the critical role of various parameters, including UAV mobility, NOMA power allocation strategies, and SWIPT PS ratios, and their impact on system performance aspects such as energy harvesting and service quality at the destination. The findings reveal the intertwined effects of these factors on the overall EE of the system. The proposed methodology underscores the significance of meticulous power management, the potential for energy harvesting, and the optimization of UAV mobility in enhancing the efficiency and sustainability of future wireless communication infrastructures.

This chapter is based on:

J. Jalali, A. Khalili, H. Tabassum, R. Berkvens, J. Famaey and W. Saad, "Energy-Efficient THz NOMA for SWIPT-aided Miniature UAV Networks", *IEEE Communications Letters*, Early Access, p. 1–5, Mar. 2024. <https://doi.org/10.1109/LCOMM.2024.3372471>

8.1 Introduction

MINIATURIZED Unmanned Aerial Vehicles (UAVs) have become a focal point of research and application due to their unparalleled ability to navigate through unexplored or complex terrains and their agility in tight spaces [257, 258]. These UAVs, with their compact form factor, can act as airborne Base Stations (BSs) or relays, significantly enhancing wireless network coverage. The predominance of Line of Sight (LoS) links between UAVs and ground users is a critical factor that facilitates high data rate transmissions, marking a significant advantage in utilizing UAVs for communication purposes [259].

However, the smaller scale of these UAVs introduces substantial challenges, particularly in terms of energy constraints. Unlike their larger counterparts, which can house more extensive battery systems and more sophisticated energy solutions, miniature UAVs must adopt exceedingly efficient energy management strategies to sustain operational viability. This necessity propels the relevance of Simultaneous Wireless Information and Power Transfer (SWIPT) in the realm of miniature UAV operations. SWIPT emerges as a pivotal innovation, not only elevating the rate of information exchange but also prolonging the operational lifespan of these UAVs through enhanced energy transfer efficiencies. The integration of SWIPT into the UAV network fabric significantly marks a significant leap forward in the Energy Efficiency (EE) frontier, offering a promising avenue to tackle the energy limitations inherent to miniature UAVs [260, 261].

The integration of SWIPT into UAV networks, despite facing challenges like the low energy conversion efficiency of contemporary energy harvesting technologies, offers targeted advantages in several key areas. These applications leverage the unique capabilities of miniature UAV, enhanced by SWIPT, to fulfill specific needs that outweigh the limitations imposed by energy conversion inefficiencies. The relevance and value of these applications are particularly pronounced in scenarios where traditional approaches may fall short. Here are some of the notable application contexts:

1. **Small-Scale and Short-Range Operations:** The compact nature and agility of miniature UAVs, combined with the capabilities of SWIPT, make them ideal for operations that require detailed attention within limited geographical extents. Tasks such as surveillance of critical infrastructure, environmental monitoring in sensitive or inaccessible regions, and providing communication links in complex urban or indoor environments are where these systems shine. Their ability to operate in confined or challenging areas, where larger UAVs or traditional infrastructure cannot reach, presents a significant advantage.
2. **Research and Experimental Applications:** In scenarios where the establishment of permanent infrastructure is impractical, economically unfeasible, or where existing networks have been compromised due to natural disasters or human-induced events, miniature UAV networks equipped with SWIPT technology offer a rapid, flexible solution. They can swiftly provide essential communication and monitoring capabilities, facilitating disaster response efforts, or extending services to remote and underserved populations temporarily.

3. Research and Experimental Applications: Miniature UAVs equipped with SWIPT technology offer significant potential for research and experimental applications. They provide a flexible and cost-effective platform for testing new technologies, sensors, and communication protocols in real-world scenarios. These systems can be used to gather data, validate theoretical models, and explore innovative uses in various fields, such as environmental science, urban planning, and emergency management, making them invaluable tools for scientific advancement and technological innovation.

While acknowledging the limitations imposed by the current state of energy harvesting technology, it is important to recognize that these limitations also drive innovation. The development of more efficient energy harvesting circuits and the integration of emerging technologies can gradually expand the scope of application for such systems. In the following paragraphs, we explore the possible synergy between miniature UAVs and emerging technologies.

Non-Orthogonal Multiple Access (NOMA) has emerged as a transformative approach to significantly enhance Spectral Efficiency (SE) within UAV networks. By allowing multiple user signals to occupy the same resource block simultaneously, NOMA ensures a more effective utilization of available network resources. This capability is particularly beneficial in UAV-assisted communications, where the dynamic nature of UAVs and the varying demands of users necessitate efficient resource management. The authors in [262] explored the effect of NOMA in a multiuser UAV network context, focusing on addressing the max-min fairness problem. This problem is analyzed taking into account various constraints, including the UAV's flight trajectory, total power and bandwidth limitations, as well as antenna beamwidth considerations, contributing valuable insights into achieving equitable resource distribution among users under the operational constraints inherent to UAV networks. Another innovative approach in resource allocation policy design is presented in [263], where a hybrid UAV-assisted network framework is proposed. This model uniquely combines Time Division Multiple Access (TDMA) utilized by the UAV with a terrestrial BS employing NOMA to serve ground users. The study emphasizes optimizing the sum data rate by carefully managing user scheduling, the UAV's flight trajectory, and precoding strategies at the BS, demonstrating the potential of hybrid access schemes in maximizing network throughput. Further extending the applicability of NOMA in UAV networks, the research outlined in [264] positions the UAV as a NOMA relay. This strategic deployment aims to extend the coverage area of the UAV network, confirming NOMA's ability to enhance connectivity and service reach in UAV-assisted communication systems. Moreover, the work presented in [265] addresses the challenge of minimizing delay in the context of millimeter-Wave (mmWave) and Multiple-Input Multiple-Output (MIMO)-NOMA resource allocation within UAV-assisted caching networks, indicating the critical importance of efficient resource allocation in reducing latency, particularly in networks where high data rate transmissions and caching capabilities are paramount.

Beyond the pursuit of high SE, EE has emerged as an equally crucial performance metric for the evolution of sixth-Generation (6G) mobile networks. The limited energy resources of UAVs and the ground users they serve necessitate innovative solutions like SWIPT to enhance network sustainability. As discussed earlier, SWIPT technology not only promises to extend the operational lifespan of these devices by improving battery life but also contributes significantly to the overall EE of the network [266, 9]. Research such as the study in [267] has highlighted the potential of UAV-SWIPT in Internet of Things (IoT) scenarios, focusing

on maximizing the minimum harvested energy of IoT devices through the joint optimization of UAV's transmission power and flight trajectory. The realm of secure transmission within UAV-assisted SWIPT systems has also been explored, as evidenced by [268], which aimed to maximize the secrecy rate for the intended Information Decoding (ID) receiver despite the presence of multiple eavesdroppers. Additionally, the application of SWIPT in UAV networks extends to enhancing network throughput, where UAVs harvest power wirelessly and utilize this power for retransmitting information to destination nodes, thereby creating an efficient cooperative link between the source and destination [269]. Such strategies underscore the dual role of UAVs in these networks, where they not only act as relay points but also as dynamic power sources. Further studies have delved into the integration of UAV-assisted cooperative communication with SWIPT, where the maximization of cooperative throughput was achieved by leveraging UAV's mobility and the power harvested [270]. Moreover, the intricate performance dynamics of downlink mmWave NOMA in UAV-assisted SWIPT systems have been analyzed, focusing on critical aspects such as security, reliability, and coverage [271]. This comprehensive examination reveals the multifaceted benefits of incorporating SWIPT in UAV networks, including enhanced EE, extended device lifespans, and improved network performance across various dimensions.

Furthermore, the exploration into the TeraHertz (THz) frequency band communication, ranging from 0.1 to 10 THz, underscores its potential to revolutionize future wireless communications. This band's capability for providing extensive bandwidth and supporting data transmission rates of Tbps, coupled with minimal latency, positions it as a decisive technology for the next generation of high-speed wireless networks [272]. Despite its promise, the application of THz technology within UAV communications remains a relatively untapped area, with a limited number of studies charting this novel territory [273, 274, 275].

Research endeavors have begun to shed light on the possibilities and challenges associated with incorporating THz frequencies in UAV-assisted wireless systems. For instance, the study in [273] outlines a framework for optimizing UAVs deployment strategies, power assignments, and bandwidth allocations within THz frequency bands. Concurrently, the investigation detailed in [274] presents a novel approach for enhancing THz downlink networks through a cooperative recharging-transmission strategy facilitated by wirelessly powered UAVs. Furthermore, the exploration in [275] ventures into the realm of NOMA-based UAV integration within THz networks, highlighting the potential synergies and performance enhancements that could be realized. Despite these advancements, a comprehensive analysis of NOMA's application in THz-enabled UAV networks is conspicuously absent. The alliance of NOMA principles with THz UAV communications could unlock new dimensions of spectral and energy efficiency, exploiting the vast bandwidths available in the THz band to accommodate a higher number of users and simultaneously transmit multiple signals more effectively. This gap in research presents a significant opportunity for further study, particularly in understanding how NOMA techniques can be tailored to leverage the unique properties of the THz band, thereby enhancing UAV network capabilities, improving throughput, and reducing latency in ultra-high-speed wireless communications.

Inspired by the aforementioned studies, we contemplate the use of a UAV-assisted cooperative system within the THz frequency band, aiming to improve the reliability of communication. In particular, a UAV-assisted cooperative network establishes a direct link between the source and destination nodes. This chapter focuses on the integration of SWIPT in minia-

ture UAV-assisted networks, addressing the unique challenges posed by their size and energy constraints. By optimizing energy use through SWIPT, we aim to maximize the operational capabilities of miniature UAVs, which is a significant departure from the strategies typically employed in larger UAV networks. The main contributions of this chapter are as follows:

- We study the EE performance analysis of UAV-assisted SWIPT-enabled cooperative NOMA network in the THz frequency band. The system architecture is carefully designed, wherein the source node transmits a composite signal that encapsulates data intended for both the UAV and the destination node, each differentiated by distinct NOMA power allocation coefficients. The signal received at the UAV is then split into two components using a Power-Splitting (PS) SWIPT mechanism. One portion of the signal is dedicated to Energy Harvesting (EH), while the other is used for Information Decoding (ID). The goal of maximizing EE in this framework encompasses a broad spectrum of considerations including the UAV's flight dynamics, the Quality of Service (QoS) requirements of the destination node, the PS ratio, and the distribution of NOMA power allocation coefficients.
- Subsequently, we introduce and formulate the problem of maximizing the network's EE. To navigate through the complexities of this optimization problem, we conceive an iterative algorithm specifically designed to deconstruct the primary problem into three more manageable sub-problems, each focusing on a distinct aspect: the PS ratio, UAV's flight trajectory, and NOMA power management strategies.
- These sub-problems are rigorously analyzed and proven to possess convex properties, enabling their resolution through established optimization techniques. The empirical results from this methodological approach underscore a substantial performance enhancement, revealing up to a 30.3% increase in EE. These findings eloquently emphasize the critical role of adept power management, efficient energy harvesting, and the adaptive mobility of the UAV in significantly elevating the cooperative EE performance of the NOMA-SWIPT network in the THz band.

Through this comprehensive analysis in this chapter, we not only elucidate the technical feasibility and efficiency gains achievable with such an integrated system but also lay the groundwork for future research directions. By highlighting the nuanced interplay between NOMA power allocation, SWIPT mechanisms, and UAV trajectory optimization, this study provides valuable insights into the design and operation of future high-frequency miniature UAV-assisted communication networks, aiming for unparalleled energy efficiency and service quality.

This chapter is organized as follows: Section 8.2 introduces the system model for a UAV-aided SWIPT-NOMA network and defines its performance metric. In Section 8.3, we formulate the EE maximization problem. Section 8.4 presents a two-stage solution to the EE problem in a miniature UAV network, with a discussion of computational complexity in Section 8.5. Section 8.6 contains the simulation results, and Section 8.7 outlines the applicability of our design in an IRS-aided network. Finally, Section 8.8 concludes the chapter.

Notations: Matrices and column vectors are denoted by boldface uppercase and lowercase letters, such as \mathbf{A} and \mathbf{a} . The Euclidean norm of vector \mathbf{a} is expressed as $|\mathbf{a}|$, while the

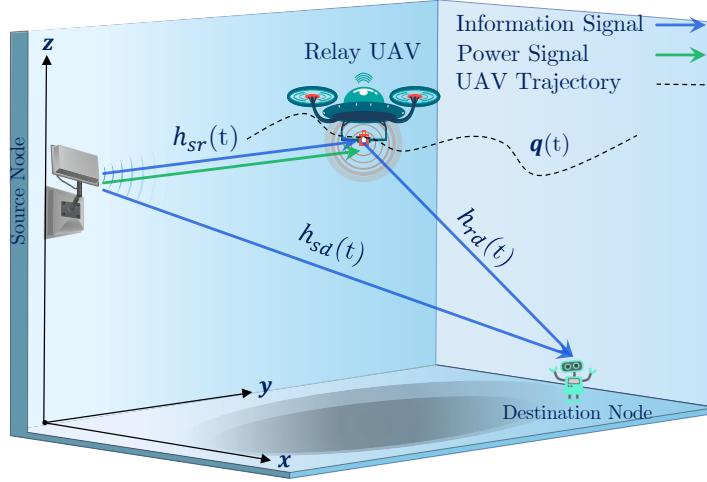


Figure 8.1: Miniature UAV-assisted cooperative THz NOMA-SWIPT network. The channel power gain between source-to-UAV and UAV-to-destination is denoted as $h_{sr}(t)$ and $h_{rd}(t)$, respectively. The miniature UAV acts as an energy-harvesting user in the first phase. In the second phase, the miniature UAV is a relay that uses the harvested energy from the previous phase.

magnitude of a complex number a is denoted by $|a|$. The transpose and Hermitian of a matrix are represented as $(\cdot)^T$ and $(\cdot)^H$, respectively. Additionally, $\text{Tr}(\mathbf{A})$ and $\text{rank}(\mathbf{A})$ define the trace and rank of matrix \mathbf{A} , respectively. The notation $\text{diag}(\mathbf{A})$ denotes a vector whose elements consist of the main diagonal elements of the matrix \mathbf{A} . If $\mathbf{A} \succeq 0$, then \mathbf{A} is a positive semidefinite matrix. $\mathbb{C}^{N \times M}$ refers to the space of $N \times M$ matrices with complex entries. The set containing elements a and b is represented as $\{a, b\}$. Matrix \mathbf{I}_M denotes an $M \times M$ identity matrix. $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ describes the distribution of a CSCG random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, where \sim indicates "with the distribution of."

8.2 System Model and Performance Metric

We consider a downlink transmission UAV-aided SWIPT-NOMA system, as shown in Fig. 8.1. The source node transmits information to two nodes, i.e., a UAV and a destination node. It is assumed that the UAV node has a better channel condition since it is closer to the source than the destination node. Hence, the UAV can act as an EH aerial relay to ensure the high targeted rate of the destination node. Frequency Division Duplex (FDD) mode with equal bandwidth of B and a 3D Cartesian coordinate system are considered where the source and destination nodes are placed at $\mathbf{s}(t) = [s_x(t), s_y(t), H_1]^T \in \mathbb{R}^{3 \times 1}$ and $\mathbf{d}(t) = [d_x(t), d_y(t), 0]^T \in \mathbb{R}^{3 \times 1}$, respectively, where $[\cdot]^T$ is the transpose operation. The destination node is static on the ground, while the UAV and source are at a fixed height above the ground. The instantaneous coordinates of the UAV are given by $\mathbf{q}(t) = [x(t), y(t), H_2]^T \in \mathbb{R}^{3 \times 1}$ at time $0 < t < T$. The first and final positions of the miniature UAV are represented by \mathbf{q}_s

and \mathbf{q}_e , respectively. We assume that the UAV's trajectory regularly varies over time, so the time range T is divided into N evenly separated time slots, i.e., $T = N\varpi$. In particular, ϖ represents the duration of each time slot and $\mathcal{N} \triangleq \{1, \dots, N\}$ is the set of time slots, where $\mathbf{q}[n]$ is the sampled trajectory. The constraints related to the position of the miniature UAV and its maximum speed can be written as:

$$\mathbf{q}[1] = \mathbf{q}_s, \quad (8.1a)$$

$$\mathbf{q}[N+1] = \mathbf{q}_e, \quad (8.1b)$$

$$\|\mathbf{q}[n+1] - \mathbf{q}[n]\| \leq \varpi V_{\max}, \forall n \in \mathcal{N}, \quad (8.1c)$$

where V_{\max} is the maximum flying speed of the miniature UAV. The channel power gain between source-to-UAV and UAV-to-destination are denoted by $h_{sr}[n]$ and $h_{rd}[n]$, respectively, which are assumed to follow the free-space path loss model and given by:

$$h_{sr}[n] = \frac{\beta_0}{\|\mathbf{q}[n] - \mathbf{s}[n]\|} e^{-\frac{\xi(f)}{2} \|\mathbf{q}[n] - \mathbf{s}[n]\|}, \forall n \in \mathcal{N}, \quad (8.2)$$

$$h_{rd}[n] = \frac{\beta_0}{\|\mathbf{q}[n] - \mathbf{d}[n]\|} e^{-\frac{\xi(f)}{2} \|\mathbf{q}[n] - \mathbf{d}[n]\|}, \forall n \in \mathcal{N}, \quad (8.3)$$

where the exponential term is the path loss caused by molecular absorption, in which $\xi(f)$ is a molecular absorption coefficient that is influenced by the operating frequency f and the concentration of water vapor and oxygen molecules [276]. To simplify the notation, we will henceforth denote $\xi(f)$ as ξ . Moreover, β_0 denotes the reference power gain and is equal to $c/4\pi f$, where c is the speed of light [274]. Finally, the channel power gain between the source-to-destination follows the same structure, as in (8.2) and (8.3), and is denoted by $h_{sd}[n]$ ¹.

Our cooperative system is studied in two phases. In the first phase, the UAV employs SWIPT to harvest energy and decode information from the source node while the destination node receives its respective data. In the second phase, the UAV acts as a decode-and-forward (DF) aerial relay to re-transmit the destination node's data by utilizing the harvested power of the first phase.

8.2.1 Phase One⁽¹⁾: Direct Transmission

In this phase, the source transmits the information to both the miniature UAV and destination node by exploiting power-domain NOMA, as shown in Fig. 8.2. Hence, the transmit signal is given by:

$$\mathbf{s}[n] = \sqrt{\alpha_1[n]} \mathbf{s}_1[n] + \sqrt{\alpha_2[n]} \mathbf{s}_2[n], \forall n \in \mathcal{N}, \quad (8.4)$$

where $\mathbf{s}_1[n]$ and $\mathbf{s}_2[n]$ are transmit symbols during each time slot and assumed to be independently Circularly Symmetric Complex Gaussian (CSCG) distributed with zero mean and unit variance. Moreover, $\sqrt{\alpha_1[n]}$ and $\sqrt{\alpha_2[n]}$ represent the NOMA power allocation coefficients

¹The instantaneous channel power gain between source-to-UAV and UAV-to-destination, denoted by $h_{sr}(t)$ and $h_{rd}(t)$, $0 < t < T$, are sampled at the rate δ_t to generate the discrete instantaneous channel power gain $h_{sr}[n]$ and $h_{rd}[n]$, $\forall n \in \mathcal{N}$.

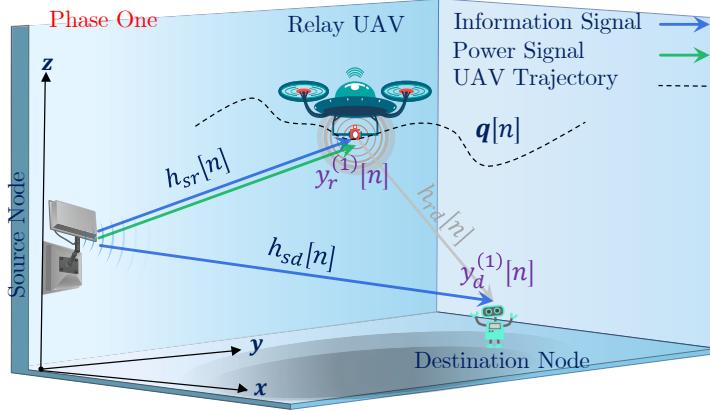


Figure 8.2: Phase One: Miniature UAV-assisted cooperative THz NOMA-SWIPT network. The gray link means an inactive link. The source transmits the information to both the miniature UAV and destination node by exploiting power-domain NOMA. The miniature UAV has not only its own data in this phase but also the destination node's data. The miniature UAV decodes the data intended for the destination node but relays it in the next phase.

in the n -th time slot, which need to satisfy the two following constraints:

$$\alpha_1[n] + \alpha_2[n] \leq P_{\text{peak}}, \forall n \in \mathcal{N}, \quad (8.5a)$$

$$\frac{1}{N} \sum_{n=1}^N \alpha_1[n] + \alpha_2[n] \leq P_{\text{max}}. \quad (8.5b)$$

Here P_{peak} refers to the peak power that can be transmitted by the source node. It denotes the maximum instantaneous power output that the source is capable of delivering at any given moment. On the other hand, P_{max} represents the maximum power constraint of the source node over a longer period, which could be a limitation of the total energy consumption of the source. This parameter is crucial in energy-efficient network designs as it ensures that the power usage by the source does not exceed a certain threshold, thereby optimizing the energy consumption over time. The received signal at the UAV can be expressed as:

$$y_r^{(1)}[n] = h_{sr}[n]s[n] + z_1^{(1)}[n], \forall n \in \mathcal{N}, \quad (8.6)$$

where $z_1^{(1)}[n] \sim \mathcal{N}(0, \sigma_1^2)$ is the received CSCG noise at the UAV node. By adopting a PS-SWIPT architecture, the received signal for ID and EH from the Radio Frequency (RF) source can be expressed as:

$$y_{\text{EH}}^{(1)}[n] = \sqrt{\rho[n]}(y_r^{(1)}[n]), \quad \forall n \in \mathcal{N}, \quad (8.7)$$

$$y_{\text{ID}}^{(1)}[n] = \sqrt{1-\rho[n]}(y_r^{(1)}[n]) + z_2^{(1)}[n], \forall n \in \mathcal{N}, \quad (8.8)$$

where $0 < \rho[n] < 1$ is the PS ratio, and $z_2[n] \sim \mathcal{N}(0, \sigma_2^2)$ is the additional noise caused by the ID receiver. The UAV node employs a successive interference cancellation (SIC) receiver

to decode the signals. In fact, the UAV first decodes the data of the destination node and then removes it from its received signal to obtain its own data in a successive manner. The received Signal-to-Interference-plus-Noise Ratio (SINR) at the UAV to detect $s_2[n]$ can be stated as:

$$\gamma_{d \leftarrow r}^{(1)}[n] = \frac{(1 - \rho[n])\alpha_2[n]|h_{sr}[n]|^2}{(1 - \rho[n])\alpha_1[n]|h_{sr}[n]|^2 + N}, \forall n \in \mathcal{N}, \quad (8.9)$$

where $N = (1 - \rho[n])\sigma_1^2[n] + \sigma_2^2[n]$ is the equivalent total noise. Then, the corresponding SINR to decode UAV's data can be described as:

$$\gamma_r^{(1)}[n] = \frac{(1 - \rho[n])\alpha_1[n]|h_{sr}[n]|^2}{N}, \forall n \in \mathcal{N}. \quad (8.10)$$

In our novel approach, the UAV is not just a relay in the communication system; it also acts as an independent user in Phase One. This means that, unlike typical systems where the UAV only relays data, in our model, the UAV has its own data to be processed during the first phase while simultaneously harvesting energy. In the next phase, this harvested energy is then used to retransmit data in a cooperation.

According to (8.6) and (8.7), the RF harvested power at the UAV by ignoring the noise power can be expressed as:

$$E[n] = \eta\rho[n]|h_{sr}[n]|^2\tau[n], \forall n \in \mathcal{N}, \quad (8.11)$$

where $\eta \in (0, 1]$ is the energy conversion efficiency and $\tau[n]$ is the transmission time fraction for the first phase during the n -th time slot. Besides, it is assumed that the transmission duration is the same for two phases, i.e., $\tau[n] = \frac{1}{2}$. Consequently, the transmit power at the UAV can be described by [270]:

$$P_t[n] = \frac{E[n]}{1 - \tau[n]}, \quad \forall n \in \mathcal{N}. \quad (8.12)$$

The received signal at the destination can be stated as:

$$y_d^{(1)}[n] = h_{sd}[n]s[n] + \nu_1^{(1)}[n], \quad \forall n \in \mathcal{N}, \quad (8.13)$$

where $\nu_1^{(1)}[n] \sim \mathcal{N}(0, \delta_1[n])$ is the received noise at the destination node in the first phase. The SINR at the destination node to decode its own data can be written as:

$$\gamma_d^{(1)}[n] = \frac{\alpha_2[n]|h_{sd}[n]|^2}{\alpha_1[n]|h_{sd}[n]|^2 + \delta_1^2[n]}, \forall n \in \mathcal{N}. \quad (8.14)$$

8.2.2 Phase Two⁽²⁾: Cooperative Transmission

In the first phase, the UAV harvests energy and decodes information from the source node, while the destination node simultaneously receives its data. This phase employs power-domain NOMA, where the source transmits the information to both the miniature UAV and the destination node. The transmitted signal, $s[n]$, is a combination of two symbols, $s_1[n]$ and $s_2[n]$, using power allocation coefficients $\alpha_1[n]$ and $\alpha_2[n]$.

The second phase is distinct in its purpose and functionality. Here, the miniature UAV acts as an aerial relay to re-transmit the destination node's data using the power harvested in

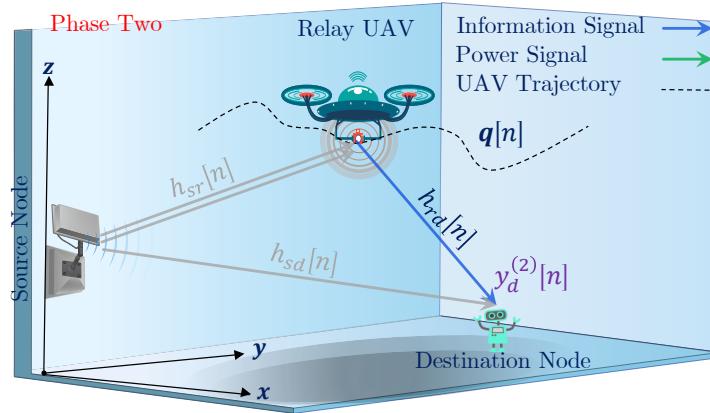


Figure 8.3: Phase Two: Miniature UAV-assisted cooperative THz NOMA-SWIPT network. The gray links mean inactive links. The miniature UAV relays the data intended to the destination node in this phase, using the harvested energy from the previous phase.

the first phase. This phase is designed to enhance the efficiency and reliability of the data transmission to the destination node, particularly in scenarios where the direct link from the source to the destination might be weak or unreliable. The reason the source does not continue transmitting signals directly to the destination in Phase Two is to capitalize on the enhanced transmission capabilities of the miniature UAV as a relay. Using the UAV as a relay in Phase Two allows for more efficient use of the harvested energy and a potentially stronger and more reliable link to the destination².

Beyond the power consumption associated with source node transmission, if the source node were to continue transmitting (using NOMA) in the second phase, it would inadvertently introduce interference, detracting from the network's overall performance due to the simultaneous transmissions from both the miniature UAV and the source node. Consequently, it is strategically advantageous to forego transmission from the source node in Phase Two, as the UAV already possesses the data intended for the destination node. This approach mitigates potential interference, thereby enhancing network efficiency and performance.

In this part of the process, which is illustrated in Figure 8.3, the miniature UAV makes use of the energy it has harvested to send the data to the intended recipient, the destination node. As a result, the formula for the signal that the destination node receives can be written as follows:

$$y_d^{(2)}[n] = \sqrt{P_t[n]} h_{rd}[n] s_2[n] + \nu_2^{(2)}[n], \quad \forall n \in \mathcal{N}, \quad (8.15)$$

where $\nu_2^{(2)}[n] \sim \mathcal{N}(0, \delta_2^2[n])$ is the received noise at the destination node. The corresponding

²In summary, the design choice in Phase Two of our protocol is based on the strategic use of the miniature UAV's capabilities as a relay to enhance data transmission to the destination node, making full use of the energy harvested in Phase One. This approach is tailored to maximize the efficiency and reliability of the network under the constraints and objectives of our proposed system model.

SINR reads as:

$$\gamma_d^{(2)}[n] = \frac{\eta\rho[n]|h_{sr}[n]|^2|h_{rd}[n]|^2}{\delta_2^2[n]}, \quad \forall n \in \mathcal{N}. \quad (8.16)$$

In the final step of the process, the destination node employs a technique known as Maximal Ratio Combining (MRC) to merge the signals transmitted during the two phases. This method optimizes the signal quality by (weighting and) summing multiple received signals based on their SINR. As a result of this integration, the combined corresponding SINR can be mathematically represented as follows:

$$\gamma_d^{\text{MRC}}[n] = \gamma_d^{(1)}[n] + \gamma_d^{(2)}[n], \quad \forall n \in \mathcal{N}. \quad (8.17)$$

In the context of our study, the miniature UAV is assigned a dual function that is quite distinctive. During the initial phase, it operates as a conventional user, involved in processing its own data. In the subsequent phase, it transitions into a relay, tasked with forwarding the data intended for the destination user. This dual functionality necessitates a unique approach to calculating the overall data transmission rate. We take into account not only the miniature UAV's personal data handling in the first phase but also its special role in transmitting the destination user's data during the second phase. To achieve this comprehensive evaluation, we incorporated the MRC technique. MRC is utilized to aggregate the data directed towards the destination user from both phases along with the UAV's own data from the initial phase. This method of data integration is crucial for our calculation of the total data rate, as detailed in the equation presented in (8.17). This approach allows for a more accurate representation of the system's performance, accounting for the UAV's multifaceted contribution to the network's data throughput.

8.2.3 Energy Efficiency Metric in a Cooperative Network

This concept of EE is particularly valuable for our miniature UAVs network for several reasons. Firstly, these UAV often operate on limited power resources; thus, optimizing energy usage extends their operational time and ensures they can cover more area or transmit more data before needing a recharge. Secondly, by maximizing the amount of data transmitted per unit of power consumed, we can achieve more efficient communication, which is crucial for applications requiring real-time or near-real-time data exchange. Finally, focusing on energy efficiency encourages the development of more sustainable UAV networks, reducing the environmental impact associated with their power consumption. We now define the network's EE as the ratio of the sum data rate to the total network's power consumption. That is:

$$\eta_{EE}[n] = \frac{R_{\text{sum}}[n]}{P_{\text{sum}}[n]}, \quad \forall n \in \mathcal{N}, \quad (8.18)$$

where the sum data rate and total consumed power of the network are given as follows:

$$R_{\text{sum}}[n] = \log_2(1 + \gamma_r^{(1)}[n]) + \log_2(1 + \gamma_d^{\text{MRC}}[n]), \quad \forall n \in \mathcal{N}, \quad (8.19)$$

$$P_{\text{sum}}[n] = \alpha_1[n] + \alpha_2[n] + P_c - P_t[n], \quad \forall n \in \mathcal{N}. \quad (8.20)$$

In the total power consumption formula, (8.20) P_c represents the constant power consumption at the source, and $P_t[n]$ represents the power used by the miniature UAV in Phase

Two, *the cooperative transmission phase*, which is indeed derived from the energy harvested in Phase One. Please note $P_t[n]$ reduces the source's energy consumption. The energy used in Phase Two by the UAV is the energy harvested from the source in Phase One, and therefore, it does not constitute an additional energy expenditure by the source. That is why we are subtracting it from the consumed power.

8.3 EE Maximization Problem Formulation

So far, we have studied how the two-phase system works. We explained the source node transmits data in Phase One to both the destination node and the miniature UAV. We learned although the miniature UAV has its own data in Phase One, it receives and decodes the data of the destination node while also harvesting energy. The miniature UAV then uses the harvested energy in Phase Two to re-transmit data to the destination node. In this section, we aim to maximize the EE of the miniature UAV network by optimizing the NOMA power allocation coefficients, PS ratio, and miniature UAV trajectory. Therefore, the EE optimization problem is formulated as:

$$P_1 : \max_{\rho[n], \alpha_1[n], \alpha_2[n], \mathbf{q}[n]} \sum_{n=1}^N \eta_{EE}[n] \quad (8.21a)$$

$$s.t. : \frac{1}{N} \sum_{n=1}^N P_t[n] \geq \frac{1}{N} \sum_{n=1}^N P[n], \quad (8.21b)$$

$$\gamma_{d \leftarrow r}^{(1)}[n] \geq \gamma_{\min}[n], \quad \forall n \in \mathcal{N}, \quad (8.21c)$$

$$\gamma_d^{\text{MRC}}[n] \geq \Gamma_{\min}[n], \quad \forall n \in \mathcal{N}, \quad (8.21d)$$

$$0 < \rho[n] < 1, \quad \forall n \in \mathcal{N}, \quad (8.21e)$$

$$P[n] \geq 0, \quad \forall n \in \mathcal{N}, \quad (8.21f)$$

$$\alpha_1[n] + \alpha_2[n] \leq P_{\text{peak}}, \quad \forall n \in \mathcal{N}, \quad (8.21g)$$

$$\frac{1}{N} \sum_{n=1}^N \alpha_1[n] + \alpha_2[n] \leq P_{\text{max}}, \quad (8.21h)$$

$$\mathbf{q}[1] = \mathbf{q}_s, \quad (8.21i)$$

$$\mathbf{q}[N+1] = \mathbf{q}_e, \quad (8.21j)$$

$$\|\mathbf{q}[n+1] - \mathbf{q}[n]\| \leq V_{\max} \delta, \quad \forall n \in \mathcal{N}. \quad (8.21k)$$

To ensure the effectiveness and reliability of the miniature UAV network, several constraints are put in place in the EE optimization problem (P_1) to manage power consumption, signal decoding, and SINR, alongside managing the UAV's position and power allocation strategy. Constraint (8.21b) stipulates that the power harvested by the UAV across all time slots must exceed a predetermined minimum level, denoted as $P[n] = P_{\text{EH}}$. This is crucial to guarantee that the UAV has sufficient power to operate effectively throughout its mission. Constraint (8.21c) requires that the UAV must be capable of decoding the destination node's

data with a signal quality of at least $\gamma_{\min}[n]$. This ensures the reliability of the data relay process from the UAV to the destination node. Constraint (8.21d) demands that the SINR at the destination node must surpass a specific threshold, $\Gamma_{\min}[n]$. This criterion is vital for ensuring that the destination node can accurately receive and interpret the signals sent from the UAV, maintaining the integrity of the communication link. Constraint (8.21e) and (8.21f) illustrate the control over the PS ratio and validate the transmitted power's feasibility from the miniature UAV. These constraints help optimize the UAV's energy use and ensure its transmissions are within operational limits. Constraints (8.21g) and (8.21h) regulate the NOMA power allocation coefficients. This is key for managing how power is distributed among different communications to maximize efficiency and reduce interference. Finally, constraints (8.21i), (8.21j), and (8.21k) govern the UAV's positioning and mobility, ensuring it operates within a designated area and adheres to flight dynamics and safety regulations.

The observation that a UAV cannot transmit more data to its destination than it has previously received, can be captured by the relationship $\gamma_{d \leftarrow r}^{(1)}[n] \leq \gamma_d^{(2)}[n]$. This principle forms what is known as the relay constraint. Addressing this constraint directly within our primary optimization problem, denoted as P_1 , significantly complicates the solution process. This complexity arises from the involvement of product terms of two THz channel gains, which are influenced by the UAV's trajectory and the NOMA coefficients. To circumvent this issue, we employ the following approach. The optimization problem P_1 integrates two distinct constraints aimed at ensuring QoS of the miniature UAV cooperative network:

$$\gamma_{d \leftarrow r}^{(1)}[n] \geq \gamma_{\min}[n], \quad \forall n \in \mathcal{N}, \quad (8.22)$$

$$\gamma_d^{\text{MRC}}[n] \geq \Gamma_{\min}[n], \quad \forall n \in \mathcal{N}. \quad (8.23)$$

Through a mathematical manipulation, we implicitly account for the relationship $\gamma_{d \leftarrow r}^{(1)}[n] \leq \gamma_d^{(2)}[n]$ by ensuring that $\Gamma_{\min}[n] \geq \gamma_{\min}[n]$. This approach allows us to navigate the complexities associated with directly addressing the relay constraint in our optimization framework.

8.4 A Two-Stage Solution to EE Problem

The difficulty with problem (P_1) stems from its non-convex nature, which arises from the coupling interactions between the variables to be optimized. This complexity poses significant challenges to the straightforward application of traditional optimization techniques. In particular, the objective function of (P_1) is characterized by a sum of ratios, a format that complicates the use of the Dinkelbach method. This method is commonly employed for solving problems involving ratios but struggles with scenarios involving sum-of-ratios, as highlighted in the referenced literature [3, 209].

To overcome these hurdles, a two-stage algorithm is proposed. This algorithm is designed to decouple the intertwined optimization variables, allowing for their independent optimization. The key innovation here is the introduction of a new strategy aimed at transforming the sum-of-ratios problem into a subtractive form. By doing so, the complex fractional non-linear problems are broken down into more manageable parts. Specifically, this involves separating

the numerators and denominators of these fractions, thereby simplifying the optimization process.

The motivation behind the two-stage optimization approach in this chapter is primarily due to the nature of the EE problem we are addressing. The problem (P_1) is non-convex and NP-hard due to the coupling of optimization variables, rendering conventional solutions like the Dinkelbach method ineffective. Thus, the proposed two-stage solution facilitates the independent optimization of each variable, offering a pragmatic pathway to tackle the intricacies of (P_1). In the first stage of our two-stage approach, we focus on optimizing the PS ratio and the UAV's trajectory, keeping the NOMA power allocation coefficients fixed. In the second stage, we then optimize the NOMA power allocation coefficients.

One might ask the question of whether a two-stage optimization causes performance degradation. While it is true that a two-stage optimization might lead to certain performance losses compared to a holistic approach, it is necessary in this context due to the complexity and non-convexity of the problem. The decoupling of variables in this manner simplifies the optimization process and makes it more tractable. There certainly are various methods, such as other convex optimization algorithms or AI algorithms, that could potentially address this problem. However, these methods might also face challenges with the complexity and non-convex nature of the problem. AI algorithms, like deep learning models or reinforcement learning strategies, could provide alternative solutions, but they often require extensive training data and computational resources. Moreover, the integration of non-convex and non-linear constraints within AI models presents another layer of complexity. The computational cost required for training sophisticated AI models is another critical consideration. AI-based approaches may not always guarantee convergence to the optimal solution, especially in non-convex settings. Deep learning models, for instance, often require substantial computational power and storage, necessitating access to high-performance computing systems — resources that are usually scarce or entirely unavailable. This can further make AI-driven solutions less accessible, especially in scenarios where no training data is initially available. Therefore, optimization-driven solutions can be preferred depending on the application at hand.

Our two-stage solution is a deliberate effort to balance the need for effective problem resolution with the practical limitations of miniature UAV resources. This strategic approach enables us to manage the complexity of the optimization problem effectively while achieving a balance between performance and computational tractability.

8.4.1 Stage-one: Optimizing PS ratio and Miniature UAV trajectory

In this phase of the optimization process, we focus on iteratively designing the PS ratio and the trajectory of the miniature UAV, while keeping the NOMA power allocation coefficients constant. However, the challenge arises from the sum data rate function's non-convex nature, attributed to the interaction between the PS ratio and the UAV's trajectory variables. This coupling complicates direct optimization.

To address this, the optimization problem is approached in an iterative manner. This method allows for a step-by-step refinement of both the PS ratio and the UAV's trajectory, gradually

moving towards an optimal solution despite the initial complexity.

However, before iterative optimization can effectively begin, a critical step must be undertaken: transforming the non-linear fractional objective function into a more manageable subtractive form. This transformation is crucial as it simplifies the problem, making the subsequent iterative optimization process feasible. The method for this transformation is grounded in a theorem from the study by Jong et al. [154]. This theorem provides a mathematical basis for decomposing the fractional objective function into a difference of two terms, thereby detangling the complex relationship between the numerator and the denominator of the fractions involved. This approach of transforming and then iteratively optimizing represents a workaround to the non-convexity and coupling issues, enabling the efficient tackling of what would otherwise be an intractable optimization problem.

Theorem: Let suppose that $\rho^*[n]$ and $\mathbf{q}^*[n]$ are the optimal solutions to the problem (P_1) . Then, the following optimization problem can provide an optimal solution in the existence of two vectors, namely, $\boldsymbol{\lambda} = [\lambda_1^*, \dots, \lambda_N^*]^T$ and $\boldsymbol{\psi} = [\psi_1^*, \dots, \psi_N^*]^T$ as follows:

$$\max_{\rho[n], \mathbf{q}[n]} \sum_{n=1}^N \lambda_n^* [R_{\text{sum}}[n] - \psi_n^*(P_{\text{sum}}[n])]. \quad (8.24)$$

Furthermore, $\rho^*[n]$ and $\mathbf{q}^*[n]$ meet these two following equations:

$$R_{\text{sum}}^*[n] - \psi_n^*(P_{\text{sum}}[n]) = 0, \forall n \in \mathcal{N}, \quad (8.25)$$

$$1 - \lambda_n^*(P_{\text{sum}}[n]) = 0, \quad \forall n \in \mathcal{N}. \quad (8.26)$$

Proof 8 Please refer to [154]. ■

Specifically, the equivalent subtractive form in (8.24) with the additional parameters $\{\boldsymbol{\lambda}^*, \boldsymbol{\psi}^*\}$ has the same optimal solution as (P_1) for given $\alpha_1[n]$ and $\alpha_2[n]$. In particular, the problem (8.24) can be solved iteratively with a two-layer approach, i.e., inner and outer layers. In the inner layer, (8.24) is solved under given $\boldsymbol{\lambda}$ and $\boldsymbol{\psi}$. Then, the two equations (8.25) and (8.26) are updated in the outer layer to obtain $\{\boldsymbol{\lambda}^*, \boldsymbol{\psi}^*\}$.

8.4.1.1 Inner-layer Problem

In the inner layer of our optimization strategy, addressing the non-convex nature of the problem requires a segmented approach to manage its complexity effectively. Initially, the focus is on optimizing the PS ratio, under the assumption that the trajectory of the miniature UAV and the NOMA power allocation coefficients are predetermined and fixed. This step involves formulating a specific optimization problem for the PS ratio, carefully designed to explore the best possible settings for energy harvesting and data transmission efficiency, given the constraints of the UAV's current flight path and the allocated power coefficients for NOMA communications. The formulation of the optimization problem for the PS ratio can be given as follows:

$$P_2 : \max_{\rho[n]} \sum_{n=1}^N \lambda_n^* [R_{\text{sum}}[n] - \psi_n^*(P_{\text{sum}}[n])] \quad (8.27a)$$

s.t. : (8.21b) – (8.21e).

Upon focusing on the PS ratio, $\rho[n]$, within the optimization framework, problem (P_2) is revealed to possess a convex structure with respect to this variable. This characteristic significantly simplifies the problem, making it amenable to efficient solution techniques commonly used in convex optimization. Such efficiency arises from the convex problem's property, where any local minimum is also a global minimum, ensuring that optimal solutions for $\rho[n]$ can be determined straightforwardly and reliably.

Following the determination of the optimal PS ratio, attention shifts to the miniature UAV's trajectory optimization. This subsequent step maintains the previously optimized PS ratio settings, integrating them into the conditions under which the UAV's flight path is refined. By doing so, the approach systematically decouples the coupling between the PS ratio and the UAV's trajectory, allowing for an iterative optimization process that sequentially finds the best configuration for each set of variables under the framework of the given optimization problem. This methodical procedure ensures that both energy efficiency and communication. Therefore, we optimize the trajectory under the optimal PS ratio as follows:

$$P_3 : \max_{\mathbf{q}[n]} \sum_{n=1}^N \lambda_n^* [R_{\text{sum}}[n] - \psi_n^*(P_{\text{sum}}[n])] \quad (8.28a)$$

$$\text{s.t.} : \sum_{n=1}^N \frac{\eta \rho[n] \beta_0^2 e^{-\xi(\|\mathbf{q}[n] - \mathbf{s}[n]\|)}}{\|\mathbf{q}[n] - \mathbf{s}[n]\|^2} \geq \sum_{n=1}^N P[n], \quad (8.28b)$$

$$\frac{\alpha_2[n]}{\alpha_1[n] + x \|\mathbf{q}[n] - \mathbf{s}[n]\|^2 e^{\xi(\|\mathbf{q}[n] - \mathbf{s}[n]\|)}} \geq \gamma_{\min}[n], \quad \forall n \in \mathcal{N}, \quad (8.28c)$$

$$\frac{\alpha_2[n] |h_{sd}[n]|^2}{\alpha_1[n] |h_{sd}[n]|^2 + \delta_1^2[n]} + \frac{\eta \rho[n] \beta_0^4}{\delta_2^2[n]} \cdot \frac{e^{-\xi(\|\mathbf{q}[n] - \mathbf{s}[n]\| + \|\mathbf{q}[n] - \mathbf{d}[n]\|)}}{\|\mathbf{q}[n] - \mathbf{s}[n]\|^2 \|\mathbf{q}[n] - \mathbf{d}[n]\|^2} \geq \Gamma_{\min}[n], \quad \forall n \in \mathcal{N}, \quad (8.28d)$$

$$(8.21f), (8.21i) – (8.21k),$$

where

$$x = \frac{N}{1 - \rho[n] \beta_0^2}. \quad (8.29)$$

When addressing the trajectory optimization in problem (P_3) , it is acknowledged that this problem maintains a non-convex nature, which poses significant challenges for direct optimization techniques. To navigate this issue, a strategic transformation of problem (P_3) is employed, focusing on reformulating it into an equivalent form that is more tractable for

optimization. This involves the introduction of slack optimization variables.

These slack variables act as a mathematical tool that simplifies the problem by breaking down complex non-linear relationships into more manageable components. By doing so, constraints and objectives that were previously non-convex can often be redefined in ways that align better with convex optimization methods, or at least become amenable to efficient approximation techniques. This transformation does not alter the essence of the problem but reshapes it into a form where advanced optimization algorithms can be applied more effectively, enhancing the feasibility of finding an optimal or near-optimal solution to the UAV's trajectory optimization challenge under the given constraints. Hence, we transform the problem (P_3) into its equivalent form by introducing slack optimization variables as follows:

$$P_4 : \max_{q[n], v[n], t[n], a[n], b[n]} \sum_{n=1}^N \lambda_n^* [R_{\text{sum}}[n] - \psi_n^*(P_{\text{sum}}[n])] \quad (8.30a)$$

$$\text{s.t. : } \sum_{n=1}^N \frac{\eta \rho[n] \beta_0^2}{e^{a[n]}} \geq \sum_{n=1}^N P[n], \quad (8.30b)$$

$$\frac{\alpha_2[n]}{\alpha_1[n] + x e^{a[n]}} \geq \gamma_{\min}[n], \quad \forall n \in \mathcal{N}, \quad (8.30c)$$

$$\frac{\alpha_2[n] |h_{sd}[n]|^2}{\alpha_1[n] |h_{sd}[n]|^2 + \delta_1^2[n]} + \frac{\eta \rho[n] \beta_0^4}{\delta_2^2[n] e^{a[n] + b[n]}} \geq \Gamma_{\min}[n], \quad \forall n \in \mathcal{N}, \quad (8.30d)$$

$$v[n] \leq \frac{\|q[n] - s[n]\|^2}{e^{-\xi \|q[n] - s[n]\|}}, \quad \forall n \in \mathcal{N}, \quad (8.30e)$$

$$t[n] \leq \frac{\|q[n] - d[n]\|^2}{e^{-\xi \|q[n] - d[n]\|}}, \quad \forall n \in \mathcal{N}, \quad (8.30f)$$

$$v[n] \leq e^{a[n]}, \quad \forall n \in \mathcal{N}, \quad (8.30g)$$

$$t[n] \leq e^{b[n]}, \quad \forall n \in \mathcal{N}, \quad (8.30h)$$

$$(8.21i) - (8.21k), (8.21f),$$

where

$$\begin{aligned} R_{\text{sum}}[n] = & \log_2 \left(1 + \left(\frac{(1 - \rho[n]) \alpha_1[n] \beta_0^2}{N} \cdot \frac{1}{e^{a[n]}} \right) \right) \\ & + \log_2 \left(1 + \gamma_d^{(1)}[n] + \left(\frac{\eta \rho[n] \beta_0^4}{\delta_2^2[n]} \cdot \frac{1}{e^{a[n] + b[n]}} \right) \right), \quad \forall n \in \mathcal{N}. \end{aligned} \quad (8.31)$$

With the transformation of problem (P_3) through the introduction of slack optimization variables, the resulting objective function and constraints are recast as convex functions. However, despite this transformation into convexity, the problem, now referred to as (P_4) , remains intractable due to its complexity and the computational difficulty in directly solving

it. To effectively approach this challenge, the method of Successive Convex Approximation (SCA) is employed, which is particularly adept at handling such scenarios.

SCA is a powerful technique that iteratively refines approximations of the original problem by employing convex over- or under-estimations of the non-convex parts, allowing for the application of convex optimization methods at each step. Specifically, for (P₄), SCA-based first-order Taylor expansions are utilized to create these approximations. These expansions provide a way to linearize the non-linear parts of the objective function and constraints, essentially converting them into a series of linear problems that can be more easily solved³. The first-order lower bounds are given by:

$$e^{a[n]} \geq e^{a^{(k)}[n]} \left(1 + e^{a[n]} - e^{a^{(k)}[n]} \right) \triangleq \tilde{e}^{a[n]}, \forall n \in \mathcal{N}, \quad (8.32)$$

$$e^{b[n]} \geq e^{b^{(k)}[n]} \left(1 + e^{b[n]} - e^{b^{(k)}[n]} \right) \triangleq \tilde{e}^{b[n]}, \forall n \in \mathcal{N}, \quad (8.33)$$

$$\begin{aligned} \frac{\|\mathbf{q}[n] - \mathbf{s}[n]\|^2}{e^{-\xi\|\mathbf{q}[n] - \mathbf{s}[n]\|}} &\geq \frac{\|\mathbf{q}^{(k)}[n] - \mathbf{s}[n]\|^2}{e^{-\xi\|\mathbf{q}^{(k)}[n] - \mathbf{s}[n]\|}} + 2 \frac{(\mathbf{q}^{(k)}[n] - \mathbf{s}[n])^T}{e^{-\xi\|\mathbf{q}^{(k)}[n] - \mathbf{s}[n]\|}}. \\ \frac{(\mathbf{q}[n] - \mathbf{q}^{(k)}[n])}{e^{-\xi\|\mathbf{q}^{(k)}[n] - \mathbf{s}[n]\|}} &\triangleq \frac{\|\tilde{\mathbf{q}}[n] - \mathbf{s}[n]\|^2}{e^{-\xi\|\mathbf{q}[n] - \mathbf{q}^{(k)}[n]\|}}, \quad \forall n \in \mathcal{N}, \end{aligned} \quad (8.34)$$

$$\begin{aligned} \frac{\|\mathbf{q}[n] - \mathbf{d}[n]\|^2}{e^{-\xi\|\mathbf{q}[n] - \mathbf{d}[n]\|^2}} &\geq \frac{\|\mathbf{q}^{(k)}[n] - \mathbf{d}[n]\|^2}{e^{-\xi\|\mathbf{q}^{(k)}[n] - \mathbf{d}[n]\|}} + 2 \frac{(\mathbf{q}^{(k)}[n] - \mathbf{d}[n])^T}{e^{-\xi\|\mathbf{q}^{(k)}[n] - \mathbf{d}[n]\|}}. \\ \frac{(\mathbf{q}[n] - \mathbf{q}^{(k)}[n])}{e^{-\xi\|\mathbf{q}^{(k)}[n] - \mathbf{d}[n]\|}} &\triangleq \frac{\|\tilde{\mathbf{q}}[n] - \mathbf{d}[n]\|^2}{e^{-\xi\|\mathbf{q}[n] - \mathbf{q}^{(k)}[n]\|^2}}, \quad \forall n \in \mathcal{N}, \end{aligned} \quad (8.35)$$

where

$$e^{a^{(k)}[n]} = \|\mathbf{q}^{(k)}[n] - \mathbf{s}[n]\|^2 \cdot e^{\xi\|\mathbf{q}^{(k)}[n] - \mathbf{d}[n]\|}, \quad (8.36)$$

$$e^{b^{(k)}[n]} = \|\mathbf{q}^{(k)}[n] - \mathbf{d}[n]\|^2 \cdot e^{\xi\|\mathbf{q}^{(k)}[n] - \mathbf{d}[n]\|}. \quad (8.37)$$

$e^{a^{(k)}[n]}$ and $e^{b^{(k)}[n]}$ express the Taylor points at iteration k . According to the above transformation, the problem (P₄) can be approximated as:

³By using the first-order Taylor expansions, first-order lower bounds for the non-convex parts are generated. These bounds are crucial as they provide a convex underestimation of the original functions, maintaining the feasibility and optimality conditions within a controlled approximation error. Through iterative updates based on these linear approximations, the algorithm converges towards the optimal solution of (P₄) by solving a sequence of convex optimization problems, each drawing the solution closer to the optimal point of the original non-convex problem. This method ensures a pragmatic and efficient pathway to optimize the UAV's trajectory and other related variables within the complex operational constraints of the network.

$$P_5 : \max_{q[n], v[n], t[n], a[n], b[n]} \sum_{n=1}^N \lambda_n^* [\tilde{R}_{\text{sum}}[n] - \psi_n^*(P_{\text{sum}}[n])] \quad (8.38a)$$

$$\text{s.t. : } \sum_{n=1}^N \frac{\eta \rho[n] \beta_0^2}{\tilde{e}^a[n]} \geq \sum_{n=1}^N P[n], \quad (8.38b)$$

$$\frac{\alpha_2[n]}{\alpha_1[n] + x \tilde{e}^a[n]} \geq \gamma_{\min}[n], \quad \forall n \in \mathcal{N}, \quad (8.38c)$$

$$\begin{aligned} & \frac{\alpha_2[n] |h_{sd}[n]|^2}{\alpha_1[n] |h_{sd}[n]|^2 + \delta_1^2[n]} \\ & + \frac{\eta \rho[n] \beta_0^4}{\delta_2^2[n] \tilde{e}^{a[n]+b[n]}} \geq \gamma_{\min}[n], \quad \forall n \in \mathcal{N}, \end{aligned} \quad (8.38d)$$

$$v[n] \leq \frac{\|\tilde{q}[n] - s[n]\|^2}{e^{-\xi \|\tilde{q}[n] - s[n]\|}}, \quad \forall n \in \mathcal{N}, \quad (8.38e)$$

$$t[n] \leq \frac{\|\tilde{q}[n] - d[n]\|^2}{e^{-\xi \|\tilde{q}[n] - d[n]\|}}, \quad \forall n \in \mathcal{N}, \quad (8.38f)$$

$$v[n] \leq \tilde{e}^a[n], \quad \forall n \in \mathcal{N}, \quad (8.38g)$$

$$t[n] \leq \tilde{e}^b[n], \quad \forall n \in \mathcal{N}, \quad (8.38h)$$

$$(8.21f), (8.21i) - (8.21k),$$

where

$$\tilde{R}_{\text{sum}}[n] = R_{\text{sum}}[n] \Big|_{\substack{e^a[n] = \tilde{e}^a[n], e^b[n] = \tilde{e}^b[n}}} , \quad \forall n \in \mathcal{N}. \quad (8.39)$$

The problem (P_5) can be solved at iteration k by employing convex optimization solvers, e.g., CVX [3], a widely recognized tool for solving convex problems. We denote each solution at k -th iteration as $F^{(k)}$. The proposed SCA iterative methodology is incorporated into the principal architectural design of our algorithm, as will become evident later.

8.4.1.2 Outer-layer Problem

In this phase of the optimization process, the damped Newton method comes into play as a sophisticated technique to find the optimal values of $\{\lambda, \psi\}$. This method is well-regarded for its efficiency in handling non-linear equations by iteratively refining guesses until convergence to the solution. To facilitate this, two functions $\phi_n(\psi_n)$ and $\phi_{N+j}(\lambda_j)$ are defined as follows:

$$\phi_n(\psi_n) = R_{\text{sum}}^*[n] - \psi_n^*(P_{\text{sum}}[n]), \quad (8.40)$$

$$\phi_{N+j}(\lambda_j) = 1 - \lambda_j^*(P_{\text{sum}}[j]), j \in \{1, \dots, N\}, \quad (8.41)$$

where $\phi_n(\psi_n)$ represents the difference between the optimal sum data rate $R_{\text{sum}}^*[n]$ and activated total power consumption $P_{\text{sum}}[n]$ whereas $\phi_{N+j}(\lambda_j)$ quantifies how far total power consumption $P_{\text{sum}}[j]$ is from unity. It is demonstrated in [9] that the optimal solution

$\{\boldsymbol{\lambda}^*, \boldsymbol{\psi}^*\}$ is found if and only if $\phi(\boldsymbol{\lambda}, \boldsymbol{\psi}) = [\phi_1, \phi_2, \dots, \phi_{2N}]^T = 0$. This condition effectively means that the optimal solutions $\{\boldsymbol{\lambda}^*, \boldsymbol{\psi}^*\}$ are those that equilibrate the specified functions, indicating a balance between the sum data rate and power consumption across the system. Accordingly, the updated value of the $\boldsymbol{\lambda}^{i+1}$ and $\boldsymbol{\psi}^{i+1}$ in the iteration i can be obtained by:

$$\boldsymbol{\lambda}^{i+1} = \boldsymbol{\lambda}^i + \zeta^i \mathbf{w}_{N+1:2N}^i, \quad (8.42)$$

$$\boldsymbol{\psi}^{i+1} = \boldsymbol{\psi}^i + \zeta^i \mathbf{w}_{1:N}^i, \quad (8.43)$$

where $\mathbf{w} = [\dot{\phi}(\boldsymbol{\lambda}, \boldsymbol{\psi})]^{-1} \phi(\boldsymbol{\lambda}, \boldsymbol{\psi})$, and $\dot{\phi}(\boldsymbol{\lambda}, \boldsymbol{\psi})$ is the Jacobian matrix of $\phi(\boldsymbol{\lambda}, \boldsymbol{\psi})$. This iterative procedure for updating $\boldsymbol{\lambda}$ and $\boldsymbol{\psi}$ at each iteration is meticulously crafted to edge closer to this equilibrium.

Moreover, ζ^i is the largest value of Ξ^m satisfying

$$\|\phi(\boldsymbol{\lambda}^i + \Xi^m \mathbf{w}_{N+1:2N}^i, \boldsymbol{\psi}^i + \Xi^m \mathbf{w}_{1:N}^i)\| \leq (1 - \varepsilon \Xi^m) \|\phi(\boldsymbol{\lambda}, \boldsymbol{\psi})\|, \quad (8.44)$$

where $m \in \{1, 2, \dots\}$, $\Xi^m \in (0, 1)$, and $\varepsilon \in (0, 1)$. By employing the damped Newton method, adjustments are made based on the current estimates and the method's insights into the problem's curvature, gradually refining these parameters. This iterative adjustment ensures a precise and convergent path towards identifying the optimal set of $\{\boldsymbol{\lambda}^*, \boldsymbol{\psi}^*\}$, marking a critical step in optimizing the network's overall efficiency and effectiveness.

8.4.2 Stage-two: Optimizing power coefficients

In the second stage of the optimization process, attention shifts towards the design of power allocation coefficients, the last component for enhancing the network's performance, particularly in terms of EE in this study. The groundwork for this new framework is laid by formulating the problem as a sum-fraction problem, which will be adopted later to design power allocation coefficients. Consider the sum-fraction problem formulated as follows:

$$\min_{\boldsymbol{\Omega} \in C} \sum_{j=1}^J \frac{B_j(\boldsymbol{\Omega})}{A_j(\boldsymbol{\Omega})}, \quad (8.45)$$

where J is the maximum number of fractional terms, and $\boldsymbol{\Omega}$ is the optimization variable vector with the domain of C . $A_j(\boldsymbol{\Omega})$ and $B_j(\boldsymbol{\Omega})$ are nominator and denominator of the n -th fractional term with positive values. The optimization problem (8.45), initially presented as a sum-fraction problem, is ingeniously restructured into an equivalent form that facilitates a more straightforward approach to finding the optimal solution. In the following, it is demonstrated that (8.45) has an equivalent form, which is:

$$\min_{\boldsymbol{\Omega} \in C, \iota_j > 0} \sum_{j=1}^J \iota_j B_j^2(\boldsymbol{\Omega}) + \sum_{j=1}^J \frac{1}{4\iota_j} \frac{1}{A_j^2(\boldsymbol{\Omega})}. \quad (8.46)$$

In fact, the solution to both (8.45) and (8.46) is the same. The equivalent form transforms the problem into minimizing the sum of two distinct series. The first series scales the square of the denominator, $B_j^2(\boldsymbol{\Omega})$, by a factor ι_j , while the second series inversely scales the reciprocal of the squared inverse of the numerator, $A_j^2(\boldsymbol{\Omega})$, by $(4\iota_j)^{-1}$. This transformation introduces an auxiliary variable, $\iota_j > 0$, for each fractional term, which plays a critical role in

bridging the original sum-fraction formulation to this new convex format. It is noteworthy that if $A_j(\boldsymbol{\Omega})$ is a concave function and $B_j(\boldsymbol{\Omega})$ is a convex one, then the problem (8.46) is convex for the given ι_j . Based on the above analysis, the convex problem (8.46) is solved for a given $\iota_j = \frac{1}{2A_j(\boldsymbol{\Omega})B_j(\boldsymbol{\Omega})}$, and then the value of ι_j will be updated in the next iteration. Consequently, with a specified PS ratio and predefined miniature UAV trajectory, problem (P_1) can be represented in the following equivalent manner:

$$P_6 : \min_{\alpha_1[n], \alpha_2[n], \iota[n] > 0} \sum_{n=1}^N \iota[n] P_{\text{sum}}^2[n] + \sum_{n=1}^N \frac{1}{4\iota[n]} \frac{1}{R_{\text{sum}}^2[n]} \quad (8.47a)$$

$$\text{s.t. : } \frac{\alpha_2[n]|h_{sr}[n]|^2}{\gamma_{\min}[n]} - \alpha_1[n]|h_{sr}[n]|^2 \geq \frac{N}{(1 - \rho[n])}, \quad \forall n \in \mathcal{N}, \quad (8.47b)$$

$$\alpha_2[n]|h_{sd}[n]|^2 - \alpha_1[n]|h_{sd}[n]|^2 \chi[n] \geq \delta_1^2[n] \chi[n], \quad \forall n \in \mathcal{N}, \quad (8.47c)$$

$$(8.21f), (8.21g), (8.21h),$$

where $\chi[n]$ and $\iota[n]$ are given as follows:

$$\chi[n] = \Gamma_{\min}[n] - \frac{\eta \rho[n] |h_{sr}[n]|^2 |h_{rd}[n]|^2}{\delta_2^2[n]}, \quad \forall n \in \mathcal{N}, \quad (8.48)$$

$$\iota[n] = \frac{1}{2P_{\text{sum}}^2[n]R_{\text{sum}}^2[n]}, \quad \forall n \in \mathcal{N}. \quad (8.49)$$

It can be observed that all constraints are linear and convex. Nevertheless, the objective function is non-convex due to the non-concavity of the sum data rate function. To deal with this issue, we utilize the result of the following corollary [3].

Corollary 2 Consider \mathcal{F} as a decreasing function, then

$$\min_{\boldsymbol{\Upsilon} \in \mathcal{C}} \sum_{j=1}^J \mathcal{F}_j \left(\frac{\mathcal{A}_j(\boldsymbol{\Upsilon})}{\mathcal{B}_j(\boldsymbol{\Upsilon})} \right), \quad (8.50)$$

is equivalent to the following problem:

$$\min_{\boldsymbol{\Upsilon} \in \mathcal{C}, \varrho_j} \sum_{j=1}^J \mathcal{F}_j \left(2\varrho_j \sqrt{\mathcal{A}_j(\boldsymbol{\Upsilon})} - \varrho_j^2 \mathcal{B}_j(\boldsymbol{\Upsilon}) \right), \quad (8.51)$$

with the updated value of $\varrho_j = \frac{\sqrt{\mathcal{A}_j(\boldsymbol{\Upsilon})}}{\mathcal{B}_j(\boldsymbol{\Upsilon})}$.

Proof 9 Please refer to [3]. ■

Adopting the results of Corollary 1 allows for the transformation of the non-convex terms within the objective function of problem (P_6) , specifically addressing the complexities introduced by the second term. Corollary 1 suggests a methodology for handling non-convex

functions within an optimization problem, making it possible to recast them in a manner that retains the problem's original intent while rendering it more amenable to solution through convex optimization techniques. The transformation typically involves introducing auxiliary variables or applying mathematical operations that expose the underlying convex structure, allowing for the application of convex optimization solvers. By adopting the result of Corollary 1, the second term of the objective function in (P_6) can be equivalently written as:

$$\min_{\alpha_1[n], \alpha_2[n], \varrho[n]} \sum_{n=1}^N \frac{1}{4\iota[n]} \frac{1}{\hat{R}_{\text{sum}}^2[n]} \quad (8.52)$$

where

$$\begin{aligned} \hat{R}_{\text{sum}}[n] = & \log_2(1 + \gamma_r^1[n]) \\ & + \log_2 \left(1 + \gamma_d^2[n] + 2\varrho[n] \sqrt{\alpha_2[n] |h_{sd}[n]|^2} - \varrho^2[n] (\alpha_1[n] |h_{sd}[n]|^2 + \delta_1^2[n]) \right), \end{aligned} \quad (8.53)$$

where

$$\varrho[n] = \frac{\sqrt{\alpha_2[n] |h_{sd}[n]|^2}}{\alpha_1[n] |h_{sd}[n]|^2 + \delta_1^2[n]}, \forall n \in \mathcal{N}. \quad (8.54)$$

The modified sum data rate $\hat{R}_{\text{sum}}[n]$ is now biconcave with respect to the power allocation coefficients and $\varrho[n]$. Accordingly, the multi-convex optimization problem can be formulated as:

$$\begin{aligned} P_7 : \min_{\alpha[n], \varrho[n]} & \sum_{n=1}^N \iota[n] P_{\text{sum}}^2[n] + \sum_{n=1}^N \frac{1}{4\iota[n]} \frac{1}{\hat{R}_{\text{sum}}^2[n]} \\ \text{s.t. :} & (8.21f), (8.21g), (8.21h), (8.47b), (8.47c), \end{aligned} \quad (8.55a)$$

where $\alpha[n] = [\alpha_1[n], \alpha_2[n]] \in \mathbb{R}^{2 \times 1}$.

It is worth noting that $P_{\text{sum}}[n]$ is a function of power allocation coefficients, and every coefficient has its own constraint. Hence, the terms of $P_{\text{sum}}[n]$ and $\hat{R}_{\text{sum}}[n]$ are decoupled to optimize $P_{\text{sum}}[n]$ distributively. As a result, the Augmented Lagrangian Method (ALM) is adopted where a penalty term is added to the Lagrange function of the problem (P_7) as

follows:

$$\begin{aligned}
L_{\kappa}(\alpha_1[n], \alpha_2[n], \varphi, \theta, \Theta, \mu, \vartheta) = & \sum_{n=1}^N \iota[n] P_{\text{sum}}^2[n] + \sum_{n=1}^N \frac{1}{4\iota[n]} \frac{1}{\hat{R}_{\text{sum}}^2[n]} \\
& + \frac{1}{2\kappa} \left[\left(\left[\sum_{n=1}^N \varphi_n + \kappa \left(\frac{N}{(1-\rho[n])} - \frac{\alpha_2[n]|h_{sr}[n]|^2}{\gamma_{\min}[n]} + \alpha_1[n]|h_{sr}[n]|^2 \right) \right]^+ \right)^2 \right. \\
& + \left(\left[\sum_{n=1}^N \theta_n + \kappa (\delta_1^2[n]\chi[n] - \alpha_2[n]|h_{sd}[n]|^2 + \alpha_1[n]|h_{sd}[n]|^2\chi[n]) \right]^+ \right)^2 \\
& + \left(\left[\sum_{n=1}^N \Theta_n + \kappa (\alpha_1[n] + \alpha_2[n] - P_{\text{peak}}) \right]^+ \right)^2 \\
& + \left(\left[\mu_n + \kappa \left(\frac{1}{N} \sum_{n=1}^N \alpha_1[n] + \alpha_2[n] - P_{\text{max}} \right) \right]^+ \right)^2 \\
& + \left(\left[\sum_{n=1}^N \vartheta_n - \kappa P[n] \right]^+ \right)^2 \\
& \left. - \sum_{n=1}^N \varphi_n^2 - \sum_{n=1}^N \theta_n^2 - \sum_{n=1}^N \Theta_n^2 - \mu_n^2 - \sum_{n=1}^N \vartheta_n \right], \tag{8.56}
\end{aligned}$$

Note that it has been determined that the ALM performs better than the traditional sub-gradient or dual-descent method, as highlighted by Bertsekas [277]. A key advantage of ALM is its superior performance in navigating the complex landscape of optimization problems, such as (P₇), where it not only facilitates a more efficient search for optimal solutions but also ensures robust adherence to constraints. One of the significant merits of ALM is its convergence properties. Unlike some methods that require a large penalty term to ensure convergence towards an optimal or near-optimal solution, ALM guarantees convergence without necessitating such conditions. This makes ALM particularly appealing for a wide array of optimization problems where the balance between objective optimization and constraint satisfaction is delicate and complex.

In the context of (P₇), the augmented Lagrangian incorporates a penalty factor, κ , alongside Lagrange multipliers $\{\varphi, \theta, \Theta, \mu, \vartheta\}$. These components work together to steer the optimization process towards solutions that are feasible within the problem's constraints while penalizing deviations from these constraints to maintain a strict adherence to them. Each iteration of the optimization, denoted as I , yields a solution $G(I)$, progressively refining the approach towards an optimal or sub-optimal solution. Finally, our proposed efficient low complexity sub-optimal algorithm is sketched in **Algorithm 9**.

8.5 Complexity Analysis

The overall complexity of the proposed two-stage solution is determined by the complexities of solving three optimization problems: P_2 , P_5 , and P_7 , associated with finding the optimal Power-Splitting factor, the miniature UAV trajectory, and the NOMA power coefficients, respectively. For P_2 , involving $3N$ constraints and N decision variables, the complexity aligns with that of an Interior Point Method for convex optimization, expressed as $O_1 =$

Algorithm 9 Iterative Resource Allocation Algorithm for EE maximization of THz-NOMA SWIPT-aided Miniature UAV Networks

Input: Set iteration indices $i = 0, k = 0, l = 0$,
 Set the maximum convergence iteration index l_{\max} ,
 Set the tolerance to $\epsilon_1 = \epsilon_2 = 10^{-3}$,
 Initialize $\lambda, \psi, a^{(k)}[n], b^{(k)}[n], v^{(k)}[n], t^{(k)}[n], \rho, \mathbf{q}^{(k)}, \alpha$,
 Set Lagrange multipliers $\varphi^l, \theta^l, \Theta^l, \mu^l, \vartheta^l$, and the penalty factor κ^l .

1: **repeat**
 2: Given $\{\lambda, \psi, \alpha[n], \mathbf{q}[n]\}$, solve (P_2) to obtain $\rho[n]$.
 3: **while** $|F^{(k)} - F^{(k-1)}| \geq \epsilon_1$ do
 4: Given $\{\alpha[n], \rho[n]\}$, solve (P_5) to obtain $\mathbf{q}^{(k)}[n]$.
 5: Update $b^{(k)} = \ln(v^{(k)}[n])$, $a^{(k)} = \ln(t^{(k)}[n])$ according to (8.34) and (8.35).
 6: Set $k = k + 1$.
 7: **end while**
 8: **if** (8.44) is satisfied **then return** $(\rho^*[n], \mathbf{q}^*[n])$.
 9: **else** Update λ and ψ according to (8.42) and (8.43).
 10: Set $i = i + 1$.
 11: **until** (8.25) and (8.26) are satisfied or $i = l_{\max}$.
 12: **while** $|G^{(l)} - G^{(l-1)}| \geq \epsilon_2$ do
 13: Given $\{\mathbf{q}[n], \rho[n]\}$, solve (P_7) to obtain $\alpha[n]$.
 14: Update the Lagrange multipliers $\varphi_n^{l+1}, \theta_n^{l+1}, \Theta_n^{l+1}, \mu_n^{l+1}$, and ϑ_n^{l+1} .
 15: Update the penalty factor $\kappa^{l+1} = 2\kappa^l$.
 16: Set $l = l + 1$.
 17: **end while**
 18: **return** $(\rho^*[n], \mathbf{q}^*[n], \alpha^*[n])$.

$\mathcal{O}(N(3N)^2)$. P_5 has $(8N+3)$ constraints and $5N$ decision variables. Its complexity, based on the Successive Convex Approximation methodology, is $\mathcal{O}_2 = \mathcal{O}((8N+3)(5N)^3)$. The complexity of P_7 , following the Augmented Lagrangian Method, is $\mathcal{O}_3 = \mathcal{O}(N^2)$. Hence, the total complexity of the proposed solution is the sum of the individual complexities: $\mathcal{O}_{\text{total}} = \mathcal{O}_1 + \mathcal{O}_2 + \mathcal{O}_3 = \mathcal{O}(9N^3 + (8N+3)(5N)^3 + N^2)$, indicating a polynomial time complexity of degree four.

8.6 Simulation Results and Discussions

In the context of our simulation framework, we consider a scenario within a defined square area, each side measuring 30 meters, designed to contain a single user and one miniature UAV, both of which are positioned in a random manner within this space. To mitigate the occurrence of peaks in path loss, the selection of the carrier frequency was carefully made at $f = 1.2$ THz, coupled with a chosen transmission bandwidth of 10 GHz. Acknowledging the critical role that water vapor plays in influencing molecular absorption loss in THz channels, our model accounts for the frequency-dependent absorption coefficient, $\xi(f)$, by attributing it exclusively to the presence of water vapor molecules, as referenced in the literature [274, 278, 279]. Furthermore, $V_{\max} = 1$ m/s, $\varpi = 0.1$ Sec, $T = 45$ Sec, $\sigma_2^2 = \delta_1^2[n] = \delta_2^2[n] = -174$ dBm/Hz, $H_1 = 2, H_2 = 3$ meters, $P_{\text{peak}} = P_{\max} = 1, P_c = 0.52$ W,

Table 8.1: Simulation Parameters for EE maximization of THz-NOMA SWIPT-aided Miniature UAV Networks.

Parameter	Value
Area side length	30 meters
Carrier frequency	1.2 THz
Transmission bandwidth	10 GHz
Absorption coefficient, $\xi(f)$	Frequency-dependent
Maximum flying speed of the miniature UAV, V_{\max}	1 meter/Sec
Duration of each time slot, τ	0.1 Sec
Miniature UAV Operation time, T	45 Sec
Noise power spectral density	-174 dBm/Hz
Source Node altitude, H_1	2 meters
Miniature UAV altitude, H_2	3 meters
Peak power, P_{peak}	1 W
Circuit power, P_c	0.52 W

following the guidelines set forth in referenced studies[270, 274].

We have derived all statistical results based on the aggregation of data obtained from an extensive series of simulation trials, generating numerous random realizations of the channel gains. This methodical approach allows for a comprehensive understanding of the dynamics involved in the deployment and operation of the miniature UAV within the specified environmental conditions, offering valuable insights into the optimization of UAV-assisted communication networks. All simulation parameters are also summarized in *Table 8.1*.

To rigorously evaluate the effectiveness of our proposed resource allocation algorithm, we conducted a comparative analysis against a set of benchmark methods, each designed to highlight different aspects of system performance under varied conditions:

- Method A: Evaluates the proposed algorithm under a NOMA framework with a static NOMA power coefficient.
- Method B: Compares the system's performance under dual access mechanisms (NOMA vs. Orthogonal Multiple Access (OMA)) to determine which is superior.
- Method C: Analyzes the proposed algorithm assuming a pre-defined UAV flight course.
- Method D: Considers a scenario with uniform PS factors ($\rho[n] = 0.5, \forall n \in \mathcal{N}$).
- Method E: Utilizes a fractional programming approach from [280] without optimizing PS factors.

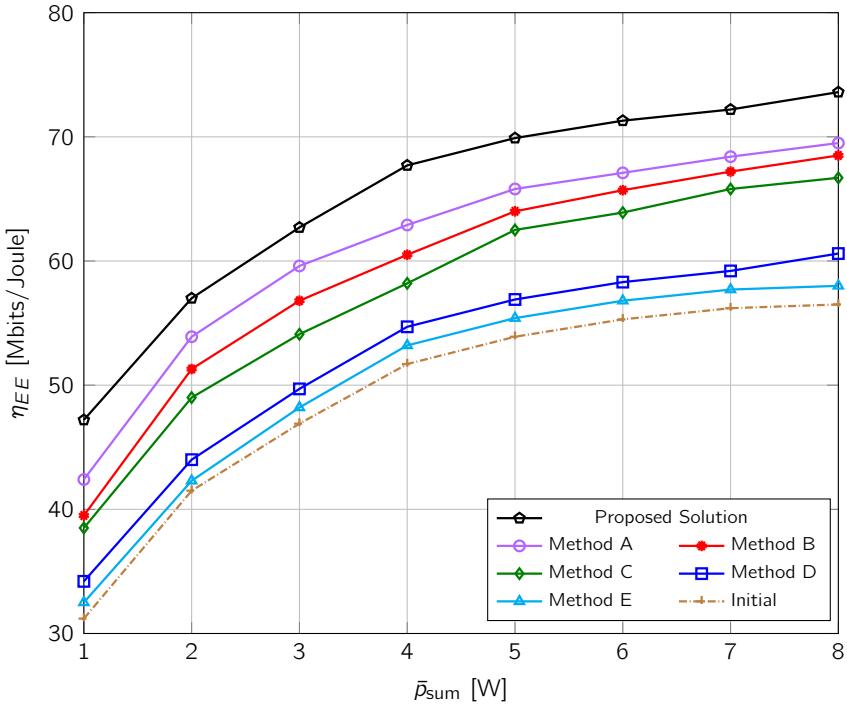


Figure 8.4: The effect of average network transmit power, \bar{p}_{sum} , on EE of THz-NOMA SWIPT-aided miniature UAV networks.

Fig. 8.4 presents the EE dynamics as influenced by the average network transmission power, denoted by $\bar{p}_{sum} = P_{max} + P_{peak} + P_c - P_{EH}$. Within this representation, the curve labeled 'Initial' sketches out the EE performance following an initial, non-optimized (random) setting of the miniature UAV flight path. A noteworthy observation from our investigation is the consistent outperformance of our proposed algorithm over various benchmark algorithms. This superiority becomes even more pronounced with an increase in \bar{p}_{sum} , indicating a relative expansion in performance disparity.

In a parallel comparison depicted in Fig. 8.4, we examined the average EE performance across these five distinct methodologies. The results demonstrated the superiority of our proposed strategy, showing performance enhancements of:

- 30.3% compared to Method A (static NOMA power coefficient)
- 23.0% compared to Method B (NOMA vs. OMA)
- 21.2% compared to Method C (pre-defined UAV flight course)
- 18.1% compared to Method D (uniform PS factors)
- 7.26% compared to Method E (fractional programming without PS optimization).

This empirical evidence firmly establishes the proposed algorithm's capability to significantly

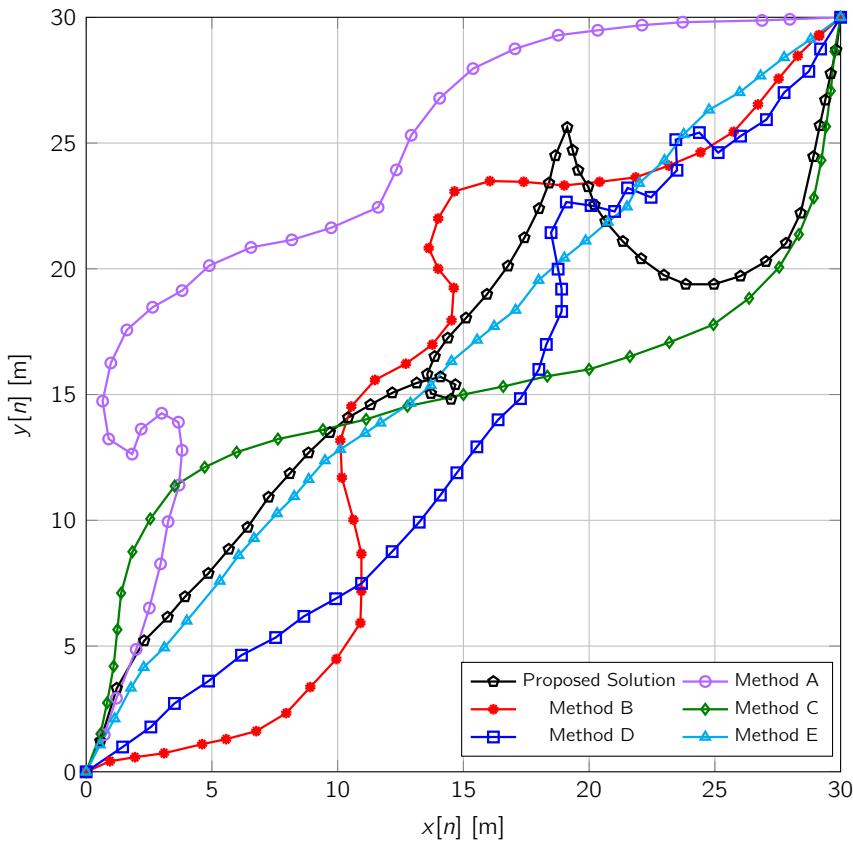


Figure 8.5: The trajectory of miniature UAV in the cooperative THz NOMA-SWIPT network.

enhance EE, highlighting its utility and effectiveness within miniature UAV-assisted communication frameworks.

Fig. 8.5 presents the results of trajectory optimization for the miniature UAV, showcasing the paths that have been refined for enhanced operational efficiency. Moving forward, Fig. 8.6 embarks on a detailed analysis of how the duration of the mission, i.e., the miniature UAV's operational time, denoted by the parameter T , impacts the EE of various benchmark schemes. An interesting pattern emerges from this examination: as mission time extends, there's a notable improvement in EE for schemes operating under fixed trajectories (Method D) and those initiated with non-optimal, feasible configurations (labelled 'Initial'). This upward trend in EE, attributable to prolonged communication opportunities and the flexibility to adjust flight parameters over time, is not uniformly observed across all methods. Specifically, Methods A, B, and C do not exhibit this consistent rise in EE with an increase in T . Quantitatively, the extension of mission time is associated with significant boosts in EE

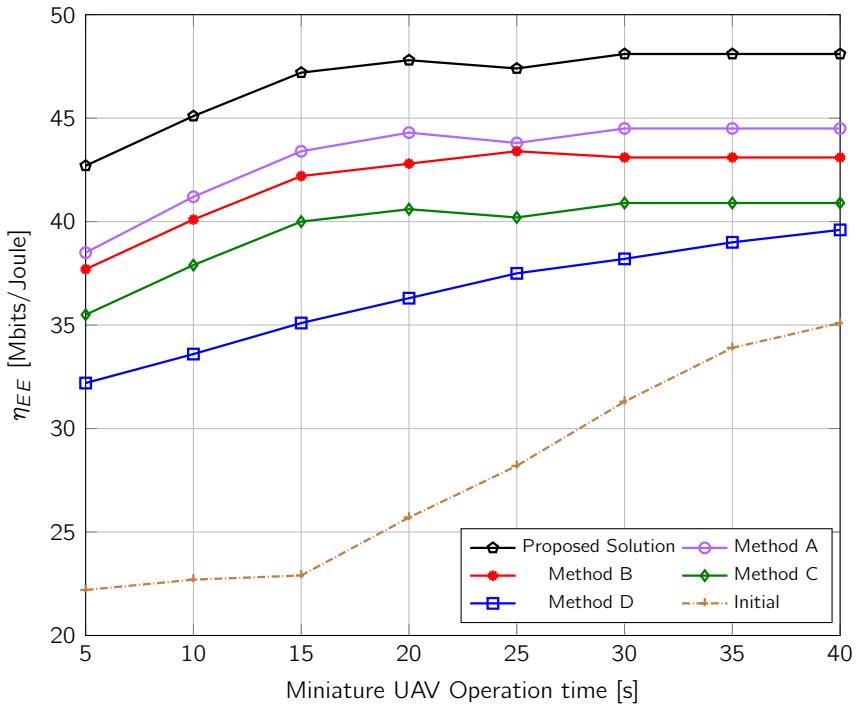


Figure 8.6: The EE vs. the miniature UAV's operational time in the THz-enabled SWIPT network.

performance, enhancing it by at least 37.1%, 26.8%, 22.8%, and 16.5%, 12.8% respectively. This enhancement suggests that longer mission times offer a strategic advantage by allowing for more optimal communication strategies and flight parameters' adjustment, thereby improving overall efficiency. However, the relationship between mission time and EE is nuanced. The interaction among optimization variables results in a non-linear, albeit generally increasing, trend in EE as mission time is extended. This phenomenon indicates the complex dynamics at play when optimizing for EE, where certain adjustments can lead to significant gains in efficiency. The observation that mission time can have such a profound impact on EE underscores an intriguing challenge: *minimizing the task completion time for UAV relay systems to meet specific EE criteria*. This challenge points towards a delicate balance that must be struck between operational efficiency and the urgency of mission completion, highlighting a fertile ground for further research and exploration in optimizing UAV-based communication networks.

Figure 8.7 demonstrates the intricacies of how the molecular absorption coefficient — a fundamental factor in THz communication systems — affects EE under a variety of environmental conditions. A salient observation from this figure is the evident inverse relationship between the molecular absorption coefficient and the EE across different communication schemes. Specifically, as the absorption coefficient increases, signifying higher propagation losses, there is a noticeable decrement in EE for all analyzed methods. This trend is primarily ascribed to the aggravated signal attenuation caused by environmental variables such as humidity and temperature, which intensify molecular absorption. This increase in molecular absorption, while potentially minimizing information leakage from the miniature

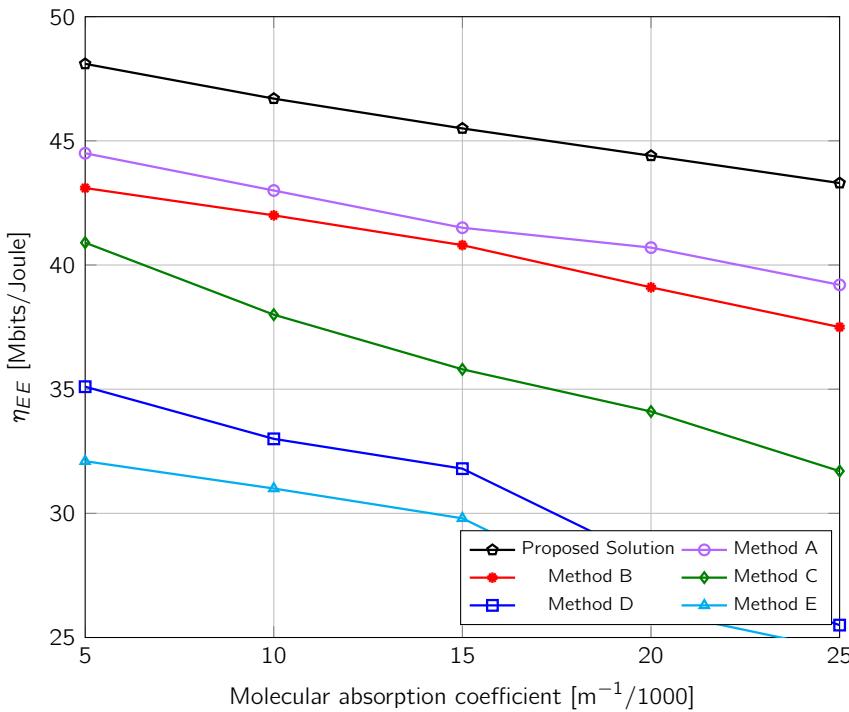


Figure 8.7: The EE vs. molecular absorption in the THz-enabled miniature UAV network.

UAV, simultaneously impairs the quality of signal reception at the intended destination node. Such a degradation has a direct adverse impact on the overall EE. Remarkably, even under these challenging environmental conditions, the proposed solution distinctly surpasses the performance of baseline methods. This underscores the resilience and effectiveness of our proposed algorithm, adeptly counteracting the detrimental effects of increased molecular absorption to sustain higher EE levels in networks that are empowered by THz technology and UAV integration. This consistency in outperforming baseline approaches underlines the adaptability and superiority of the proposed solution in maintaining optimal communication efficiency, regardless of the environmental constraints encountered.

8.7 IRS-based UAV with Underlaid D2D Users in THz Networks

Throughout our journey in this chapter, we have pioneered a groundbreaking concept, the 'miniature' UAV, envisioned as a mobile relay akin to an IRS with the unique ability to change its location dynamically. Recognizing that an IRS fundamentally acts as a relay by reflecting and manipulating signal paths, the miniature UAV (as a relay but not generally) takes this concept airborne, offering a new dimension of flexibility and adaptability in network topology.

This dissertation has been devoted to a deep dive into IRS technology, probing its capabilities,

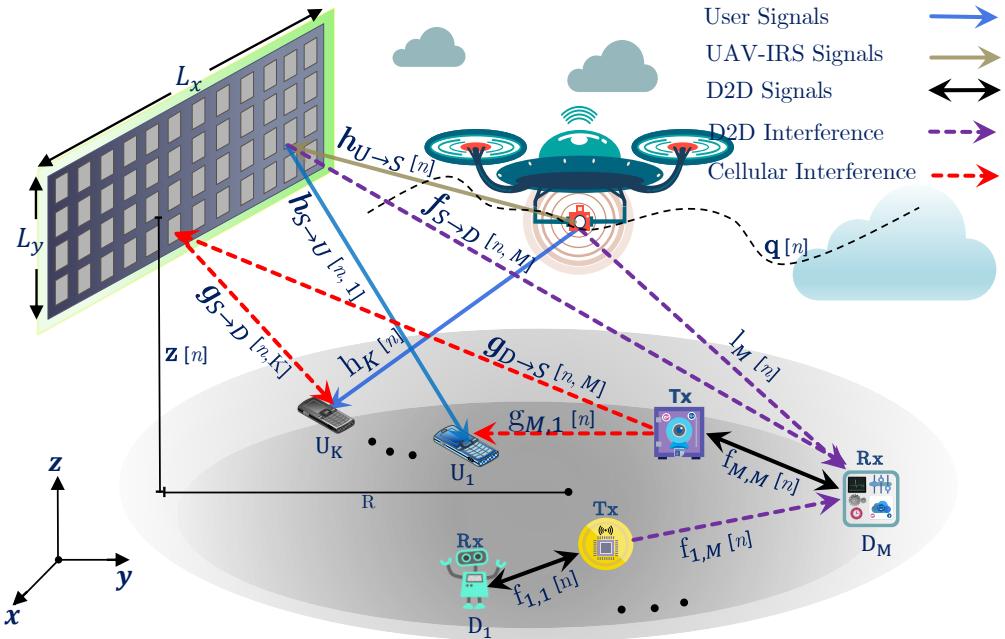


Figure 8.8: Model of the UAV with underlaid device-to-device (D2D) communications network. The system model of an uplink single macro cell OFDMA-based cellular network with one base station to serve M CUs and K DUs. In this figure, the green arrow shows the cellular transmission link between the base station and the CUs, while the dotted line indicates the D2D links.

limitations, and applications. Hence, a natural progression for our work involves examining the practical integration of such miniature UAVs into IRS-aided networks. The intriguing question we pose is: How can we leverage the agility and dynamism of miniature UAVs to enhance IRS functionality within complex network environments?

To avoid redundancy, we have deliberately chosen not to repeat system models previously discussed. Instead, we propose that future studies should build upon the established system model, extending it to incorporate the unique characteristics of miniature UAVs. This innovative approach opens up a plethora of opportunities for improving network coverage, optimizing signal quality in real-time, and delivering robust wireless communication in scenarios where fixed IRS installations might fall short.

In essence, the inclusion of miniature UAVs into our system model ushers in a novel era of dynamic, aerial IRS solutions that could transform the infrastructure of modern wireless networks. It invites a fascinating research avenue that promises to push the boundaries of current wireless communication technologies and pave the way for unprecedented advancements in the field.

Taking Device-to-Device (D2D) communication as an example to explore the possible integration of miniature UAVs, its integration with UAVs can redefine the operational efficiency of wireless communication systems, particularly in regions where conventional communication infrastructure is sparse or non-existent for D2D services [1]. UAVs serve not only as agile

aerial platforms enhancing coverage but also as spectrum-sharing facilitators that can significantly boost network throughput. This innovative model allows for a dual-communication modality where some users can establish direct D2D links while others are served through the UAV, resulting in a versatile and high-performing network.

Recent research has studied the optimization of UAV-D2D integrated networks [281, 282, 283, 284, 125, 285, 286, 287, 288, 289]. Studies have highlighted the delicate balance between leveraging UAVs for extended coverage and the efficient use of D2D communication for localized data exchange. These investigations consider crucial parameters such as data rate, coverage, power control, and the strategic deployment of UAVs to augment the network's capabilities [281, 284]. For example, one study examined the pivotal interplay between the altitude of UAV and the density of D2D pairs, unpacking their collective influence on network coverage and data-rate performance [282]. Another research effort provided insights into the multi-hop capabilities of D2D pairs under the watchful guidance of a UAV, particularly within the context of IoT networks [289]. Further research optimized the power control in UAV-aided D2D communications, focusing on maximizing throughput while adhering to interference constraints [283]. Additionally, the efficacy of employing NOMA within UAV-aided networks was scrutinized, revealing the potential for power control optimization to mitigate outage probabilities for both cellular users and D2D pairs, thus enhancing overall communication reliability [283].

These pioneering studies underscore the potential of integrating UAVs with D2D communication, a synergy that not only extends the reach of networks but also fine-tunes their performance across diverse scenarios. This research trajectory is especially pertinent as we steer towards a future where the harmonization of aerial and ground-level communication frameworks will be paramount. The continued exploration and advancement in this domain promise to fortify the robustness of wireless systems, ensuring connectivity resilience even in the most challenging environments.

We propose an innovative approach that leverages miniature UAVs operating in the THz frequency network as transmitters (the miniature UAV we developed earlier is not a relay anymore but rather an aerial base station). In this scheme, the IRS is employed as a strategic enhancer of signal coverage. This proposition stands at the confluence of advanced aerial mobility and cutting-edge reflective technology, potentially ushering in an era of communication networks characterized by expanded reach and improved signal fidelity. To explore further, we consider a downlink UAV-NOMA system with an underlying D2D communication network, as shown in Fig 8.8. As seen, the operational region is a circular domain of radius R . The network hosts single-antenna CUs and DU pairs, succinctly represented by sets $\mathcal{K} = \{U_1, \dots, U_K\} = \{1, \dots, K\}$ and $\mathcal{M} = \{D_1, \dots, D_M\} = \{1, \dots, M\}$, respectively. The users are randomly placed within the circular region, with CU users engaged in communication with a THz-enabled UAV. This UAV, pivotal in our system model, is complemented by an IRS to enhance its communication capabilities. Ensuring unobstructed interaction between the IRS, the UAV, and the terrestrial CUs, the IRS is strategically positioned at the service area's periphery, enabling an omnidirectional perspective over the ground users. The IRS is constructed from $L = L_x \times L_y$ Passive Reflection Units (PRUs), which collectively form a Uniform Linear Array (ULA). The ensemble of IRS elements is encapsulated by the set $\mathcal{L} = \{1, \dots, L\}$.

The architecture of our system further entails an underlay strategy for D2D communications,

where each D2D receiver is subject to interference originating not only from the UAV but also from other D2D transmitters. Conversely, CU must contend with interference from all D2D transmitters. A three-dimensional Cartesian coordinate framework serves as our mathematical canvas, allowing for precise representation of the nodes' positions without sacrificing generality. The UAV's trajectory in this spatial tapestry is articulated as $q(t) = [x(t), y(t), z(t)]^T \in \mathbb{R}^{3 \times 1}$, while the coordinates of CUs and DUs are defined by $c_k(t) = [c_{x,k}(t), c_{y,k}(t), 0]^T \in \mathbb{R}^{3 \times 1}$, $\forall k \in \mathcal{K}$ and $v_m(t) = [v_{x,m}(t), v_{y,m}(t), 0]^T \in \mathbb{R}^{3 \times 1}$, $\forall m \in \mathcal{M}$ for the duration $0 < t < T$, respectively. The IRS location is denoted by $s = [s_x, s_y, s_z]^T \in \mathbb{R}^{3 \times 1}$.

We postulate that the incorporation of IRS into the miniature UAV THz network infrastructure will significantly enhance network performance, benefiting both DUs and CUs in terms of spectral efficiency. This improvement is not limited to data transmission rates alone; we also anticipate a notable boost in the overall energy efficiency of the network. The IRS achieves this by smartly redirecting and focusing the signal energy, which otherwise might dissipate or not reach the intended users effectively, thereby optimizing the use of available spectral resources and reducing the energy required for transmissions.

This hypothesis rests on the ability of the IRS to manipulate electromagnetic waves in a controlled manner, thus enabling a more directed and efficient signal propagation. For DUs, this means an enhanced user experience with faster data rates and reduced latency. For CUs, the benefits include improved coverage and reliability, even at the network's edge or in traditionally challenging environments for signal penetration. The cumulative effect of these enhancements is a network that not only performs better in delivering services to its users but does so with greater energy parsimony.

The anticipated improvements in spectral and energy efficiency underscore the transformative potential of IRS technology. By bridging the gap between the increasing demand for high-quality wireless communication services and the imperative for energy conservation, the IRS emerges as a pivotal technology in the sustainable evolution of wireless network infrastructures. This is also expected in a miniature UAV THz network.

8.8 Conclusion

In this chapter, we have tackled the intricacies of enhancing the efficiency of a cooperative TeraHertz (THz) Non-Orthogonal Multiple Access (NOMA) supported miniature Unmanned Aerial Vehicle (UAV) network integrated with Simultaneous Wireless Information and Power Transfer (SWIPT). This work commenced with the formulation of an Energy Efficiency (EE) optimization problem aimed at refining the network's resource allocation scheme. A novel deployment strategy for the miniature UAV was introduced, tailored to augment wireless THz connectivity. This strategy is distinctive in its meticulous consideration of the molecular absorption phenomenon, a critical factor in the THz-enabled UAV path loss channel gain model.

Building upon this foundation, we crafted an optimization problem with the goal of improving EE within a NOMA-SWIPT cooperative UAV framework, ensuring compliance with stringent Quality of Service (QoS) benchmarks. The crux of this optimization problem was the strategic modulation of decision variables, including the UAV's positioning, the Power-

Splitting (PS)-SWIPT ratio, and the NOMA power allocation coefficients, all of which play pivotal roles in the network's operational efficiency.

To navigate the complexities of the EE optimization, we devised an iterative solution methodology. This method decomposes the problem into three more manageable sub-problems, solved using a two-stage framework. This procedural innovation not only facilitated a more streamlined solution process but also showcased its effectiveness through compelling numerical outcomes. These results proved the resource allocation algorithm's superiority by comparing it against baseline scenarios that did not incorporate optimizations related to miniature UAV trajectory, NOMA power, or SWIPT PS. Such comparative analysis illuminated our proposed strategy's significant contributions to enhancing the network's performance efficiency. It underscored notable advancements in miniature UAV endurance and battery longevity, marking a substantial leap forward in the operational capabilities of THz-NOMA-enabled miniature UAV networks with SWIPT. This work not only expands the frontier of (miniature) UAV network optimization but also sets a precedent for future research in the domain, promising more resilient, efficient, and sustainable aerial communication networks.

Chapter 9

Conclusion

In this dissertation, we embarked on an in-depth exploration of Intelligent Reflecting Surfaces (IRSs) and their transformative impact on the enhancement of wireless network capabilities. Our examination of IRS technology proved its revolutionary role in enhancing networks' key performance indicators (KPIs) (such as data rate, power, spectral, and energy efficiencies) and optimizing resource allocation, thereby illuminating its substantial influence on the evolving landscape of wireless communications. Our analytical (sometimes mathematically involved) approach revealed the unique benefits that laser-aided systems offer, notably through the advanced beamforming capabilities for amplifying desired signals and the dynamic suppression of co-channel interference. This is achieved by the IRS's ability to sense the surrounding wireless environment and adjust its reflection coefficients in real time to optimize signal quality.

Throughout this research, we ventured into the integration of IRS with a variety of service types, including Machine Type Communications (MTC), Ultra-Reliable Low-Latency Communication (URLLC), Internet of Things (IoT) deployments, and Mobile Edge Computing (MEC). This exploration demonstrated IRS setups could open up new perspectives to enhance the efficiency and reliability of these services. We introduced and developed several innovative, low-complexity algorithms designed to establish (sub-)optimal resource allocation policies in GLSS-aided networks. Our analytical insights were validated through extensive simulations, providing a strong foundation for the practical implementation of IRS in real-world scenarios.

As we conclude this dissertation, we present a comprehensive summary of our findings, reflections on the journey undertaken, and the broader conclusions drawn from our research. Moreover, we offer a few potential future directions for IRS-aided wireless communication networks, highlighting the untapped possibilities and emerging challenges that await. Our discussion extends to envisage the integration of IRS with emerging technologies, anticipating the innovative applications and solutions that IRS technology could unlock in the quest for more efficient, reliable, and versatile wireless communication systems. This final chapter serves not only as a closure to our current research endeavors but also as a beacon guiding future explorations in the ever-evolving domain of IRS-enhanced wireless networks.

9.1 Main findings

Now we can look at an Intelligent Reflecting Surface (IRS) as an innovative technology designed to enhance wireless network performance by smartly manipulating the propagation of ElectroMagnetic (EM) waves. Comprising numerous passive or active reflecting elements, an IRS can adjust the phase shifts of incoming signals, thereby directing them toward specific receivers to improve signal strength, coverage, and overall network efficiency. This dissertation has delved deeply into the exploration, analysis, and application of IRS technology, recognizing its potential to revolutionize the field of wireless communications. Focusing on the myriad ways IRS can be integrated into existing or future networks to address current challenges and optimize performance, the study has unfolded various dimensions of this cutting-edge technology.

In the following paragraphs, we will outline the key contributions of this study, highlighting the significant strides made in advancing our understanding and application of IRS within wireless communication systems.

Contribution 1

Optimizing Power Efficiency in SWIPT Networks through Advanced Beamforming and Antenna Selection

The initial contribution of this dissertation sets the stage for subsequent explorations by focusing on the optimization of power efficiency within single-cell networks that incorporate multi-antenna and multi-user configurations, specifically focusing on the integration of Simultaneous Wireless Information and Power Transfer (SWIPT). By tackling the dual objectives of maximizing energy harvested and minimizing power consumption, we introduced an optimization strategy that navigates the complexities of beamforming and antenna selection. This foundational contribution not only establishes the theoretical background necessary for optimizing SWIPT-enabled networks but also demonstrates, through extensive simulations, the potential enhancements in power and energy efficiencies. The development and validation of low-complexity, locally optimal solutions signify a major step towards balancing energy conservation and operational efficiency in future wireless communication networks, serving as the baseline upon which further IRS-related investigations are based.

Contribution 2

Enhancing URLLC with IRS: A Leap Towards Ultra-Reliable Communications

The second major contribution is the utilization of IRS within multi-user Multiple-Input Single-Output (MISO) systems to significantly enhance Ultra-Reliable Low-Latency Communication (URLLC) services. By aiming to reduce total transmission power through the simultaneous optimization of active and passive beamformers, we designed an advanced Alternating Optimization (AO) algorithm. Our approach not only illuminates the intricate rela-

tionship between active and passive beamforming but also highlights the substantial promise the IRS holds for refining URLLC systems. Supported by simulation studies, this contribution underscores the efficiency and effectiveness of the proposed solutions, marking a pivotal step towards harnessing the full capabilities of the IRS in future wireless communications.

Contribution 3

IRS-Assisted MTC Systems: Bridging IoT Connectivity with Energy Efficiency

We explored the dynamics of integrating Machine Type Communication (MTC) technology within a multi-user MISO system, enhanced through the deployment of an IRS. We optimized the system's total Energy Efficiency (EE) while maximizing the serviceability of IoT users through a strategic joint optimization of active and passive beamformers. The introduction of another AO algorithm highlighted the IRS's significant impact on improving system efficiency and its ability to accommodate a larger number of users within IoT frameworks. Through this exploration, a delicate balance emerges between EE and Spectral Efficiency (SE), revealing the potential of IRS technology in optimizing the performance of MISO MTC-enabled networks.

Contribution 4

Leveraging IRS for Computational Offloading in MEC-Enabled Multiuser MTC Networks

The next contribution of this dissertation delved into the synergy between MEC and multiuser MTC scenarios. We underlined the pivotal role of the IRS in facilitating computational offloading to improve latency and reliability for MTC devices. By optimizing the joint radio resource allocation and edge offloading decisions within an IRS-aided network, we presented a novel approach toward enhancing the efficiency of edge computing for MTC environments. Through the development of an efficient iterative algorithm and supported by simulation results, we extended the utility of IRS beyond signal enhancement, illustrating its capacity to significantly impact the computational aspects of wireless networks.

Contribution 5

Active IRS in mmWave Networks: Pushing the Boundaries of High-Frequency Wireless Communication

While the previous contributions primarily focused on Frequency Range 1 (FR1), delving into the potentials and challenges within sub-6 GHz bands, we ventured into uncharted territory with our fifth contribution. Our next contribution marks our inaugural exploration into Frequency Range 2 (FR2), specifically within the high-frequency spectrum of mmWave

wireless networks. To this end, we investigated an active IRS-assisted MISO system operating in mmWave frequencies. We addressed the challenge of optimizing the system sum rate by exploring the unique advantages offered by active IRS configurations. By formulating a comprehensive optimization problem and introducing two low-complexity algorithms, we showcased the potential of active IRS in enhancing network performance and also set a new benchmark for the application of IRS technology in mmWave spectra, opening new avenues for research and development.

Contribution 6

UAV-Assisted NOMA Networks in the THz Band: Maximizing Energy Efficiency through Strategic Deployment

Pushing the boundaries even further in the frequency spectrum, this contribution explores the realm of Unmanned Aerial Vehicle (UAV)-assisted Non-Orthogonal Multiple Access (NOMA) networks that operate in the TeraHertz (THz) band, a domain known for its potential to revolutionize high-speed wireless communication. We introduced a novel framework designed to optimize energy efficiency through meticulously planned miniature UAV trajectory deployment and the formulation of an effective network resource allocation strategy. We studied the critical role of miniature UAV mobility, NOMA power allocation strategies, and SWIPT Power-Splitting (PS) ratios in influencing the overall system performance. By introducing an innovative optimization approach and demonstrating the intertwined effects of these parameters on energy efficiency, this contribution advances the understanding of THz band communication and its implications for future wireless networks.

Each of these contributions represents a significant advancement in the field of wireless communications, reflecting four years of dedicated research, continuous effort, and fruitful collaboration with globally recognized scientists. Together, they encapsulate a comprehensive exploration of IRS technology and its potential to redefine the landscape of wireless network capabilities.

9.2 Future Research Directions

In this final section, we discuss potential directions for future research based on the results obtained in this dissertation.

- **Frequency-Dependent Response Model for IRSs:** The assumption of a uniform response from IRS elements across the spectrum is increasingly untenable as we extend the frequency range of next-generation wireless networks. Beginning in the sub-6 GHz range, where IRS technology has traditionally proposed, the expansion into mmWave, terahertz, and even higher frequency domains, 430 THz to 790 THz, such as Optical Wireless Communications (OWC) and Visible Light Communication (VLC) presents new challenges and opportunities. The interaction between electromagnetic waves and

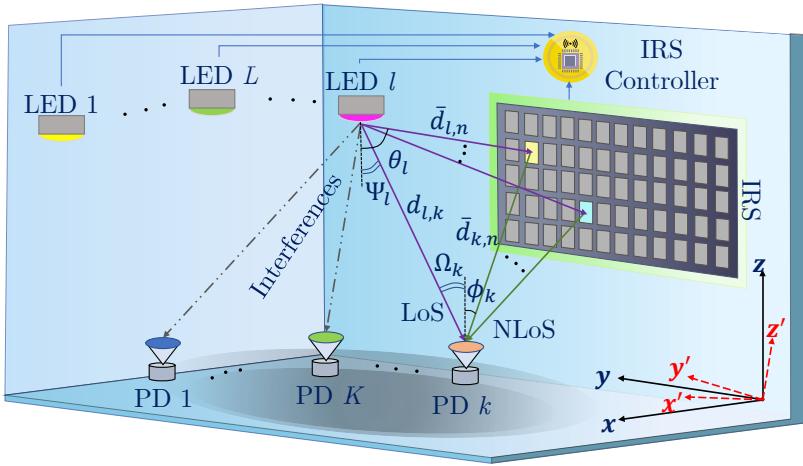


Figure 9.1: Illustration of the Optical IRS (OIRS)-supported OWC network, where the l -th LED and its reflection are symmetrically positioned across the $x'y'z'$ plane. We consider the downlink of an OIRS-aided cell-free OWC system, where L LEDs serve K PD users (or photodetectors), with an OIRS with N units enhancing communication.

IRS elements in these higher frequencies is vastly different, necessitating sophisticated, empirically informed IRS models that can accurately reflect frequency-dependent behaviors. Such advanced modeling is essential for fine-tuning the IRS's phase shift capabilities across these diverse frequency bands, optimizing performance from the well-established sub-6 GHz spectrum to the frontier territories of OWC and VLC. In these higher bands, the potential of IRS to reshape wireless communications is particularly pronounced, with applications ranging from enhancing indoor penetrations to providing high-capacity, low-latency communications in densely populated urban centers. Fig 9.1 shows a generic optical IRS setup, where the LEDs transmits information symbols both directly and indirectly through the IRS to PDs. Research in this area must continue to push the boundaries of frequency utilization, ensuring that the deployment of IRS technology keeps pace with the rapid evolution of wireless standards and the growing demand for bandwidth and connectivity. By addressing the unique propagation characteristics and interaction mechanisms of each frequency band, IRS technology can be harnessed to its full potential, facilitating a seamless wireless future that spans across a wide spectrum of frequencies [290, 291, 292, 293, 294, 295, 296, 297, 298, 299].

- **Near-Field Modeling for IRS:** As the deployment scenarios for the IRS become more varied, including close proximity indoor environments like manufacturing plants or offices, the traditional far-field models become less applicable. Near-field effects, characterized by spherical wavefronts rather than planar ones, significantly impact the performance of the IRS. Future investigations should delve into analyzing IRS' aperture at high frequencies and creating accurate near-field models for IRS, focusing on the distinct propagation characteristics and their implications on IRS-assisted communication. Understanding near-field interactions is crucial for accurately configuring

IRS in environments where users are within a few wavelengths of the surface, ensuring optimal signal enhancement or suppression as required [300, 301, 302, 303, 304, 305, 306, 307, 308].

- **Low-Complexity Optimization Techniques for IRS:** Current optimization algorithms for configuring IRS elements often come with high computational complexity, making them less viable for dynamic or real-time applications with user mobility. The next leap in IRS research should involve the development and integration of low-complexity, possibly machine learning-driven, optimization techniques. These algorithms should aim at real-time adaptability, enabling the IRS to adjust dynamically to changing environmental conditions and user demands with minimal computational overhead. Such advancements are essential for the seamless integration of the IRS into future networks, where flexibility and responsiveness are key [309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319].
- **Leveraging Statistical CSI in IRS-Enhanced Networks:** In dynamic wireless environments, acquiring instantaneous Channel State Information (CSI) can be challenging and resource-intensive, particularly for the IRS, which has a large number of elements. A promising research direction involves leveraging statistical CSI to guide the configuration of the IRS. This approach can mitigate the need for constant channel updates, reduce overhead, and enable more efficient IRS operations. Developing frameworks and algorithms that can effectively utilize statistical CSI to optimize IRS settings will be crucial. This strategy not only enhances network performance under practical constraints but also aligns with the envisioned 6-th Generation (6G) networks' emphasis on intelligence and efficiency [170, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333].
- **Advanced Material Science for IRS Enhancement — Expanding the Boundaries of Wireless Communication:** The exploration of novel materials and metamaterials presents an exciting frontier in enhancing the efficiency and responsiveness of IRS. By delving into materials boasting superior refractive indices, minimal loss factors, and the ability to dynamically alter electromagnetic properties in response to external stimuli, such as electrical or thermal changes, we stand on the verge of revolutionizing IRS technology. The key lies in understanding the intricate interactions of these advanced materials across diverse frequency spectra, particularly in the mmWave and terahertz domains. This endeavor could usher in a new era of IRS designs, offering unparalleled precision and control over the behavior of electromagnetic waves. As we envision the future integration of IRS technology into the very fabric of our built environment, one transformative application emerges: the incorporation of IRS into the facades of bundling or skyscrapers, as shown in Fig 9.2. Consider the glass industry, a sector that has seen minimal innovation over decades. The potential for IRS to become an integral component of high-rise building glass represents a profound shift, not just for wireless communications but also for architectural design and functionality. Such integration would not only enhance the aesthetic appeal of these structures but also turn them into active participants in the wireless ecosystem, significantly boosting signal quality and network coverage in urban landscapes. This vision for IRS-embedded building facades invites us to reimagine the possibilities for urban

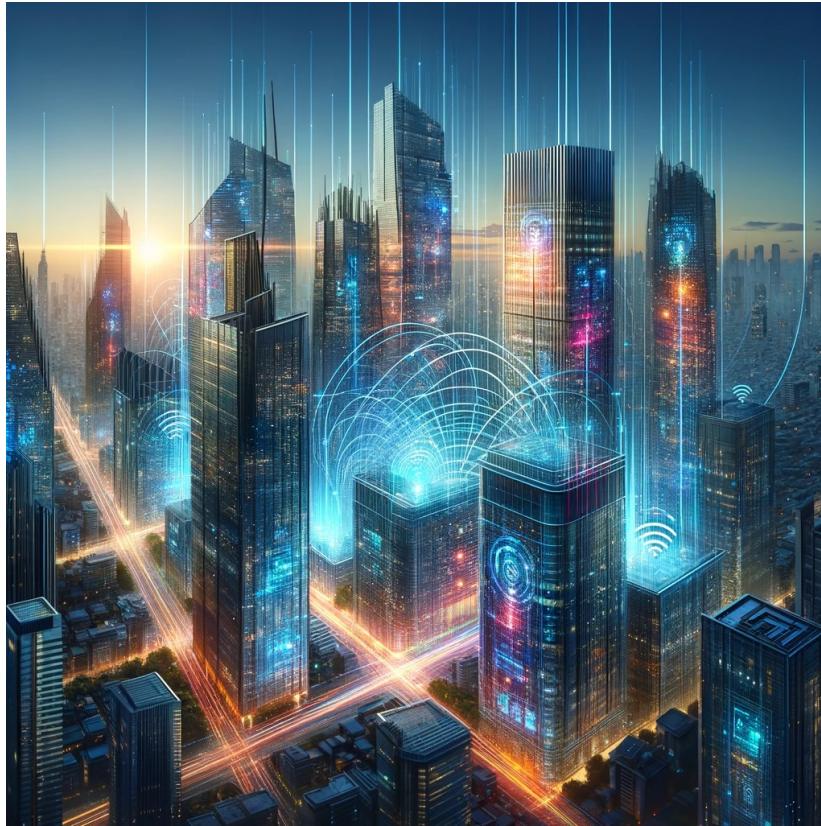


Figure 9.2: *Envisioning the Future of Connectivity: A cityscape where skyscrapers are embedded with Intelligent Reflecting Surfaces (IRS), enhancing the coverage and the boundaries of next-generation wireless networks.* — Generated by OpenAI.

development and connectivity. By transforming skyscrapers into giant, interactive nodes within the wireless network, we could effectively eliminate coverage dead zones and dramatically improve the efficiency of urban communications infrastructure. This is the future we should strive for, a world where advanced material science and IRS technology converge to redefine the boundaries of wireless communication and urban living [334, 335, 336, 337, 338, 339, 340, 341].

- **Integration of IRS with Non-Terrestrial Networks (NTNs):** Exploring the potential of IRS to enhance Non-Terrestrial Networks (NTNs), including satellite and UAV communication systems. IRS could be used to improve ground-satellite or UAV communication links by mitigating signal attenuation and interference, thus enabling more robust and extensive coverage. Research could focus on the design and placement of IRS on terrestrial structures or directly on airborne/spaceborne platforms to optimize the link quality in NTNs [342, 343, 344].

- **Energy Harvesting and Self-Sustaining IRS Systems:** Developing self-sustaining IRS panels through the integration of energy-harvesting technologies. This includes leveraging solar panels, piezoelectric materials, or RF energy harvesting to power the IRS's active components, such as tunable elements and sensors. Research in this area would contribute to deploying IRS in remote or power-constrained environments, expanding their applicability and sustainability. Enhancing WPT and energy harvesting techniques with IRS technology could significantly improve wireless networks' sustainability and autonomy. Research could focus on developing IRS-assisted WPT systems that maximize energy transfer efficiency across various environments and distances. Additionally, investigating the IRS's role in ambient energy harvesting from diverse sources (e.g., solar, electromagnetic) could further augment the energy efficiency of wireless devices, extending their operational lifespan and reducing their environmental impact. This could include reinforcement learning algorithms that continuously learn and improve IRS settings for optimal performance, predictive modeling to anticipate network demands and unsupervised learning techniques for clustering and anomaly detection within IRS-enhanced networks [345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357].
- **Advanced Optimization Techniques for IRS Placement and Configuration — Polyhedron Approach and AI-Driven Solutions:** In our pursuit to further refine and enhance the operational efficiency of IRS within wireless networks, we have identified and are exploring several cutting-edge optimization techniques. Beyond the traditional methods, the authors have envisioned the use of the *Polyhedron* technique as a particularly intriguing approach for determining optimal IRS placements. This novel Polyhedron approach, as shown in Fig. 9.3, which has not been studied in the realm of IRS optimization, is currently under rigorous investigation in our team. Its potential extends beyond merely strategic IRS placements; it could revolutionize the planning and deployment of other network infrastructures, such as access points, base stations, UAVs, and drones. By employing this method, we anticipate uncovering new dimensions of network optimization that have remained undeveloped, thereby enhancing overall network performance and efficiency.

Furthermore, we recognize the power of Artificial Intelligence (AI) and Machine Learning (ML) in optimizing IRS configurations. AI and ML algorithms are at the forefront of enabling intelligent, dynamic, and autonomous optimization of IRS configurations. The development of AI-driven predictive models is central to our approach, allowing for the accurate anticipation of optimal IRS settings in response to fluctuating environmental conditions, user movements, and varying network demands. By leveraging AI and ML, we aim to substantially reduce the complexity and computational demands traditionally associated with managing and optimizing large-scale IRS deployments. These AI-driven optimization strategies are not only fascinating for their ability to adapt and learn but also for their potential to offer innovative solutions for resource allocation decision policies across the network. Combining the exploratory potential of the Polyhedron approach with the dynamic adaptability of AI and ML methods opens up a broad spectrum of possibilities for IRS optimization. These advancements promise to elevate the capabilities of wireless networks, ensuring that they are more

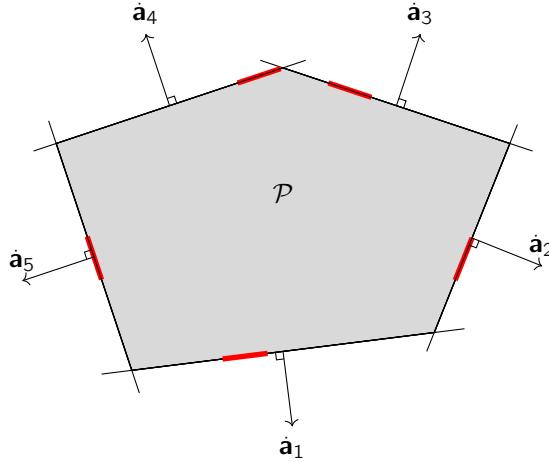


Figure 9.3: Illustration of a polyhedron \mathcal{P} as the intersection of five half-spaces, with outward normal vectors $\dot{\mathbf{a}}_1, \dots, \dot{\mathbf{a}}_5$. The red segments are the possible locations of IRSs in each half-space.

efficient, responsive, and capable of meeting the ever-growing demands of modern communication systems. As we delve deeper into these optimization techniques, our goal is to pave the way for a future where wireless networks are not only more interconnected but also significantly more intelligent and adaptable to the needs of their users [358, 359, 360, 361, 362, 363, 364, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383].

- **Integration of IRS with Quantum Communication:** Investigating the potential integration of IRS with quantum communication technologies presents a novel research avenue. Quantum communication promises unparalleled security and data transmission capabilities. Exploring how IRS can be optimized to improve quantum signal fidelity and extend quantum communication range could potentially enable secure and efficient wireless networks that are resistant to eavesdropping and offer high data integrity. This integration could involve studying the effects of IRS on quantum entanglement distribution and Quantum Key Distribution (QKD) protocols [384, 377, 385, 386, 387, 388, 389, 390].

These research directions are a short list of possible continuations or improvements with respect to the contribution of this dissertation. They highlight the importance of advancing IRS technology to meet the evolving demands of future wireless networks. By addressing the unique challenges associated with high-frequency operation, near-field interactions, optimization complexity, and CSI utilization, the research community can pave the way for more sophisticated, efficient, and adaptable IRS solutions. These advancements will undoubtedly contribute to the broader goal of creating more robust, high-performing wireless ecosystems that leverage the full potential of IRS technology.

Optimization Techniques

In this appendix, we introduce the foundational principles of convex functions, which are central to the thematic exploration of this thesis. Convex functions are distinguished by a defining characteristic: any line segment joining two points on the graph of the function remains either above or precisely on the graph, a property that significantly influences their application in optimization and mathematical modeling. These functions are integral in various fields, including economics, engineering, and machine learning, due to their desirable properties that facilitate problem-solving and analysis.

Understanding the basic properties of convex functions, such as the non-negativity of the second derivative (in the one-dimensional case or positive semi-definiteness of the Hessian in the high-dimensional cases), Jensen's inequality, and the Duality theorem, equips us with the analytical tools needed to approach complex optimization problems with greater efficacy. These properties ensure that local minima are also global minima, simplifying the optimization process and making it more predictable and efficient.

The exploration of convex optimizations in this context is inspired by the seminal work of Boyd and Vandenberghe (2004) in their comprehensive guide on convex optimization [197]. Their work provides a solid foundation for appreciating the importance of convexity in optimization theory and its applications across various disciplines.

A.1 Convex Analysis

Convex analysis is a branch of mathematics that studies convex sets and convex functions. At its core, convex analysis focuses on the properties and behaviors of convex functions, which are functions where the line segment between any two points on the graph of the function lies above or on the graph itself. This characteristic leads to numerous useful properties, such as the existence of unique global minima for optimization problems under certain conditions [391, 197].

A.1.1 Definitions

Let $f(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}$ be a convex function. Then, $f(\cdot)$ is convex, if for each $\lambda \in [0, 1]$, we have:

$$f(\lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2) \leq \lambda f(\mathbf{x}_1) + (1 - \lambda) f(\mathbf{x}_2), \quad (\text{A.1})$$

for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$. Geometrically speaking, the above inequality states that the line segment between $(\mathbf{x}_1, f(\mathbf{x}_1))$ and $(\mathbf{x}_2, f(\mathbf{x}_2))$, which is the chord from \mathbf{x}_1 to \mathbf{x}_2 , lies on top of the graph of $f(\cdot)$. A function $f(\cdot)$ is said to be *strictly convex* if (A.1) holds with strict inequality. Moreover, supposing that $f(\cdot)$ is differentiable, i.e., its gradient exists, then each convex function satisfies the following inequality:

$$f(\mathbf{x}) \geq f(\tilde{\mathbf{x}}) + \nabla_{\mathbf{x}} f(\tilde{\mathbf{x}})^T (\mathbf{x} - \tilde{\mathbf{x}}), \quad (\text{A.2})$$

where $\nabla_{\mathbf{x}}$ is the gradient vector with respect to \mathbf{x} at $\tilde{\mathbf{x}}$, and \square^T is the transpose operation on \square . The inequality (A.2), known as the first-order condition for convexity, asserts that for a convex function, a global underestimator of the function can be easily derived via its first-order Taylor approximation. Consequently, the first-order Taylor approximation of a convex function is always a global underestimator of the function. This inequality additionally confirms that global information of a convex function can be obtained through its local information, i.e., its value and derivative at a point.

A.2 Duality Theorem

Duality is a powerful concept in mathematical optimization that provides a framework for understanding and solving optimization problems from a different perspective. The theory of duality involves the formulation of a secondary problem, known as the dual problem, which is intrinsically linked to the original optimization problem, referred to as the primal problem. The relationships between the solutions of these two problems offer deep insights into the nature of the optimization problem in question.

The duality theorem, a cornerstone of this theory, states that under certain conditions, the solution to the dual problem provides bounds on the solution to the primal problem. For convex optimization problems, these conditions are often satisfied, and the duality theorem assures that the optimal value of the primal problem is equal to the optimal value of the dual problem. This equivalence is known as strong duality. Let's now mathematically define this theorem.

We consider the following optimization problem, also known as the *primal problem*, written in its general form as:

$$\begin{aligned} & \min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) \\ & \text{s.t. : } g_i(\mathbf{x}) \leq 0, \quad \forall i = 1, \dots, I, \\ & \quad h_i(\mathbf{x}) = 0, \quad \forall i = 1, \dots, L, \end{aligned} \quad (\text{A.3})$$

where $f(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}$ is the objective function, and $\mathbf{x} \in \mathbb{R}^n$ is the vector of optimization variables inside the feasible set \mathcal{X} . This optimization problem has I inequality constraints

and L equality constraints. Furthermore, we also refer to p^* as the optimal value of the optimization problem in (A.3).

The Lagrangian duality of the objective function of (A.3) is given by:

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\nu}) = f(\mathbf{x}) + \sum_{i=1}^I \mu_i g_i(\mathbf{x}) + \sum_{l=1}^L \nu_l h_l(\mathbf{x}), \quad (\text{A.4})$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ are called the vector of *Lagrangian multiplier* or the *dual variables* with respect to inequality and equality constraints associated with the problem (A.3) that have μ_i 's and ν_l 's as the elements of the corresponding vectors. The essential purpose of Lagrangian duality is to get somehow rid of the constraints in (A.3) by adding a weighted sum of the constraint functions to the objective function. We can now define the corresponding *Lagrange dual function* (or just *dual function*), which is formally stated as:

$$\mathcal{D}(\boldsymbol{\mu}, \boldsymbol{\nu}) = \inf_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\nu}). \quad (\text{A.5})$$

Note that even though the primal problem could be non-convex, the dual problem is always a convex optimization problem since the dual function is a point-wise infimum. This infimum can be seen as the greatest lower bound of a family of affine functions with respect to $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$.

The Lagrange dual function in (A.5) gives us a lower bound on the optimal value p^* of the primal problem (A.3). In order to find the best lower bound for the primal problem, the following optimization problem can be defined from the Lagrange dual function:

$$\max_{\boldsymbol{\mu}, \boldsymbol{\nu}} \mathcal{D}(\boldsymbol{\mu}, \boldsymbol{\nu}). \quad (\text{A.6})$$

This problem is known as the *Lagrange dual problem* corresponding to the primal problem. Moreover, if $\boldsymbol{\mu}^*$ and $\boldsymbol{\nu}^*$ are the optimal values for the Lagrange dual problem in (A.6), they are traditionally called *dual optimal* or *optimal Lagrange*. It should also be noted since the objective to be maximized is concave in (A.6), the Lagrange dual problem is a convex optimization problem no matter the primal problem in (A.3) is convex or not.

A.2.1 Weak Duality and Duality Gap

Weak duality refers to the relationship between the optimal solutions of the primal and dual problems. Specifically, in the context of a minimization problem, it states that the objective function value of any feasible solution to the dual problem provides an upper bound on the objective function value of any feasible solution to the primal problem. For maximization problems, this relationship is reversed, and the dual provides a lower bound on the primal. Mathematically, let \mathbf{x}^* be a feasible solution for the primal problem, i.e., p^* and $(\boldsymbol{\mu}^*, \boldsymbol{\nu}^*)$ are a feasible solution to the dual problem, that is, d^* . According to weak duality, we have the following inequality for a general (possibly non-convex) problem:

$$d^* \leq p^*. \quad (\text{A.7})$$

It must be noted that the weak duality inequality also holds when d^* and p^* are infinite. On the other hand, the difference between the primal optimal value and dual optimal value, i.e.,

$p^* - d^*$ is called the *optimal duality gap*. It should be stated that the optimal duality gap is always non-negative. Since the dual problem is always convex, and often can be solved efficiently to determine d^* , the inequality in (A.7) is quite useful in finding a lower bound on the optimal value of a problem that is difficult to solve¹.

A.2.2 Strong Duality and Slater condition

The significance of strong duality and the Slater condition lies in their ability to guarantee that the optimal value of a primal optimization problem is equal to the optimal value of its dual problem. This equivalence facilitates the solution of the primal problem by solving its dual, which can be computationally more efficient or analytically more tractable in certain cases. Furthermore, strong duality provides a powerful framework for sensitivity analysis and for deriving optimality conditions, which are pivotal for designing and analyzing algorithms for solving optimization problems. Let us now define them formally.

If the duality gap is zero, i.e., $p^* = d^*$, the *strong duality* holds. The strong duality indicates that the best bound that can be achieved from the Lagrange dual function is tight. Moreover, in strong duality, since the gap between primal and dual is zero, solving the dual problem is equivalent to solving the primal problem.

A sufficient condition for strong duality to hold for a convex optimization problem is the *Slater condition* or *Slater's condition*. In particular, if the Slater condition holds for the primal problem, then the duality gap is zero, which implies strong duality for convex problems. And if the dual optimal value is finite, then it is attained, i.e., a dual feasible (μ^*, ν^*) exists that satisfies $\mathcal{D}(\mu^*, \nu^*) = d^* = p^*$. In general, there exist many results that establish conditions on the optimization problem that yield strong duality. These conditions are coined *constraint qualifications*, where the Slater condition is only a simple specific example of many.

A.3 Lagrangian

In order to study duality in optimization models, two approaches exist historically, and the duality results are manifested as referred to as: i) *Classical Lagrangian* and ii) *Abstract Lagrangian*.

Classical Lagrangian: In classical optimization, particularly in the context of calculus of variations and classical mechanics, the Lagrangian refers to a function that describes the dynamics of a system. It is defined as the difference between the kinetic and potential energies of the system. In optimization, the classical Lagrangian $L(x, \lambda)$ for a problem is constructed by adding the product of Lagrange multipliers (λ) and the constraint functions

¹In an ideal scenario, especially for convex optimization problems, this gap is zero, indicating that the solutions to the primal and dual problems coincide (this is known as strong duality). However, in non-convex problems or in instances where certain regularity conditions are not met, there may be a positive duality gap. This means that the best solutions to the primal and dual problems do not achieve the same objective value, and the gap quantifies the difference between these two values. A nonzero duality gap indicates a lack of optimality or a limitation in the tightness of the dual problem as a bound on the primal problem.

to the objective function. This approach is used to find the stationary points of L under the constraints, which correspond to the optimal solutions of the original problem.

Abstract Lagrangian: The concept of an abstract Lagrangian extends the classical idea to a broader set of problems, including those in convex analysis and functional analysis. It involves constructing a Lagrangian function that incorporates both the objective function and the constraints in a more general form, allowing for the analysis and solution of more complex or abstract optimization problems.

Among these two forms, the classical Lagrangian form is more extensively used in the literature. What we have discussed so far is indeed the classical Lagrangian form of duality. As seen, classical Lagrangian typically starts from a primal problem while the Lagrangian and the Dual Lagrangian problems are established. However, at a more abstract level, an abstract Lagrangian function is used to derive the primal and dual optimization problems. Here, we briefly discuss an abstract version of Lagrangian duality that is elaborated in more significant detail in [392]. In this version, through a certain real-valued abstract Lagrangian function, the primal and dual costs are taken into account, such that:

$$\begin{aligned} \text{(Primalproblem)} \quad & \min_{\mathbf{x} \in \mathcal{X}} \mathcal{F}(\mathbf{x}) \quad \text{where } \mathcal{F}(\mathbf{x}) = \sup_{\mathbf{y} \in \mathcal{Y}} \mathcal{L}(\mathbf{x}, \mathbf{y}), \\ \text{(Dualproblem)} \quad & \max_{\mathbf{y} \in \mathcal{Y}} \mathcal{G}(\mathbf{y}) \quad \text{where } \mathcal{G}(\mathbf{y}) = \inf_{\mathbf{x} \in \mathcal{X}} \mathcal{L}(\mathbf{x}, \mathbf{y}), \end{aligned}$$

where $\mathcal{L} : \mathcal{X} \times \mathcal{Y} \longrightarrow R$ is the abstract Lagrangian function pertaining to \mathcal{X} and \mathcal{Y} as appropriate domains defined in some primal and dual spaces, respectively. Moreover, the supremum can be seen as the least upper bound of a family of affine functions with respect to \mathbf{x} and \mathbf{y} . This approach to duality is based on conjugate duality, where a convexity assumption is always made [393]. This approach also puts a strong emphasis on the minimax and saddle point theorems, which are given below.

- **Minimax Theorem:** This theorem provides the condition that guarantees the *strong max-min property* or the *saddle point* as follows:

$$\sup_{\mathbf{y} \in \mathcal{Y}} \inf_{\mathbf{x} \in \mathcal{X}} \mathcal{H}(\mathbf{x}, \mathbf{y}) = \inf_{\mathbf{x} \in \mathcal{X}} \sup_{\mathbf{y} \in \mathcal{Y}} \mathcal{H}(\mathbf{x}, \mathbf{y}). \quad (\text{A.8})$$

It should be noted that the above equality, strong max-min property, holds only in special cases. This is, in particular, true, when for example, $\mathcal{H} : \mathcal{X} \times \mathcal{Y} \longrightarrow R$ is the Lagrangian of a problem where the strong duality holds.

- **Saddle Point Theorem:** Under suitable conditions, there exists a *saddle point* for $\mathcal{S}(\cdot)$ referred to as a pair $(\mathbf{x}^*, \mathbf{y}^*) \in \mathcal{X} \times \mathcal{Y}$ such that for all $(\mathbf{x}, \mathbf{y}) \in \mathcal{X} \times \mathcal{Y}$:

$$\mathcal{S}(\mathbf{x}^*, \mathbf{y}) \leq \mathcal{S}(\mathbf{x}^*, \mathbf{y}^*) \leq \mathcal{S}(\mathbf{x}, \mathbf{y}^*). \quad (\text{A.9})$$

In (A.9), $\mathcal{S} : \mathcal{X} \times \mathcal{Y} \longrightarrow R$ is the Lagrangian of a problem where the strong duality holds. In other words, $\mathcal{S}(\mathbf{x}^*, \mathbf{y}^*) = \sup_{\mathbf{y} \in \mathcal{Y}} \mathcal{S}(\mathbf{x}^*, \mathbf{y})$, and $\mathcal{S}(\mathbf{x}^*, \mathbf{y}^*) = \inf_{\mathbf{x} \in \mathcal{X}} \mathcal{S}(\mathbf{x}, \mathbf{y}^*)$. This indicates that the strong max-min property (A.8) holds with the common value of $\mathcal{S}(\mathbf{x}^*, \mathbf{y}^*)$.

A.3.1 Augmented Lagrangian Method

The augmented Lagrangian method is an extension of the classical Lagrangian technique, designed to improve the convergence properties when solving constrained optimization prob-

lems, especially those with equality and inequality constraints. The augmented Lagrangian function includes an additional term that penalizes violations of the constraints, effectively augmenting the original Lagrangian with a quadratic penalty term. This method works by iteratively solving an unconstrained optimization problem for the augmented Lagrangian and updating the Lagrange multipliers and penalty parameters until convergence to an optimal solution of the original constrained problem is achieved [394]. Consider a constrained optimization problem of the form:

$$\begin{aligned} & \min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) \\ \text{s.t. : } & g_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m, \\ & h_j(\mathbf{x}) = 0, \quad j = 1, \dots, p, \end{aligned} \quad (\text{A.10})$$

where $f(\mathbf{x})$ is the objective function to be minimized over the variable $\mathcal{X} \in \mathbb{R}^n$, $g_i(\mathbf{x})$ are inequality constraints, and $h_j(\mathbf{x})$ are equality constraints. The Augmented Lagrangian $\mathcal{L}_A(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}, r)$ for this problem is given by:

$$\begin{aligned} \mathcal{L}_A(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}, r) &= f(\mathbf{x}) \\ &+ \sum_{i=1}^m \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^p \mu_j h_j(\mathbf{x}) \\ &+ \frac{r}{2} \sum_{i=1}^m \max(0, g_i(\mathbf{x}))^2 + \frac{r}{2} \sum_{j=1}^p h_j(\mathbf{x})^2, \end{aligned} \quad (\text{A.11})$$

where $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)$ and $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)$ are vectors of Lagrange multipliers for the inequality and equality constraints, respectively, and $r > 0$ is a penalty parameter. The key features of the Augmented Lagrangian method include:

Penalty for Constraint Violation: The addition of $\frac{r}{2} \max(0, g_i(\mathbf{x}))^2$, the quadratic penalty terms, for inequality constraints and $\frac{r}{2} h_j(\mathbf{x})^2$ for equality constraints. These terms impose a penalty on the violation of constraints, which becomes more severe as the value of r increases.

Adaptive Penalty Parameter: The penalty parameter r is typically updated (usually increased) iteratively, which helps in driving the solution towards feasibility with respect to the constraints.

Dual Update: The Lagrange multipliers $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are updated at each iteration based on the degree of violation of the constraints, facilitating the convergence towards the optimal dual variables.

The Augmented Lagrangian method iteratively solves a sequence of unconstrained or easier-to-handle constrained optimization problems, adjusting the penalty parameter r and updating the Lagrange multipliers as it progresses. This approach effectively bridges the gap between feasibility and optimality, ensuring that the solution satisfies both the objective function minimization and the constraints. The augmented Lagrangian method is particularly useful for problems where direct application of the classical Lagrangian method is difficult due to the nature of the constraints or where the convergence of classical methods is slow. By penalizing constraint violations more strongly, the augmented Lagrangian method often leads to faster convergence and more robust solutions, making it a powerful tool in the optimization toolkit.

A.4 Complementary Slackness and KKT Optimality Conditions

Complementary Slackness and the Karush-Kuhn-Tucker (KKT) are integral to both the theoretical understanding of optimization and its practical applications. Complementary Slackness provides a direct link between the primal and dual formulations of an optimization problem. It defines that for every constraint in the problem, the product of the constraint's Lagrange multiplier and the slack in the constraint must equal zero. This implies that if a constraint is not active (meaning it does not directly influence the optimal solution), its corresponding Lagrange multiplier is zero, highlighting which constraints are critical at the optimum.

Mathematically, suppose that both the primal and dual optimal values exist and are equal. This means the strong duality holds. We further assume that \mathbf{x}^* and $(\boldsymbol{\mu}^*, \boldsymbol{\nu}^*)$ to be a primal optimal and a dual optimal point, respectively. Therefore, we have

$$f(\mathbf{x}^*) = \mathcal{D}(\boldsymbol{\mu}^*, \boldsymbol{\nu}^*) \leq f(\mathbf{x}^*) + \sum_{i=1}^I \mu_i^* g_i(\mathbf{x}^*) + \sum_{l=1}^L \nu_l^* h_l(\mathbf{x}^*) \leq f(\mathbf{x}^*). \quad (\text{A.12})$$

The first inequality in (A.12) holds since the infimum of the Lagrangian over \mathbf{x} is less than or equal to its value at $\mathbf{x} = \mathbf{x}^*$. However, the last inequality follows from $\mu_i^* \geq 0$, $g_i(\mathbf{x}^*) \leq 0$, $\forall i = \{1, \dots, I\}$, and $h_l(\mathbf{x}^*) \leq 0$, $\forall l = 1, \dots, L$. An important conclusion that one can make from (A.12) is that

$$\mu_i^* g_i(\mathbf{x}^*) = 0, \quad \forall i = 1, \dots, I. \quad (\text{A.13})$$

This condition is called the *complementary slackness*. It confirms that one can go from the optimal primal solution to the optimal dual solution, and vice versa, if the strong duality holds. Moreover, the complementary slackness verifies that a solution is optimal, by checking if there is a dual solution [395, 396].

The KKT conditions extend the idea of Lagrange multipliers to include inequality constraints, offering a comprehensive set of criteria that must be met for a solution to be deemed optimal. These conditions encapsulate stationarity, primal and dual feasibility, and complementary slackness. In essence, the KKT conditions ensure that the gradient of the objective function, adjusted for the weighted gradients of the constraints, vanishes at the optimum, affirming that no feasible direction can lead to improvement. They also ensure that all constraints are satisfied (primal feasibility), the multipliers for inequality constraints are non-negative (dual feasibility), and complementary slackness holds.

Mathematically, suppose that all the functions both in the objective and the constraints in (A.3) are differentiable. Just the same as was assumed in (A.12), let's also suppose the primal and dual variables at the optimum point, for which strong duality obtains, are \mathbf{x}^* and

$(\boldsymbol{\mu}^*, \boldsymbol{\nu}^*)$, respectively. The KKT conditions have the following properties

$$g_i(\mathbf{x}^*) \leq 0, \quad \forall i = 1, \dots, I, \quad (\text{A.14a})$$

$$h_l(\mathbf{x}^*) = 0, \quad \forall l = 1, \dots, L, \quad (\text{A.14b})$$

$$\mu_i^* \geq 0, \quad \forall i = 1, \dots, I, \quad (\text{A.14c})$$

$$\mu_i^* g_i(\mathbf{x}^*) = 0, \quad \forall i = 1, \dots, I, \quad (\text{A.14d})$$

$$\nabla_{\mathbf{x}} f(\mathbf{x}^*) + \sum_{i=1}^I \mu_i^* \nabla_{\mathbf{x}} g_i(\mathbf{x}^*) + \sum_{l=1}^L \nu_l^* \nabla_{\mathbf{x}} h_l(\mathbf{x}^*) = 0, \quad (\text{A.14e})$$

where μ_i^* and ν_l^* are the elements of Lagrangian vectors $\boldsymbol{\mu}^*$ and $\boldsymbol{\nu}^*$, respectively. Also, $\nabla_{\mathbf{x}}$ denotes the gradient of a function with respect to \mathbf{x} in (A.14e). Note that the KKT conditions are necessary and sufficient conditions for the optimality of the convex optimization problem with differentiable objective and constraint functions. However, if the problem is non-convex, the KKT conditions would only provide the necessary conditions for optimality, given that the objective and constraints are differentiable.

In practice, the KKT conditions and complementary slackness are used to identify optimal solutions to constrained optimization problems. By examining these conditions, one can determine whether a candidate solution is truly optimal. In algorithm design, for instance, these conditions guide the iterative steps towards an optimal solution, ensuring that adjustments to the variables respect the constraints and move towards satisfying the KKT conditions. Moreover, in sensitivity analysis, they provide insights into how changes in the problem's parameters might affect the solution, based on the relationship between the primal and dual problems elucidated by complementary slackness. The KKT conditions and complementary slackness not only facilitate the identification and verification of optimal solutions but also enrich the understanding of the problem's structure and the interplay between its constraints and objectives [397].

A.5 Interior-Point Methods

Interior-Point Methods are a class of algorithms designed to solve linear and nonlinear optimization problems, particularly those involving constraints. These methods are distinguished from other optimization techniques, like simplex methods for linear programming, by their approach of traversing the interior of the feasible region to reach an optimal solution, rather than moving along the boundary.

The term “interior-point” refers to the strategy of these methods to start from a point within the interior of the feasible region of the optimization problem and iteratively move towards the optimal solution, while staying within the feasible region. This is in contrast to methods that operate on the boundary of the feasible region or explore the vertices of the feasible region, such as the simplex method in linear programming.

Developed in the mid-1980s, Interior-Point Methods gained prominence through Karmarkar's algorithm for linear programming. They have since been extended to various types of optimization problems, including nonlinear programming, semidefinite programming, and convex optimization. The key idea behind these methods is to solve a sequence of approximations of

the original problem that makes the barrier to leaving the feasible region infinite, effectively "pushing" the solution towards optimality while remaining in the interior [398].

To understand this approach, let's consider a basic form of an optimization problem and how Interior-Point Methods are applied, especially focusing on linear optimization for clarity. A typical linear optimization (Linear Programming (LP)) problem can be formulated as follows:

$$\begin{aligned} & \min_{x \in \mathcal{X}} c^T x \\ & \text{s.t. : } Ax = b, \\ & \quad x \geq 0, \end{aligned} \tag{A.15}$$

where $x \in \mathbb{R}^n$ is the vector of decision variables, $c \in \mathbb{R}^n$ is the coefficient vector for the objective function, A is a $m \times n$ matrix representing the linear constraints, and $b \in \mathbb{R}^m$ is the right-hand side vector. The goal is to find the vector x that minimizes the objective function while satisfying the constraints.

Interior-Point Methods solve such a problem by starting from a point within the feasible region (hence "interior") and iteratively moving towards the optimum. One key concept used in these methods is the logarithmic barrier function, which allows the method to handle the non-negativity constraints $x \geq 0$ by incorporating them into the objective function. An example of a barrier function that might be added to the objective is:

$$-\mu \sum_{i=1}^n \ln(x_i), \tag{A.16}$$

where x_i are the components of the decision variable vector x . The modified objective function becomes:

$$\min \quad c^T x - \mu \sum_{i=1}^n \ln(x_i), \tag{A.17}$$

where μ is a positive parameter controlling the influence of the barrier. As the algorithm progresses, μ is gradually reduced, steering the solution toward the boundary of the feasible region and the optimal solution. At each iteration, the method solves a system of equations derived from the KKT conditions, adjusting the variables and μ to move closer to the optimum. The KKT conditions for the modified problem incorporate both the gradient of the objective function (including the barrier term) and the feasibility conditions. The central path, which is a trajectory that the solutions follow as μ decreases, leads to the optimal solution as $\mu \rightarrow 0$, ensuring that the iterates remain in the interior of the feasible region.

This mathematical framework allows Interior-Point Methods to efficiently navigate the feasible space, exploiting the curvature of the barrier function to avoid the pitfalls of boundary navigation and to leverage more direct paths toward the optimum. This approach is not only applicable to linear programming but also extends to nonlinear and convex optimization problems, where similar barrier methods are utilized to handle constraints and guide the solution process.

Algorithm 10 Interior-Point Method Algorithm with Barrier Function for LP1: **Initialize**

Choose an initial feasible point \mathbf{x}_0 such that $\mathbf{A}\mathbf{x}_0 = \mathbf{b}$ and $\mathbf{x}_0 > 0$.

Set an initial barrier parameter $\mu_0 > 0$ and a parameter $\theta \in (0, 1)$ to reduce μ in each iteration.

2: **while** not converged **do**

3: For the current value of μ , solve (A.17).

4: Reduce the barrier parameter for the next iteration, typically using $\mu \leftarrow \theta\mu$.

5: **until** some convergence criterion is met

6: **return** optimal \mathbf{x}

A.6 MM Approach and DC Programming

MM algorithms are an appropriate tool to reduce a given optimization problem into a series of simpler problems. In this sense, an MM algorithm is not an algorithm, but rather an appropriate principal way of designing optimization algorithms for high dimensional settings, where the classical methods of optimization do not work well. MM algorithms are not new. The celebrated Expectation Maximization algorithm is a particular case of MM algorithms that is extensively used in electrical engineering applications and in other fields. The reason for selecting the MM acronym is two-fold. An MM algorithm operates on a more straightforward and simpler surrogate function that majorizes/minorizes (the first M of MM) the objective function in a minimization/maximization (the second M of MM) optimization problem. Thus, the MM stands for either *Majorization Minimization* or *Minorization Maximization*, depending on the application. In the next few paragraphs, we consider a majorization minimization problem to explain how the algorithm works. The process involves two steps:

Majorization: Given the current estimate, find a surrogate function that is easier to optimize and that upper bounds the original objective function while being tangent to it at the current estimate.

Minimization: Minimize this surrogate function to obtain a new estimate. This sequence of majorization and minimization is repeated until convergence. The MM algorithm is particularly useful because it can convert non-convex problems into convex optimization problems at each iteration, making it easier to find global or local optima.

Consider the following optimization problem

$$\min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}), \quad (\text{A.18})$$

where \mathbf{x} is the optimization variable vector belonging to the feasible set \mathcal{X} . In order to majorize the function $f(\mathbf{x})$ at \mathbf{x}^n , there exists a surrogate function $g(\mathbf{x}|\mathbf{x}^n)$ that satisfies two conditions

$$f(\mathbf{x}^n) = g(\mathbf{x}^n|\mathbf{x}^n), \quad (\text{A.19})$$

$$f(\mathbf{x}) \leq g(\mathbf{x}|\mathbf{x}^n), \quad \mathbf{x} \neq \mathbf{x}^n. \quad (\text{A.20})$$

The first condition (A.19) is called the tangency condition at the current iteration step. This condition guarantees $g(\mathbf{x}^n|\mathbf{x}^n)$ is tangent to $f(\mathbf{x})$ at \mathbf{x}^n . The second condition, on the

other hand, (A.20) makes sure the $g(\mathbf{x}|\mathbf{x}^n)$ is dominant in a sense that it always lies above the surface of $f(\mathbf{x})$ except at \mathbf{x}^n . Besides, if a function $g(\mathbf{x}|\mathbf{x}^n)$ majorizes the function $f(\mathbf{x})$ at \mathbf{x}^n , it can be easily perceived that $-g(\mathbf{x}|\mathbf{x}^n)$ minorizes $-f(\mathbf{x})$.

Another very important result of the MM algorithms is the descent property. Starting from $\mathbf{x}^0 \in \mathcal{X}$ as an initial point for the feasible set \mathcal{X} , an MM algorithm generates a sequence of feasible point \mathbf{x}^n . At point \mathbf{x}^n in the majorization step, a continuous surrogate function is constructed that satisfies the domination condition in (A.20)

$$g(\mathbf{x}|\mathbf{x}^n) \geq f(\mathbf{x}) + g(\mathbf{x}^n|\mathbf{x}^n) - f(\mathbf{x}^n), \quad \mathbf{x} \neq \mathbf{x}^n. \quad (\text{A.21})$$

Hence, in the minimization step, the following update rule can be applied

$$\mathbf{x}^{n+1} \in \min_{\mathbf{x} \in \mathcal{X}} g(\mathbf{x}|\mathbf{x}^n). \quad (\text{A.22})$$

It is easy to show that the generated sequence $f(\mathbf{x}^n)$ is non-increasing. Thus, we have

$$f(\mathbf{x}^{n+1}) \leq g(\mathbf{x}^{n+1}|\mathbf{x}^n) - g(\mathbf{x}^n|\mathbf{x}^n) + f(\mathbf{x}^n) \leq g(\mathbf{x}^n|\mathbf{x}^n) - g(\mathbf{x}^n|\mathbf{x}^n) + f(\mathbf{x}^n) = f(\mathbf{x}^n), \quad (\text{A.23})$$

where the first inequality comes from (A.21), and the second inequality is the direct consequence of (A.22). The property in (A.23), the descent property, gives a remarkable numerical stability to MM algorithms. Hence, instead of minimizing the cost function $f(\mathbf{x})$ directly, the MM algorithms stably optimize a sequence of tractable approximate surrogate objective functions $g(\mathbf{x}|\mathbf{x}^n)$ that minorize $f(\mathbf{x})$ as tightly as possible.

The MM algorithms can easily be connected to other algorithmic frameworks [207, 399, 400]. One of the application areas of the MM algorithms is in *Difference of Convex functions* (DC) programming problems. The general form of DC functions is

$$\min_{\mathbf{x}} f_0(\mathbf{x}) - h_0(\mathbf{x}) \quad (\text{A.24})$$

$$\text{s.t.}: f_i(\mathbf{x}) - h_i(\mathbf{x}) \leq 0, \quad \forall i = 1, \dots, m, \quad (\text{A.25})$$

where f_i 's and h_i 's are all convex functions. We further assume that f_i 's and h_i 's are twice differentiable, and are strictly convex without loss of generality according to (A.1). Among various algorithms having desirable properties for the solution of DC problems, the MM scheme, which solves a sequence of convex problems acquired by linearizing non-convex parts in the objective function as well as the constraints, is preferred. Accordingly, an approximate solution can be found that iteratively solves (A.24) through defining the following convex subproblem

$$\min_{\mathbf{x}} g_0(\mathbf{x}|\mathbf{x}^n) \quad (\text{A.26})$$

$$\text{s.t.}: g_i(\mathbf{x}|\mathbf{x}^n) \leq 0, \quad \forall i = 1, \dots, m, \quad (\text{A.27})$$

where

$$g_i(\mathbf{x}|\mathbf{x}^n) = f_i(\mathbf{x}) - \left(h_i(\mathbf{x}^n) + \nabla_{\mathbf{x}} h_i(\mathbf{x}^n)^T (\mathbf{x} - \mathbf{x}^n) \right), \quad \forall i \in \{0, \dots, m\}. \quad (\text{A.28})$$

The aforementioned approximation satisfies the MM principle and is a tight upper bound of $f_i - h_i$ with equality achieved at $\mathbf{x} = \mathbf{x}^n$. This technique is used several times throughout the thesis. Moreover, the solution methodology for the MM algorithm is summarized in **Algorithm 11**. A valid question to be asked at this point would be how good the convergence behaviors of the MM algorithms are. For the answer, the interested reader is referred to [401, 402, 403].

Algorithm 11 The MM (Majorization Minimization or Minorization Maximization) algorithm1: **Initialize**

Iteration index $n = 0$ with the maximum number of iteration N_{\max}
and find a feasible point \mathbf{x}^0 .

2: **repeat**

3: Find \mathbf{x}^n by solving the optimization problem (A.22) and store as \mathbf{x} .
4: Set $n = n + 1$ and $\mathbf{x}^n = \mathbf{x}$.

5: **until** some convergence criterion is met or $n = N_{\max}$

6: **return** optimal \mathbf{x}

A.7 Optimization Packages

The field of optimization encompasses a diverse array of tools and packages, each tailored to address the complexities of various optimization problems at hand. This overview briefly highlights some of the most prominent optimization packages while also shedding light on additional tools that play a crucial role in the optimization landscape.

The GNU Linear Programming Kit (GLPK) [404] stands out for its capacity to tackle large-scale Linear Programming (LP), Mixed Integer Programming (MIP), and related problem types, providing a robust framework for dealing with complex optimization tasks.

Gurobi[405] Optimizer emerges as a leading commercial solver, acclaimed for its efficiency in solving LP, Quadratic Programming (QP), and MIP challenges, including specialized forms such as Mixed Integer Linear Programming (MILP), Mixed Integer Quadratic Programming (MIQP), and Mixed Integer Quadratically Constrained Programming (MIQCP). Its reputation for speed and reliability makes it a preferred choice among professionals.

Mosek [406] offers another powerful optimization solution, with capabilities that extend to LP, QP, MIP, Second-Order Cone Programming (SOCP), and Semi-Definite Programming (SDP), making it a versatile tool for a wide range of optimization problems.

SeDuMi [407] and SDPT3 [408, 409] are distinguished for their focus on Semi-Definite Programming (SDP), offering advanced solutions for problems within this domain.

Beyond these mentioned tools, the optimization field is rich with other noteworthy packages, each designed to meet specific needs:

- **CVX** [410, 175]: A Matlab-based software for convex optimization that provides an easy-to-use interface for specifying and solving convex programs. It automates the transformation of high-level convex optimization problems into standard form, leveraging solvers like SeDuMi, SDPT3, and Mosek for efficient problem-solving.
- **CVXPY** [411]: Inspired by CVX, this Python-based framework offers similar functionalities, enabling users to formulate and solve convex optimization problems through a high-level API. It supports various solvers, including MOSEK, Gurobi, and SCIP, and extends its applicability to disciplines beyond Matlab's reach.
- **YALMIP** [412]: Primarily developed for Matlab, YALMIP is a high-level modeling tool

that simplifies the task of formulating and solving optimization problems. It acts as an interface between the user's problem formulation and the solver, capable of handling a broad spectrum of optimization tasks and integrating with several solvers, such as Gurobi, Mosek, and GLPK.

- **OSQP** [413] (Operator Splitting Quadratic Program (OSQP)): Specializes in solving large-scale quadratic programming problems, particularly useful in areas like control systems and finance. Its emphasis on scalability and efficiency makes it an essential tool for certain optimization challenges.
- **COIN-OR** [414] (Common Optimization INterface for Operations Research (COIN-OR)): An initiative to foster the development of open-source software for the operations research community, offering a collection of packages that cover various aspects of optimization, from linear and integer programming to more specialized areas.

The continuous evolution of optimization tools and packages, alongside the introduction of new solutions, ensures that researchers and practitioners have access to the cutting-edge technologies required to tackle problems of ever-increasing complexity. The goal of employing these tools is not to contribute to their development but to leverage their capabilities to address the specific optimization challenges at hand, underscoring a distinct domain of research dedicated to creating and refining these sophisticated solvers.

Bibliography

- [1] S. Bayat, J. Jalali, A. Khalili, M. R. Mili, S. Wittevrongel, and H. Steendam, "Optimal multi-objective resource allocation for D2D underlaying cellular networks in uplink communications," *IEEE Access*, vol. 9, pp. 114 153–114 166, July 2021.
- [2] A. Rezaei, A. Khalili, J. Jalali, H. Shafiei, and Q. Wu, "Energy-efficient resource allocation and antenna selection for irs-assisted multicell downlink networks," *IEEE Wireless Communications Letters*, vol. 11, no. 6, pp. 1229–1233, March 2022.
- [3] J. Jalali, A. Khalili, A. Rezaei, R. Berkvens, M. Weyn, and J. Famaey, "IRS-based energy efficiency and admission control maximization for IoT users with short packet lengths," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 9, pp. 12 379–12 384, September 2023.
- [4] J. Jalali, A. Khalili, H. Tabassum, R. Berkvens, J. Famaey, and W. Saad, "Energy-efficient THz NOMA for SWIPT-aided miniature UAV networks," *IEEE Communications Letters*, vol. 28, no. 5, pp. 1107–1111, May 2024.
- [5] J. Jalali, A. Khalili, H. Tabassum, J. Famaey, and W. Saad, "Placement, orientation, and resource allocation optimization for cell-free OIRS-aided OWC network," *IEEE Transactions on Vehicular Technology*, pp. 1–5, 2024, submitted.
- [6] J. Jalali, F. Lemic, H. Tabassum, and J. Famaey, "Is geometric polyhedron approach viable for optimal multi-IRS placement in upcoming 6G networks?" *IEEE Transactions on Vehicular Technology*, pp. 1–5, 2024, preprint and under submission.
- [7] J. Jalali, A. Kaya, M. Weyn, and R. Berkvens, "Activity monitoring at an intersection using a sub-ghz wireless sensor network," in *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*, 2021, pp. 1–6.
- [8] J. Jalali, A. Rezaei, A. Khalili, and J. Famaey, "Power-efficient joint resource allocation and decoding error probability for multiuser downlink MISO with finite block length codes," in *2022 25th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, 2022, pp. 232–237.
- [9] J. Jalali, A. Khalili, A. Rezaei, J. Famaey, and W. Saad, "Power-efficient antenna switching and beamforming design for multi-user swipt with non-linear energy harvesting," in *2023 IEEE 20th Consumer Communications & Networking Conference (CCNC)*, 2023, pp. 746–751.
- [10] J. Jalali, A. Khalili, A. Rezaei, and J. Famaey, "Is active irs Useful for mmWave wireless networks or not?" in *2023 International Conference on Computing, Networking and Communications (ICNC)*, 2023, pp. 377–382.

- [11] J. Jalali, A. Khalili, R. Berkvens, and J. Famaey, "Joint offloading policy and resource allocation in IRS-aided MEC for IoT users with short packet transmission," in *2023 IEEE 98th Vehicular Technology Conference (VTC2023-Fall)*, 2023, pp. 1–7.
- [12] J. Jalali, F. Lemic, H. Tabassum, R. Berkvens, and J. Famaey, "Toward energy efficient multiuser IRS-assisted URLLC systems: A novel rank relaxation method," in *2023 IEEE Globecom Workshops (GC Wkshps)*, 2023, pp. 1354–1360.
- [13] J. Jalali, M. Bustamante, F. Lemic, H. Tabassum, J. Struye, J. Famaey, and X. Costa-Pérez, "Location optimization and resource allocation of IRS in a multi-user indoor mmWave VR network," in *2024 IEEE Wireless Communications and Networking Conference (WCNC)*, Dubai, United Arab Emirates, April 2024, pp. 1–7.
- [14] F. Lemic, J. Jalali, G. C. Bartra, A. Amat, J. Struye, J. Famaey, and X. C. Perez, "Location-based real-time utilization of reconfigurable intelligent surfaces for mmwave communication and sensing in full-immersive multiuser VR," in *Advanced Metaverse Wireless Communication Systems*. IET, 2024, ch. 4, Accepted for Publication.
- [15] *IMT traffic estimates for the years 2020 to 2030*. Report ITU-R M.2370-0, July 2017.
- [16] S. J. Nawaz, S. K. Sharma, S. Wyne, M. N. Patwary, and M. Asaduzzaman, "Quantum machine learning for 6G communication networks: State-of-the-Art and vision for the future," *IEEE Access*, vol. 7, pp. 46 317–46 350, 2019.
- [17] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Communications Magazine*, vol. 58, no. 3, pp. 55–61, March 2020.
- [18] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, June 2014.
- [19] M. Shafi, A. F. Molisch, P. J. Smith, T. Haustein, P. Zhu, P. De Silva, F. Tufvesson, A. Benjebbour, and G. Wunder, "5G: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1201–1221, June 2017.
- [20] T. Taleb, N. Sehad, Z. Nadir, and J. Song, "VR-based immersive service management in B5G mobile systems: A UAV command and control use case," *IEEE Internet of Things Journal*, vol. 10, no. 6, pp. 5349–5363, March 2023.
- [21] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Network*, vol. 34, no. 3, pp. 134–142, May/June 2020.
- [22] H. Viswanathan and P. E. Mogensen, "Communications in the 6G era," *IEEE Access*, vol. 8, pp. 57 063–57 074, 2020.
- [23] E. Calvanese Strinati, S. Barbarossa, J. L. Gonzalez-Jimenez, D. Ktenas, N. Cassiau, L. Maret, and C. Dehos, "6G: The next frontier: From holographic messaging to artificial intelligence using subterahertz and visible light communication," *IEEE Vehicular Technology Magazine*, vol. 14, no. 3, pp. 42–50, September 2019.

[24] H. Tataria, M. Shafi, A. F. Molisch, M. Dohler, H. Sjöland, and F. Tufvesson, "6G wireless systems: Vision, requirements, challenges, insights, and opportunities," *Proceedings of the IEEE*, vol. 109, no. 7, pp. 1166–1199, July 2021.

[25] T. S. Rappaport, Y. Xing, O. Kanhere, S. Ju, A. Madanayake, S. Mandal, A. Alkateeb, and G. C. Trichopoulos, "Wireless communications and applications above 100 GHz: Opportunities and challenges for 6g and beyond," *IEEE Access*, vol. 7, pp. 78 729–78 757, 2019.

[26] F. Tariq, M. R. A. Khandaker, K.-K. Wong, M. A. Imran, M. Bennis, and M. Debbah, "A speculative study on 6G," *IEEE Wireless Communications*, vol. 27, no. 4, pp. 118–125, August 2020.

[27] M. Z. Chowdhury, M. Shahjalal, S. Ahmed, and Y. M. Jang, "6G wireless communication systems: Applications, requirements, technologies, challenges, and research directions," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 957–975, 2020.

[28] Z. Wan, Z. Gao, F. Gao, M. D. Renzo, and M.-S. Alouini, "Terahertz massive MIMO with holographic reconfigurable intelligent surfaces," *IEEE Transactions on Communications*, vol. 69, no. 7, pp. 4732–4750, July 2021.

[29] H. Taghvaei, S. Abadal, A. Pitilakis, O. Tsilipakos, A. C. Tasolamprou, C. Liaskos, M. Kafesaki, N. V. Kantartzis, A. Cabellos-Aparicio, and E. Alarcón, "Scalability analysis of programmable metasurfaces for beam steering," *IEEE Access*, vol. 8, pp. 105 320–105 334, 2020.

[30] H. Taghvaei, A. Pitilakis, O. Tsilipakos, A. C. Tasolamprou, N. V. Kantartzis, M. Kafesaki, A. Cabellos-Aparicio, E. Alarcón, and S. Abadal, "Multiwideband terahertz communications via tunable graphene-based metasurfaces in 6G networks: Graphene enables ultimate multiwideband THz waveform control," *IEEE Vehicular Technology Magazine*, vol. 17, no. 2, pp. 16–25, June 2022.

[31] R. Malleboina, J. C. Dash, and D. Sarkar, "Design of anomalous reflectors by phase gradient unit cell-based digitally coded metasurface," *IEEE Antennas and Wireless Propagation Letters*, vol. 22, no. 9, pp. 2305–2309, September 2023.

[32] A. Bagheri, M. Safaei, A. Araghi, S. M. M. Shahabi, F. Wang, M. Khalily, and R. Tafazolli, "Mathematical model and real-world demonstration of multi-beam and wide-beam reconfigurable intelligent surface," *IEEE Access*, vol. 11, pp. 19 613–19 621, 2023.

[33] C. Liaskos, A. Tsioliaridou, K. Georgopoulos, I. Morianos, S. Ioannidis, I. Salem, D. Manessis, S. Schmid, D. Tyrovolas, S. A. Tegos, P.-V. Mekikis, P. D. Diamantoulakis, A. Pitilakis, N. V. Kantartzis, G. K. Karagiannidis, A. C. Tasolamprou, O. Tsilipakos, M. Kafesaki, I. F. Akyildiz, A. Pitsillides, M. Pateraki, M. Vakalellis, and I. Spais, "XR-RF imaging enabled by software-defined metasurfaces and machine learning: Foundational vision, technologies and challenges," *IEEE Access*, vol. 10, pp. 119 841–119 862, 2022.

- [34] E. Ayanoglu, F. Capolino, and A. L. Swindlehurst, "Wave-controlled metasurface-based reconfigurable intelligent surfaces," *IEEE Wireless Communications*, vol. 29, no. 4, pp. 86–92, August 2022.
- [35] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116 753–116 773, August 2019.
- [36] S. Zhang, Q. Wu, S. Xu, and G. Y. Li, "Fundamental green tradeoffs: Progresses, challenges, and impacts on 5G networks," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 33–56, Firstquarter 2017.
- [37] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, February 2014.
- [38] Q. Wu, G. Y. Li, W. Chen, D. W. K. Ng, and R. Schober, "An overview of sustainable green 5G networks," *IEEE Wireless Communications*, vol. 24, no. 4, pp. 72–80, August 2017.
- [39] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge University Press, 2005.
- [40] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge University Press, 2005.
- [41] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design," in *2018 IEEE Global Communications Conference (GLOBECOM)*, 2018, pp. 1–6.
- [42] E. Björnson and L. Sanguinetti, "Demystifying the power scaling law of intelligent reflecting surfaces and metasurfaces," in *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2019, pp. 549–553.
- [43] E. Che, H. D. Tuan, and H. H. Nguyen, "Joint optimization of cooperative beamforming and relay assignment in multi-user wireless relay networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 10, pp. 5481–5495, October 2014.
- [44] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1436–1449, April 2013.
- [45] X. Guan, Q. Wu, and R. Zhang, "Intelligent reflecting surface assisted secrecy communication: Is artificial noise helpful or not?" *IEEE Wireless Communications Letters*, vol. 9, no. 6, pp. 778–782, June 2020.
- [46] J. Chen, Y.-C. Liang, Y. Pei, and H. Guo, "Intelligent reflecting surface: A programmable wireless environment for physical layer security," *IEEE Access*, vol. 7, pp. 82 599–82 612, 2019.

[47] D. Xu, X. Yu, Y. Sun, D. W. K. Ng, and R. Schober, "Resource allocation for secure IRS-assisted multiuser MISO systems," in *2019 IEEE Globecom Workshops (GC Wkshps)*, 2019, pp. 1–6.

[48] Q. Wu and R. Zhang, "Weighted sum power maximization for intelligent reflecting surface aided SWIPT," *IEEE Wireless Communications Letters*, vol. 9, no. 5, pp. 586–590, June 2020.

[49] Q. Wu and R. Zhang, "Joint active and passive beamforming optimization for intelligent reflecting surface assisted SWIPT under QoS constraints," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1735–1748, August 2020.

[50] C. Pan, H. Ren, K. Wang, M. Elkashlan, A. Nallanathan, J. Wang, and L. Hanzo, "Intelligent reflecting surface aided MIMO broadcasting for simultaneous wireless information and power transfer," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1719–1734, August 2020.

[51] D. Mishra and H. Johansson, "Channel estimation and low-complexity beamforming design for passive intelligent surface assisted MISO wireless energy transfer," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 4659–4663.

[52] M. Fu, Y. Zhou, and Y. Shi, "Intelligent reflecting surface for downlink non-orthogonal multiple access networks," in *2019 IEEE Globecom Workshops (GC Wkshps)*, 2019, pp. 1–6.

[53] G. Yang, X. Xu, and Y.-C. Liang, "Intelligent reflecting surface assisted non-orthogonal multiple access," in *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, 2020, pp. 1–6.

[54] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1838–1851, March 2020.

[55] *Study on Channel Model for Frequencies from 0.5 to 100 GHz (Release 16)*. 3GPP TR 38.901 V16.1.0, Technical Report, November 2020.

[56] *NR; User Equipment (UE) radio transmission and reception; Part 1: Range 1 Standalone (Release 18)*. 3GPP TR 38.101-1:NR, V18.4.0, Technical Report, December 2023.

[57] *NR; User Equipment (UE) radio transmission and reception; Part 2: Range 2 Standalone (Release 18)*. 3GPP TR 38.101-2:NR, V18.4.0, Technical Report, December 2023.

[58] *7 - 24 GHz Frequency Range (Release 16)*. 3GPP TR 38.820 V16.1.0, Technical Report, March 2021.

[59] C. Liaskos, S. Nie, A. Tsioliardou, A. Pitsillides, S. Ioannidis, and I. Akyildiz, "A new wireless communication paradigm through software-controlled metasurfaces," *IEEE Communications Magazine*, vol. 56, no. 9, pp. 162–169, September 2018.

[60] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Transactions on Wireless Communications*, vol. 18, no. 8, pp. 4157–4170, August 2019.

[61] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Communications Magazine*, vol. 58, no. 1, pp. 106–112, January 2020.

[62] M. Di Renzo, M. Debbah, D.-T. Phan-Huy, A. Zappone, M.-S. Alouini, C. Yuen, V. Sciancalepore, G. C. Alexandropoulos, J. Hoydis, H. Gacanin, J. de Rosny, A. Bounceur, G. Lerosey, and M. Fink, "Smart radio environments empowered by reconfigurable AI meta-surfaces: an idea whose time has come," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 8, pp. 4157–4170, May 2019.

[63] X. Pei, H. Yin, L. Tan, L. Cao, Z. Li, K. Wang, K. Zhang, and E. Björnson, "RIS-aided wireless communications: Prototyping, adaptive beamforming, and indoor/outdoor field trials," *IEEE Transactions on Communications*, vol. 69, no. 12, pp. 8627–8640, December 2021.

[64] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5394–5409, November 2019.

[65] Ö. Özdogan, E. Björnson, and E. G. Larsson, "Intelligent reflecting surfaces: Physics, propagation, and pathloss modeling," *IEEE Wireless Communications Letters*, vol. 9, no. 5, pp. 581–585, May 2020.

[66] Z. Zhang and L. Dai, "A joint precoding framework for wideband reconfigurable intelligent surface-aided cell-free network," *IEEE Transactions on Signal Processing*, vol. 69, pp. 4085–4101, June 2021.

[67] S. Hu, F. Rusek, and O. Edfors, "Beyond massive mimo: The potential of data transmission with large intelligent surfaces," *IEEE Transactions on Signal Processing*, vol. 66, no. 10, pp. 2746–2758, May 2018.

[68] M. Cheng and X. Meng, "Decode-and-forward vs. lossy-forward: Intelligent reflecting surface-assisted sidelink transmission," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 9, pp. 12 425–12 429, September 2023.

[69] J. D. Griffin and G. D. Durgin, "Complete link budgets for backscatter-radio and RFID systems," *IEEE Antennas and Propagation Magazine*, vol. 51, no. 2, pp. 11–25, April 2009.

[70] S. Idrees, X. Jia, S. Durrani, and X. Zhou, "Design of intelligent reflecting surface (IRS)-boosted ambient backscatter systems," *IEEE Access*, vol. 10, pp. 65 000–65 010, 2022.

[71] X.-T. Dang, H. V. Nguyen, and O.-S. Shin, "Optimization of IRS-NOMA-assisted cell-free massive MIMO systems using deep reinforcement learning," *IEEE Access*, vol. 11, pp. 94 402–94 414, 2023.

[72] O. Yurduseven, D. L. Marks, T. Fromenteze, and D. R. Smith, "Dynamically reconfigurable holographic metasurface aperture for a mills-cross monochromatic microwave camera," *Optics Express*, vol. 26, no. 5, pp. 5281–5291, March 2018. [Online]. Available: <https://opg.optica.org/oe/abstract.cfm?URI=oe-26-5-5281>

[73] J. An, C. Yuen, C. Huang, M. Debbah, H. Vincent Poor, and L. Hanzo, "A tutorial on holographic MIMO communications — Part I: Channel modeling and channel estimation," *IEEE Communications Letters*, vol. 27, no. 7, pp. 1664–1668, July 2023.

[74] J. An, C. Yuen, C. Huang, M. Debbah, H. V. Poor, and L. Hanzo, "A tutorial on holographic MIMO communications — Part II: Performance analysis and holographic beamforming," *IEEE Communications Letters*, vol. 27, no. 7, pp. 1669–1673, July 2023.

[75] C. Huang, S. Hu, G. C. Alexandropoulos, A. Zappone, C. Yuen, R. Zhang, M. D. Renzo, and M. Debbah, "Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends," *IEEE Wireless Communications*, vol. 27, no. 5, pp. 118–125, October 2020.

[76] K. Ntontin, A.-A. A. Boulogeorgos, Z. Abdullah, A. Mesodiakaki, S. Abadal, and S. Chatzinotas, "Time vs. unit cell splitting for autonomous reconfigurable intelligent surfaces," in *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, 2022, pp. 3332–3337.

[77] K. Ntontin, E. Björnson, A.-A. A. Boulogeorgos, Z. Abdullah, A. Mesodiakaki, S. Abadal, and S. Chatzinotas, "Time- and unit-cell splitting comparison for the autonomous operation of reconfigurable intelligent surfaces," *IEEE Transactions on Green Communications and Networking*, vol. 7, no. 3, pp. 1566–1582, September 2023.

[78] K. Ntontin, A.-A. A. Boulogeorgos, S. Abadal, A. Mesodiakaki, S. Chatzinotas, and B. Ottersten, "Perpetual reconfigurable intelligent surfaces through in-band energy harvesting: Architectures, protocols, and challenges," *IEEE Vehicular Technology Magazine*, vol. 19, no. 1, pp. 36–44, March 2024.

[79] V. Jamali, A. M. Tulino, G. Fischer, R. Muller, and R. Schober, "Scalable and energy-efficient millimeter massive MIMO architectures: Reflect-array and transmit-array antennas," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, 2019, pp. 1–7.

[80] X. Yu, D. Xu, Y. Sun, D. W. K. Ng, and R. Schober, "Robust and secure wireless communications via intelligent reflecting surfaces," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2637–2652, November 2020.

[81] C. Pan, H. Ren, K. Wang, W. Xu, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Multicell MIMO communications relying on intelligent reflecting surfaces," *IEEE Transactions on Wireless Communications*, vol. 19, no. 8, pp. 5218–5233, August 2020.

[82] V. Jamali, A. M. Tulino, G. Fischer, R. R. Müller, and R. Schober, "Intelligent surface-aided transmitter architectures for millimeter-wave ultra massive MIMO systems," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 144–167, 2021.

- [83] H. Guo, Y.-C. Liang, J. Chen, and E. G. Larsson, "Weighted sum-rate maximization for intelligent reflecting surface enhanced wireless networks," in *2019 IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6.
- [84] P. Wang, J. Fang, X. Yuan, Z. Chen, and H. Li, "Intelligent reflecting surface-assisted millimeter wave communications: Joint active and passive precoding design," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 14 960–14 973, December 2020.
- [85] Y. Li, F. Wang, X. Zhang, and S. Guo, "IRS-based MEC for delay-constrained qos Over RF-powered 6G mobile wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 7, pp. 8722–8737, July 2023.
- [86] D. Zhao, H. Lu, Y. Wang, H. Sun, and Y. Gui, "Joint power allocation and user association optimization for IRS-assisted mmWave systems," *IEEE Transactions on Wireless Communications*, vol. 21, no. 1, pp. 577–590, January 2022.
- [87] Y. Pan, K. Wang, C. Pan, H. Zhu, and J. Wang, "Self-sustainable reconfigurable intelligent surface aided simultaneous terahertz information and power transfer (STIPT)," *IEEE Transactions on Wireless Communications*, vol. 21, no. 7, pp. 5420–5434, July 2022.
- [88] J. Na, J. Kang, and J. Kang, "Intelligent reflecting surface-assisted uplink NOMA for eMBB and URLLC coexistence," *IEEE Transactions on Vehicular Technology*, pp. 1–6, 2023.
- [89] Y. Zhao, Q. Wu, G. Chen, W. Chen, R. Liu, M.-M. Zhao, Y. Wu, and S. Ma, "Intelligent reflecting surface aided multi-tier hybrid computing," *IEEE Journal of Selected Topics in Signal Processing*, pp. 1–15, 2023.
- [90] H. Xie, J. Xu, Y.-F. Liu, L. Liu, and D. W. K. Ng, "User grouping and reflective beamforming for IRS-aided URLLC," *IEEE Wireless Communications Letters*, vol. 10, no. 11, pp. 2533–2537, November 2021.
- [91] X. Tan, Z. Sun, D. Koutsonikolas, and J. M. Jornet, "Enabling indoor mobile millimeter-wave networks based on smart reflect-arrays," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, 2018, pp. 270–278.
- [92] S. Tewes, M. Heinrichs, R. Kronberger, and A. Sezgin, "IRS-enabled breath tracking with colocated commodity WiFi transceivers," *IEEE Internet of Things Journal*, vol. 10, no. 8, pp. 6870–6886, April 2023.
- [93] Y. Liu, J. Kelly, M. Holm, S. Gopal, S. R. Aghdam, and Y. Liu, "Unit cell design for intelligent reflecting and refracting surface (IRS) with independent electronic control capability," *IEEE Antennas and Wireless Propagation Letters*, vol. 23, no. 1, pp. 414–418, January 2024.
- [94] G. Zhou, C. Pan, H. Ren, K. Wang, M. Elkashlan, and M. D. Renzo, "Stochastic learning-based robust beamforming design for RIS-aided millimeter-wave systems in the presence of random blockages," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 1, pp. 1057–1061, January 2021.

[95] F. Yu, C. Zhang, and T. Q. S. Quek, "Blockage correlation in IRS-assisted millimeter wave communication systems," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2023.

[96] S. Cai, H. Qu, J. Zhang, X. Shi, and H. Zhu, "Symbol-level precoding design in IRS-aided secure wireless communication systems," *IEEE Wireless Communications Letters*, vol. 11, no. 11, pp. 2315–2319, November 2022.

[97] W. Wang, X. Liu, J. Tang, N. Zhao, Y. Chen, Z. Ding, and X. Wang, "Beamforming and jamming optimization for IRS-aided secure NOMA networks," *IEEE Transactions on Wireless Communications*, vol. 21, no. 3, pp. 1557–1569, March 2022.

[98] H.-M. Wang, J. Bai, and L. Dong, "Intelligent reflecting surfaces assisted secure transmission without eavesdropper's CSI," *IEEE Signal Processing Letters*, vol. 27, pp. 1300–1304, July 2020.

[99] T.-A. Nguyen, H.-V. Nguyen, D.-T. Do, and A. Silva, "Performance analysis of downlink double-IRS systems relying on non-orthogonal multiple access," *IEEE Access*, vol. 11, pp. 110 208–110 220, 2023.

[100] L. Ferdouse, I. Woungang, A. Anpalagan, and K. Yamamoto, "A resource allocation policy for downlink communication in distributed IRS aided multiple-input single-output systems," *IEEE Transactions on Communications*, vol. 71, no. 4, pp. 2410–2424, April 2023.

[101] N. Li, W. Hao, F. Zhou, Z. Chu, S. Yang, O. Muta, and H. Gacanin, "Min–Max latency optimization for IRS-aided cell-free mobile edge computing systems," *IEEE Internet of Things Journal*, vol. 11, no. 5, pp. 8757–8770, March 2024.

[102] Z. Hou, Y. Huang, J. Chen, G. Li, X. Guan, Y. Xu, R. Chen, and Y. Xu, "Joint IRS selection and passive beamforming in multiple IRS-UAV-enhanced anti-jamming D2D communication networks," *IEEE Internet of Things Journal*, vol. 10, no. 22, pp. 19 558–19 569, November 2023.

[103] J. Cheng, C. Shen, Z. Chen, and N. Pappas, "Robust beamforming design for IRS-aided URLLC in D2D networks," *IEEE Transactions on Communications*, vol. 70, no. 9, pp. 6035–6049, September 2022.

[104] Q. Chen, Y. Zhou, W. Xu, and Y. Wu, "Energy-efficient resource allocation for irs-assisted swipt-enabled D2D communications with imperfect CSI," *IEEE Communications Letters*, pp. 1–1, 2024.

[105] N. Ashraf, S. A. Sheikh, S. A. Khan, I. Sharya, and M. Jalal, "Simultaneous wireless information and power transfer with cooperative relaying for next-generation wireless networks: A review," *IEEE Access*, vol. 9, pp. 71 482–71 504, 2021.

[106] C. Tatino, N. Pappas, and D. Yuan, "QoS aware robot trajectory optimization with IRS-assisted millimeter-wave communications," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 1323–1336, 2022.

[107] T. Ohyama, Y. Kawamoto, and N. Kato, "Standalone intelligent reflecting surface with automatic update of phase-shift switching schedule for improved communication and localization," *IEEE Transactions on Vehicular Technology*, pp. 1–11, 2024.

[108] M. Noor-A-Rahim, F. Firyaguna, J. John, M. O. Khyam, D. Pesch, E. Armstrong, H. Claussen, and H. V. Poor, "Toward industry 5.0: Intelligent reflecting surface in smart manufacturing," *IEEE Communications Magazine*, vol. 60, no. 10, pp. 72–78, October 2022.

[109] R. Liu, Q. Wu, M. Di Renzo, and Y. Yuan, "A path to smart radio environments: An industrial viewpoint on reconfigurable intelligent surfaces," *IEEE Wireless Communications*, vol. 29, no. 1, pp. 202–208, February 2022.

[110] X. Mu, Y. Liu, L. Guo, J. Lin, and R. Schober, "Intelligent reflecting surface enhanced indoor robot path planning: A radio map-based approach," *IEEE Transactions on Wireless Communications*, vol. 20, no. 7, pp. 4732–4747, July 2021.

[111] H. Ren, K. Wang, and C. Pan, "Intelligent reflecting surface-aided URLLC in a factory automation scenario," *IEEE Transactions on Communications*, vol. 70, no. 1, pp. 707–723, January 2022.

[112] M. B. Goktas, Y. Dursun, and Z. Ding, "IRS and SWIPT-assisted full-duplex NOMA for 6G umMTC," *IEEE Transactions on Green Communications and Networking*, vol. 7, no. 4, pp. 1957–1970, December 2023.

[113] Q.-U.-A. Nadeem, H. Alwazani, A. Kammoun, A. Chaaban, M. Debbah, and M.-S. Alouini, "Intelligent reflecting surface-assisted multi-user MISO communication: Channel estimation and beamforming design," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 661–680, May 2020.

[114] M. Cui, G. Zhang, and R. Zhang, "Secure wireless communication via intelligent reflecting surface," *IEEE Wireless Communications Letters*, vol. 8, no. 5, pp. 1410–1414, October 2019.

[115] S. Gong, X. Lu, D. T. Hoang, D. Niyato, L. Shu, D. I. Kim, and Y.-C. Liang, "Toward smart wireless communications via intelligent reflecting surfaces: A contemporary survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 4, pp. 2283–2314, Fourthquarter (Q4) 2020.

[116] Z. Ding and H. Vincent Poor, "A simple design of IRS-NOMA transmission," *IEEE Communications Letters*, vol. 24, no. 5, pp. 1119–1123, May 2020.

[117] B. Zheng, Q. Wu, and R. Zhang, "Intelligent reflecting surface-assisted multiple access with user pairing: NOMA or OMA?" *IEEE Communications Letters*, vol. 24, no. 4, pp. 753–757, April 2020.

[118] K. Ntontin, A. A. Boulogeorgos, E. Björnson, W. A. Martins, S. Kisseleff, S. Abadal, E. Alarcón, A. Papazafeiropoulos, F. I. Lazarakis, and S. Chatzinotas, "Wireless energy harvesting for autonomous reconfigurable intelligent surfaces," *IEEE Transactions on Green Communications and Networking*, vol. 7, no. 1, pp. 114–129, March 2023.

[119] N. Ashraf, T. Saeed, H. Taghvaei, S. Abadal, V. Vassiliou, C. Liaskos, A. Pitsilides, and M. Lestas, "Intelligent beam steering for wireless communication using programmable metasurfaces," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 5, pp. 4848–4861, May 2023.

[120] H. Taghvaei, A. Pitilakis, O. Tsilipakos, A. C. Tasolamprou, N. V. Kantartzis, M. Kafesaki, A. Cabellos-Aparicio, E. Alarcón, S. Abadal, and G. Gradoni, "Tunable graphene-based metasurfaces for multi-wideband 6G communications," in *2022 Sixteenth International Congress on Artificial Materials for Novel Wave Phenomena (Metamaterials)*, 2022, pp. 434–436.

[121] T. Saeed, S. Abadal, C. Liaskos, A. Pitsillides, H. Taghvaei, A. Cabellos-Aparicio, V. Soteriou, E. Alarcón, I. F. Akyildiz, and M. Lestas, "Workload characterization and traffic analysis for reconfigurable intelligent surfaces within 6G wireless systems," *IEEE Transactions on Mobile Computing*, vol. 22, no. 5, pp. 3079–3094, May 2023.

[122] Y. Yang, S. Zhang, and R. Zhang, "IRS-enhanced OFDMA: Joint resource allocation and passive beamforming optimization," *IEEE Wireless Communications Letters*, vol. 9, no. 6, pp. 760–764, June 2020.

[123] Z. Wei, Y. Cai, Z. Sun, D. W. K. Ng, J. Yuan, M. Zhou, and L. Sun, "Sum-rate maximization for IRS-assisted UAV OFDMA communication systems," *IEEE Transactions on Wireless Communications*, vol. 20, no. 4, pp. 2530–2550, April 2021.

[124] D. Xu, X. Yu, Y. Sun, D. W. K. Ng, and R. Schober, "Resource allocation for IRS-assisted full-duplex cognitive radio systems," *IEEE Transactions on Communications*, vol. 68, no. 12, pp. 7376–7394, December 2020.

[125] Y. Cai, Z. Wei, S. Hu, C. Liu, D. W. K. Ng, and J. Yuan, "Resource allocation and 3D trajectory design for power-efficient IRS-assisted UAV-NOMA communications," *IEEE Transactions on Wireless Communications*, vol. 21, no. 12, pp. 10315–10334, June 2022.

[126] M. Hua, Q. Wu, D. W. K. Ng, J. Zhao, and L. Yang, "Intelligent reflecting surface-aided joint processing coordinated multipoint transmission," *IEEE Transactions on Communications*, vol. 69, no. 3, pp. 1650–1665, March 2021.

[127] B. Clerckx, R. Zhang, R. Schober, D. W. K. Ng, D. I. Kim, and H. V. Poor, "Fundamentals of wireless information and power transfer: From RF energy harvester models to signal and system designs," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 1, pp. 4–33, January 2019.

[128] D. W. K. Ng, E. S. Lo, and R. Schober, "Energy-efficient resource allocation in multi-cell OFDMA systems with limited backhaul capacity," *IEEE Transactions on Wireless Communications*, vol. 11, no. 10, pp. 3618–3631, October 2012.

[129] J. Jalali and A. Khalili, "Optimal resource allocation for MC-NOMA in SWIPT-enabled networks," *IEEE Communications Letters*, vol. 24, no. 10, pp. 2250–2254, October 2020.

[130] S. Jang, H. Lee, S. Kang, T. Oh, and I. Lee, "Energy efficient beamforming for multi-cell MISO SWIPT systems," in *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*, 2018, pp. 1–6.

[131] ———, "Energy efficient SWIPT systems in multi-cell MISO networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 12, pp. 8180–8194, December 2018.

[132] J. Tang, J. Luo, M. Liu, D. K. C. So, E. Alsusa, G. Chen, K. K. Wong, and J. A. Chambers, "Energy efficiency optimization for NOMA with SWIPT," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 3, pp. 452–466, June 2019.

[133] H. Zhang, M. Feng, K. Long, G. K. Karagiannidis, V. C. M. Leung, and H. V. Poor, "Energy efficient resource management in SWIPT enabled heterogeneous networks with NOMA," *IEEE Transactions on Wireless Communications*, vol. 19, no. 2, pp. 835–845, February 2020.

[134] W. Lu, X. Xu, G. Huang, B. Li, Y. Wu, N. Zhao, and F. R. Yu, "Energy efficiency optimization in SWIPT enabled WSNs for smart agriculture," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 6, pp. 4335–4344, February 2021.

[135] Y. Lu, K. Xiong, P. Fan, Z. Ding, Z. Zhong, and K. B. Letaief, "Global energy efficiency in secure MISO SWIPT systems with non-linear power-splitting EH model," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 1, pp. 216–232, January 2019.

[136] S. Zargari, A. Khalili, and R. Zhang, "Energy efficiency maximization via joint active and passive beamforming design for multiuser MISO IRS-aided SWIPT," *IEEE Wireless Communications Letters*, vol. 10, no. 3, pp. 557–561, March 2021.

[137] M. Sheng, L. Wang, X. Wang, Y. Zhang, C. Xu, and J. Li, "Energy efficient beamforming in miso heterogeneous cellular networks with wireless information and power transfer," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 954–968, April 2016.

[138] J. Jalali, "Resource allocation for SWIPT in multi-service wireless networks," Master's thesis, Ghent University, June 2020.

[139] Z. Ding, C. Zhong, D. Wing Kwan Ng, M. Peng, H. A. Suraweera, R. Schober, and H. V. Poor, "Application of smart antenna technologies in simultaneous wireless information and power transfer," *IEEE Communications Magazine*, vol. 53, no. 4, pp. 86–93, April 2015.

[140] J. Tang, D. K. C. So, A. Shojaeifard, K. Wong, and J. Wen, "Joint antenna selection and spatial switching for energy efficient MIMO SWIPT system," *IEEE Transactions on Wireless Communications*, vol. 16, no. 7, pp. 4754–4769, July 2017.

[141] Y. Al-Eryani, M. Akrout, and E. Hossain, "Antenna clustering for simultaneous wireless information and power transfer in a MIMO full-duplex system: A deep reinforcement learning-based design," *IEEE Transactions on Communications*, vol. 69, no. 4, pp. 2331–2345, April 2021.

[142] F. Benkhelifa and M.-S. Alouini, "Prioritizing data/energy thresholding-based antenna switching for SWIPT-enabled secondary receiver in cognitive radio networks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 782–800, December 2017.

[143] Z. Wei, X. Yu, D. W. K. Ng, and R. Schober, "Resource allocation for simultaneous wireless information and power transfer systems: A tutorial overview," *Proc. IEEE*, vol. 110, no. 1, pp. 127–149, October 2022.

[144] I. Krikidis, S. Timotheou, S. Nikolaou, G. Zheng, D. W. K. Ng, and R. Schober, "Simultaneous wireless information and power transfer in modern communication systems," *IEEE Communications Magazine*, vol. 52, no. 11, pp. 104–110, November 2014.

[145] L. R. Varshney, "Transporting information and energy simultaneously," in *2008 IEEE International Symposium on Information Theory*, 2008, pp. 1612–1616.

[146] R. Zhang and C. K. Ho, "MIMO broadcasting for simultaneous wireless information and power transfer," *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 1989–2001, May 2013.

[147] I. Krikidis, S. Sasaki, S. Timotheou, and Z. Ding, "A low complexity antenna switching for joint wireless information and energy transfer in MIMO relay channels," *IEEE Transactions on Communications*, vol. 62, no. 5, pp. 1577–1587, May 2014.

[148] D. W. K. Ng, E. S. Lo, and R. Schober, "Robust beamforming for secure communication in systems with wireless information and power transfer," *IEEE Transactions on Wireless Communications*, vol. 13, no. 8, pp. 4599–4615, August 2014.

[149] L. You, J. Xiong, D. W. K. Ng, C. Yuen, W. Wang, and X. Gao, "Energy efficiency and spectral efficiency tradeoff in RIS-aided multiuser MIMO uplink transmission," *IEEE Transactions on Signal Processing*, vol. 69, pp. 1407–1421, December 2021.

[150] X. Zhou, R. Zhang, and C. K. Ho, "Wireless information and power transfer: Architecture design and rate-energy tradeoff," *IEEE Transactions on Communications*, vol. 61, no. 11, pp. 4754–4767, October 2013.

[151] J. Jalali, *Resource allocation for SWIPT in multi-service wireless networks*. M.S. thesis, Department of Telecommunications and Information Processing, TELIN/IMEC, University of Ghent, Ghent, Belgium, June 2020. [Online]. Available: <https://arxiv.org/abs/2007.13676>

[152] E. Boshkovska, D. W. K. Ng, N. Zlatanov, and R. Schober, "Practical non-linear energy harvesting model and resource allocation for SWIPT systems," *IEEE Communications Letters*, vol. 19, no. 12, pp. 2082–2085, October 2015.

[153] D. W. K. Ng, E. S. Lo, and R. Schober, "Wireless information and power transfer: Energy efficiency optimization in OFDMA systems," *IEEE Transactions on Wireless Communications*, vol. 12, no. 12, pp. 6352–6370, December 2013.

[154] Y. Jong, *An efficient global optimization algorithm for nonlinear sum-of-ratios problem*. Center of Natural Science, University of Sciences, Pyongyang, DPR Korea, May 2012.

[155] M. Li, N. Cheng, J. Gao, Y. Wang, L. Zhao, and X. Shen, "Energy-efficient UAV-assisted mobile edge computing: Resource allocation and trajectory optimization," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 3, pp. 3424–3438, March 2020.

[156] M. Ali, S. Qaisar, M. Naeem, and S. Mumtaz, "Energy efficient resource allocation in D2D-assisted heterogeneous networks with relays," *IEEE Access*, vol. 4, pp. 4902–4911, 2016.

- [157] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Transactions on Communications*, vol. 69, no. 5, pp. 3313–3351, May 2021.
- [158] Z. Kang, C. You, and R. Zhang, "IRS-aided wireless relaying: Deployment strategy and capacity scaling," *IEEE Wireless Communications Letters*, vol. 11, no. 2, pp. 215–219, February 2022.
- [159] S. Xu, Y. Du, J. Liu, and J. Li, "Weighted sum rate maximization in IRS-BackCom enabled downlink multi-cell MISO network," *IEEE Communications Letters*, vol. 26, no. 3, pp. 642–646, March 2022.
- [160] X. Yu, D. Xu, and R. Schober, "Optimal beamforming for MISO communications via intelligent reflecting surfaces," in *2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2020, pp. 1–5.
- [161] M. Bennis, M. Debbah, and H. V. Poor, "Ultrareliable and low-latency wireless communication: Tail, risk, and scale," *Proceedings of the IEEE*, vol. 106, no. 10, pp. 1834–1853, October 2018.
- [162] C. She, C. Yang, and T. Q. S. Quek, "Radio resource management for ultra-reliable and low-latency communications," *IEEE Communications Magazine*, vol. 55, no. 6, pp. 72–78, June 2017.
- [163] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
- [164] C. Sun, C. She, C. Yang, T. Q. S. Quek, Y. Li, and B. Vucetic, "Optimizing resource allocation in the short blocklength regime for ultra-reliable and low-latency communications," *IEEE Transactions on Wireless Communications*, vol. 18, no. 1, pp. 402–415, January 2019.
- [165] W. R. Ghanem, V. Jamali, Y. Sun, and R. Schober, "Resource allocation for multi-user downlink MISO OFDMA-URLLC systems," *IEEE Transactions on Communications*, vol. 68, no. 11, pp. 7184–7200, November 2020.
- [166] W. R. Ghanem, V. Jamali, and R. Schober, "Joint beamforming and phase shift optimization for multicell IRS-aided OFDMA-URLLC systems," in *2021 IEEE Wireless Communications and Networking Conference (WCNC)*, 2021, pp. 1–7.
- [167] T. Bai, C. Pan, Y. Deng, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Latency minimization for intelligent reflecting surface aided mobile edge computing," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2666–2682, November 2020.
- [168] R. Hashemi, S. Ali, N. H. Mahmood, and M. Latva-aho, "Average rate and error probability analysis in short packet communications over RIS-aided URLLC systems," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 10, pp. 10 320–10 334, October 2021.

[169] Z. Huang, B. Zheng, and R. Zhang, "Transforming fading channel from fast to slow: Intelligent refracting surface aided high-mobility communication," *IEEE Transactions on Wireless Communications*, vol. 21, no. 7, pp. 4989–5003, July 2022.

[170] K. Zhi, C. Pan, H. Ren, and K. Wang, "Statistical CSI-based design for reconfigurable intelligent surface-aided massive MIMO systems with direct links," *IEEE Wireless Communications Letters*, vol. 10, no. 5, pp. 1128–1132, May 2021.

[171] M. Shen, X. Lei, P. T. Mathiopoulos, and R. Q. Hu, "Robust beamforming design for IRS-aided secure communication systems under complete imperfect CSI," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 6, pp. 8204–8209, June 2023.

[172] M. Setayesh, S. Bahrami, and V. W. Wong, "Joint PRB and power allocation for slicing eMBB and URLLC services in 5G C-RAN," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, 2020, pp. 1–6.

[173] S. Zargari, A. Khalili, Q. Wu, M. Robat Mili, and D. W. K. Ng, "Max-Min fair energy-efficient beamforming design for intelligent reflecting surface-aided SWIPT systems with non-linear energy harvesting model," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 6, pp. 5848–5864, June 2021.

[174] A. Khalili, S. Zargari, Q. Wu, D. W. K. Ng, and R. Zhang, "Multi-objective resource allocation for IRS-aided SWIPT," *IEEE Wireless Communications Letters*, vol. 10, no. 6, pp. 1324–1328, June 2021.

[175] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," <http://cvxr.com/cvx>, March 2014.

[176] M. Di Renzo, K. Ntontin, J. Song, F. H. Danufane, X. Qian, F. Lazarakis, J. De Rosny, D.-T. Phan-Huy, O. Simeone, R. Zhang, M. Debbah, G. Leroosey, M. Fink, S. Tretyakov, and S. Shamai, "Reconfigurable intelligent surfaces vs. re-laying: Differences, similarities, and performance comparison," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 798–807, 2020.

[177] C. Bockelmann, N. Pratas, H. Nikopour, K. Au, T. Svensson, C. Stefanovic, P. Popovski, and A. Dekorsy, "Massive machine-type communications in 5G: Physical and MAC-layer solutions," *IEEE Communications Magazine*, vol. 54, no. 9, pp. 59–65, September 2016.

[178] K. Singh, M.-L. Ku, and M. F. Flanagan, "Energy-efficient precoder design for down-link multi-user MISO networks with finite blocklength codes," *IEEE Transactions on Green Communications and Networking*, vol. 5, no. 1, pp. 160–173, March 2021.

[179] L. Zhao, S. Yang, X. Chi, W. Chen, and S. Ma, "Achieving energy-efficient uplink URLLC with MIMO-aided grant-free access," *IEEE Transactions on Wireless Communications*, vol. 21, no. 2, pp. 1407–1420, February 2022.

[180] M. Darabi, V. Jamali, L. Lampe, and R. Schober, "Hybrid puncturing and superposition scheme for joint scheduling of URLLC and eMBB traffic," *IEEE Communications Letters*, vol. 26, no. 5, pp. 1081–1085, May 2022.

[181] Q.-U.-A. Nadeem, A. Kammoun, A. Chaaban, M. Debbah, and M.-S. Alouini, "Asymptotic Max-Min SINR analysis of reconfigurable intelligent surface assisted MISO systems," *IEEE Transactions on Wireless Communications*, vol. 19, no. 12, pp. 7748–7764, December 2020.

[182] N. K. Kundu and M. R. Mckay, "Large intelligent surfaces with channel estimation overhead: Achievable rate and optimal configuration," *IEEE Wireless Communications Letters*, vol. 10, no. 5, pp. 986–990, May 2021.

[183] Z. Wang, L. Liu, and S. Cui, "Channel estimation for intelligent reflecting surface assisted multiuser communications: Framework, algorithms, and analysis," *IEEE Transactions on Wireless Communications*, vol. 19, no. 10, pp. 6607–6620, October 2020.

[184] T. L. Jensen and E. De Carvalho, "An optimal channel estimation scheme for intelligent reflecting surfaces based on a minimum variance unbiased estimator," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 5000–5004.

[185] Q. Wu and R. Zhang, "Beamforming optimization for intelligent reflecting surface with discrete phase shifts," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 7830–7833.

[186] B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface assisted multi-user OFDMA: Channel estimation and training design," *IEEE Transactions on Wireless Communications*, vol. 19, no. 12, pp. 8315–8329, December 2020.

[187] B. Zheng and R. Zhang, "Intelligent reflecting surface-enhanced OFDM: Channel estimation and reflection optimization," *IEEE Wireless Communications Letters*, vol. 9, no. 4, pp. 518–522, April 2020.

[188] C. You, B. Zheng, and R. Zhang, "Channel estimation and passive beamforming for intelligent reflecting surface: Discrete phase shift and progressive refinement," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2604–2620, November 2020.

[189] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, April 2003.

[190] S. Zhang and R. Zhang, "Capacity characterization for intelligent reflecting surface aided MIMO communication," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1823–1838, June 2020.

[191] P. Xu, G. Chen, Z. Yang, and M. D. Renzo, "Reconfigurable intelligent surfaces-assisted communications with discrete phase shifts: How many quantization levels are required to achieve full diversity?" *IEEE Wireless Communications Letters*, vol. 10, no. 2, pp. 358–362, February 2021.

[192] S. Zhou, W. Xu, K. Wang, M. Di Renzo, and M.-S. Alouini, "Spectral and energy efficiency of IRS-assisted MISO communication with hardware impairments," *IEEE Wireless Communications Letters*, vol. 9, no. 9, pp. 1366–1369, September 2020.

[193] S. Gong, Z. Yang, C. Xing, J. An, and L. Hanzo, "Beamforming optimization for intelligent reflecting surface-aided SWIPT IoT networks relying on discrete phase shifts," *IEEE Internet of Things Journal*, vol. 8, no. 10, pp. 8585–8602, May 2021.

[194] M.-M. Zhao, A. Liu, and R. Zhang, "Outage-constrained robust beamforming for intelligent reflecting surface aided wireless communication," *IEEE Transactions on Signal Processing*, vol. 69, pp. 1301–1316, February 2021.

[195] K. Shen and W. Yu, "Fractional programming for communication systems — Part I: Power control and beamforming," *IEEE Transactions on Signal Processing*, vol. 66, no. 10, pp. 2616–2630, May 2018.

[196] Z. Wei, D. W. K. Ng, and J. Yuan, "NOMA for hybrid mmWave communication systems with beamwidth control," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 3, pp. 567–583, June 2019.

[197] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[198] R. Jain, D.-M. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," *DEC Research, Hudson, MA, USA, Tech. Rep. TR-301*, vol. 38, September 1984. [Online]. Available: <https://doi.org/10.48550/arXiv.cs.9809099>

[199] F. Salehi, N. Neda, M.-H. Majidi, and H. Ahmadi, "Cooperative NOMA-based user pairing for URLLC: A Max–Min fairness approach," *IEEE Systems Journal*, vol. 16, no. 3, pp. 3833–3843, September 2022.

[200] H. Yin, L. Zhang, and S. Roy, "Multiplexing URLLC traffic within eMBB services in 5G NR: Fair scheduling," *IEEE Transactions on Communications*, vol. 69, no. 2, pp. 1080–1093, February 2021.

[201] H. Shariatmadari, R. Duan, Z. Li, S. Iraji, M. A. Uusitalo, and R. Jäntti, "Analysis of transmission modes for ultra-reliable communications," in *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2016, pp. 1–6.

[202] A. A. Nasir, H. D. Tuan, H. Q. Ngo, T. Q. Duong, and H. V. Poor, "Cell-free massive MIMO in the short blocklength regime for URLLC," *IEEE Transactions on Wireless Communications*, vol. 20, no. 9, pp. 5861–5871, September 2021.

[203] W. R. Ghanem, V. Jamali, and R. Schober, "Optimal resource allocation for multi-user OFDMA-URLLC MEC systems," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 2005–2023, 2022.

[204] S. Dhok, P. Raut, P. K. Sharma, K. Singh, and C.-P. Li, "Non-linear energy harvesting in RIS-Assisted URLLC networks for industry automation," *IEEE Transactions on Communications*, vol. 69, no. 11, pp. 7761–7774, October 2021.

[205] V. D. P. Souto, S. Montejo-Sánchez, J. L. Rebelatto, R. D. Souza, and B. F. Uchôa-Filho, "IRS-Aided physical layer network slicing for URLLC and eMBB," *IEEE Access*, vol. 9, pp. 163 086–163 098, December 2021.

[206] Z. Li, H. Shen, W. Xu, P. Zhu, and C. Zhao, "Resource allocation for irs-assisted uplink URLLC systems," *IEEE Communications Letters*, vol. 27, no. 6, pp. 1540–1544, June 2023.

[207] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Transactions on Signal Processing*, vol. 65, no. 3, pp. 794–816, February 2017.

[208] C. Sun and R. Dai, "An iterative rank penalty method for nonconvex quadratically constrained quadratic programs," *SIAM Journal on Control and Optimization (SICON)*, vol. 57, no. 6, pp. 3749–3766, January 2019. [Online]. Available: <https://doi.org/10.1137/17M1147214>

[209] W. Dinkelbach, "On nonlinear fractional programming," *Management science*, vol. 13, no. 7, pp. 492–498, 1967.

[210] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, March 2004. [Online]. Available: <http://www.amazon.com/exec/obidos/redirect?tag=citeulike-20&path=ASIN/0521833787>

[211] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of Things for smart cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, February 2014.

[212] Y. Zeng, R. Zhang, and T. J. Lim, "Wireless communications with unmanned aerial vehicles: Opportunities and challenges," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 36–42, May 2016.

[213] Y. Cui, F. Liu, X. Jing, and J. Mu, "Integrating sensing and communications for ubiquitous IoT: Applications, trends, and challenges," *IEEE Network*, vol. 35, no. 5, pp. 158–167, September/October 2021.

[214] L. Shen, N. Wang, Z. Zhu, W. Xu, Y. Li, X. Mu, and L. Cai, "UAV-Enabled data collection over clustered machine-type communication networks: AEM modeling and trajectory planning," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 9, pp. 10 016–10 032, September 2022.

[215] S. Barbarossa, S. Sardellitti, and P. Di Lorenzo, "Communicating while computing: Distributed mobile cloud computing over 5G heterogeneous networks," *IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 45–55, October 2014.

[216] H. Chen, R. Abbas, P. Cheng, M. Shirvanimoghaddam, W. Hardjawana, W. Bao, Y. Li, and B. Vucetic, "Ultra-reliable low latency cellular networks: Use cases, challenges and approaches," *IEEE Communications Magazine*, vol. 56, no. 12, pp. 119–125, December 2018.

[217] P. Mach and Z. Becvar, "Mobile edge computing: A survey on architecture and computation offloading," *IEEE Communications Surveys and Tutorials*, vol. 19, no. 3, pp. 1628–1656, 3rd Quart. 2017.

[218] Z. A. El Houda, B. Brik, A. Ksentini, and L. Khoukhi, "A MEC-Based architecture to secure IoT applications using federated deep learning," *IEEE Internet of Things Magazine*, vol. 6, no. 1, pp. 60–63, March 2023.

[219] J. Liang, H. Xing, F. Wang, and V. K. N. Lau, "Joint task offloading and cache placement for energy-efficient mobile edge computing systems," *IEEE Wireless Communications Letters*, vol. 12, no. 4, pp. 694–698, April 2023.

[220] Q. Gan, G. Li, W. He, Y. Zhao, Y. Song, and C. Xu, "Delay-minimization offloading scheme in multi-server MEC networks," *IEEE Wireless Communications Letters*, vol. 12, no. 6, pp. 1071–1075, June 2023.

[221] L. Li and P. Fan, "Latency and task loss probability for NOMA assisted MEC in mobility-aware vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 5, pp. 6891–6895, May 2023.

[222] H. Yu and K.-W. Chin, "Maximizing sensing and computation rate in Ad Hoc energy harvesting IoT networks," *IEEE Internet of Things Journal*, vol. 10, no. 6, pp. 5434–5446, March 2023.

[223] L. Shi, X. Chu, H. Sun, and G. Lu, "Wireless-powered OFDMA-MEC networks with hybrid active–passive communications," *IEEE Internet of Things Journal*, vol. 10, no. 12, pp. 10 484–10 496, June 2023.

[224] M. Wu, W. Qi, J. Park, P. Lin, L. Guo, and I. Lee, "Residual energy maximization for wireless powered mobile edge computing systems with mixed-offloading," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 4, pp. 4523–4528, April 2022.

[225] X. Qi, M. Peng, and H. Zhang, "Joint mmWave beamforming and resource allocation in NOMA-MEC network for Internet of Things," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 4, pp. 4969–4980, April 2023.

[226] Q. Zhang, Y. Wang, H. Li, S. Hou, and Z. Song, "Resource allocation for energy efficient STAR-RIS aided MEC systems," *IEEE Wireless Communications Letters*, vol. 12, no. 4, pp. 610–614, April 2023.

[227] G. Chen, Q. Wu, R. Liu, J. Wu, and C. Fang, "IRS aided MEC systems with binary offloading: A unified framework for dynamic IRS beamforming," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 2, pp. 349–365, February 2023.

[228] Y. Wang, M. Sheng, X. Wang, L. Wang, and J. Li, "Mobile-edge computing: Partial computation offloading using dynamic voltage scaling," *IEEE Transactions on Communications*, vol. 64, no. 10, pp. 4268–4282, October 2016.

[229] W. Zhang, Y. Wen, K. Guan, D. Kilper, H. Luo, and D. O. Wu, "Energy-optimal mobile cloud computing under stochastic wireless channel," *IEEE Transactions on Wireless Communications*, vol. 12, no. 9, pp. 4569–4581, September 2013.

[230] D. H. N. Nguyen, L. B. Le, T. Le-Ngoc, and R. W. Heath, "Hybrid MMSE precoding and combining designs for mmWave multiuser systems," *IEEE Access*, vol. 5, pp. 19 167–19 181, September 2017.

[231] M.-M. Zhao, Q. Wu, M.-J. Zhao, and R. Zhang, "IRS-aided wireless communication with imperfect CSI: Is amplitude control helpful or not?" in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, 2020, pp. 1–6.

[232] L. Li, T. J. Cui, W. Ji, S. Liu, J. Ding, X. Wan, Y. B. Li, M. Jiang, C.-W. Qiu, and S. Zhang, "Electromagnetic reprogrammable coding-metasurface holograms," *Nature Communications*, vol. 8, no. 1, p. 197, August 2017.

[233] C. Huang, G. C. Alexandropoulos, A. Zappone, M. Debbah, and C. Yuen, "Energy efficient multi-user MISO communication using low resolution large intelligent surfaces," in *2018 IEEE Globecom Workshops (GC Wkshps)*, December 2018, pp. 1–6.

[234] M.-M. Zhao, Q. Wu, M.-J. Zhao, and R. Zhang, "Exploiting amplitude control in intelligent reflecting surface aided wireless communication with imperfect CSI," *IEEE Transactions on Communications*, vol. 69, no. 6, pp. 4216–4231, March 2021.

[235] C. You and R. Zhang, "Wireless communication aided by intelligent reflecting surface: Active or passive?" *IEEE Wireless Communications Letters*, vol. 10, no. 12, pp. 2659–2663, December 2021.

[236] Z. Zhang, L. Dai, X. Chen, C. Liu, F. Yang, R. Schober, and H. V. Poor, "Active RIS vs. passive RIS: Which will prevail in 6G?" *IEEE Transactions on Communications*, vol. 71, no. 3, pp. 1707–1725, March 2023.

[237] S. Zeng, H. Zhang, B. Di, Y. Tan, Z. Han, H. V. Poor, and L. Song, "Reconfigurable intelligent surfaces in 6G: Reflective, transmissive, or both?" *IEEE Communications Letters*, vol. 25, no. 6, pp. 2063–2067, June 2021.

[238] M. H. Khoshafa, T. M. N. Ngatchad, M. H. Ahmed, and A. R. Ndjiongue, "Active reconfigurable intelligent surfaces-aided wireless communication system," *IEEE Communications Letters*, vol. 25, no. 11, pp. 3699–3703, October 2021.

[239] D. Xu, X. Yu, D. W. Kwan Ng, and R. Schober, "Resource allocation for active IRS-assisted multiuser communication systems," in *2021 55th Asilomar Conference on Signals, Systems, and Computers*, 2021, pp. 113–119.

[240] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, October 2014.

[241] Y. Zhang, C. You, and B. Zheng, "Multi-active multi-passive (MAMP)-IRS aided wireless communication: A multi-hop beam routing design," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 8, pp. 2497–2513, August 2023.

[242] Q. Qi, X. Chen, A. Khalili, C. Zhong, Z. Zhang, and D. W. K. Ng, "Integrating sensing, computing, and communication in 6G wireless networks: Design and optimization," *IEEE Transactions on Communications*, vol. 70, no. 9, pp. 6212–6227, September 2022.

[243] J. Jalali, A. Khalili, and H. Steendam, "Antenna selection and resource allocation in downlink MISO OFDMA femtocell networks," in *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, 2020, pp. 1–6.

[244] L. Muñoz-Saavedra *et al.*, "Augmented and virtual reality evolution and future tendency," *Applied sciences*, vol. 10, no. 1, p. 322, January 2020.

[245] M. Chen, W. Saad, and C. Yin, "Virtual reality over wireless networks: Quality-of-service model and learning-based resource management," *IEEE Transactions on Communications*, vol. 66, no. 11, pp. 5621–5635, November 2018.

[246] F. Lemic, S. Abadal, C. Han, J. M. Marquez-Barja, E. Alarcón, and J. Famaey, "Localization in power-constrained terahertz-operating software-defined metamaterials," *Nano Communication Networks*, vol. 30, p. 100365, December 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1878778921000260>

[247] C. Chaccour, M. N. Soorki, W. Saad, M. Bennis, and P. Popovski, "Risk-based optimization of virtual reality over terahertz reconfigurable intelligent surfaces," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.

[248] K.-L. Besser and E. A. Jorswieck, "Reconfigurable intelligent surface phase hopping for ultra-reliable communications," *IEEE Transactions on Wireless Communications*, vol. 21, no. 11, pp. 9082–9095, November 2022.

[249] Y. Zhou, C. Pan, P. L. Yeoh, K. Wang, Z. Ma, B. Vucetic, and Y. Li, "Latency minimization for secure intelligent reflecting surface enhanced virtual reality delivery systems," *IEEE Wireless Communications Letters*, vol. 11, no. 9, pp. 1770–1774, September 2022.

[250] E. R. Bachmann, E. Hodgson, C. Hoffbauer, and J. Messinger, "Multi-user redirected walking and resetting using artificial potential fields," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 5, pp. 2022–2031, May 2019.

[251] W. Tang, M. Z. Chen, X. Chen, J. Y. Dai, Y. Han, M. Di Renzo, Y. Zeng, S. Jin, Q. Cheng, and T. J. Cui, "Wireless communications with reconfigurable intelligent surface: Path loss modeling and experimental measurement," *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 421–439, January 2021.

[252] E. Björnson and L. Sanguinetti, "Power scaling laws and near-field behaviors of massive MIMO and intelligent reflecting surfaces," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 1306–1324, 2020.

[253] F. Lemic, J. Struye, T. Van Onsem, J. Famaey, and X. Costa-Pérez, "Predictive context-awareness for full-immersive multiuser virtual reality with redirected walking," *IEEE Communications Magazine*, vol. 61, no. 9, pp. 32–38, September 2023.

[254] T. Van Onsem, J. Struye, X. C. Perez, J. Famaey, and F. Lemic, "Toward full-immersive multiuser virtual reality with redirected walking," *IEEE Access*, vol. 11, pp. 24 722–24 736, 2023.

[255] K. Ntontin, D. Selimis, A.-A. A. Boulogeorgos, A. Alexandridis, A. Tsolis, V. Vlachodimitropoulos, and F. Lazarakis, "Optimal reconfigurable intelligent surface placement in millimeter-wave communications," in *2021 15th European Conference on Antennas and Propagation (EuCAP)*, 2021, pp. 1–5.

[256] V. Angelov, E. Petkov, G. Shipkovenski, and T. Kalushkov, "Modern virtual reality headsets," in *2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 2020, pp. 1–5.

[257] A. Batra, M. El-Absi, M. Wiemeler, D. Göhringer, and T. Kaiser, "Indoor THz SAR trajectory deviations effects and compensation with passive sub-mm localization system," *IEEE Access*, vol. 8, pp. 177 519–177 533, September 2020.

[258] H. Müller, V. Niculescu, T. Polonelli, M. Magno, and L. Benini, "Robust and efficient depth-based obstacle avoidance for autonomous miniaturized UAVs," *IEEE Transactions on Robotics*, vol. 39, no. 6, pp. 4935–4951, December 2023.

[259] H. Shakhatreh, A. H. Sawalmeh, A. Al-Fuqaha, Z. Dou, E. Almaita, I. Khalil, N. S. Othman, A. Khreishah, and M. Guizani, "Unmanned aerial vehicles (UAVs): A survey on civil applications and key research challenges," *IEEE Access*, vol. 7, pp. 48 572–48 634, April 2019.

[260] K. Mendes, F. Lemic, and J. Famaey, "Small UAVs-supported autonomous generation of fine-grained 3D indoor radio environmental maps," in *2022 IEEE 42nd International Conference on Distributed Computing Systems Workshops (ICDCSW)*, July 2022, pp. 296–301.

[261] P. Talarn, B. Ollé, F. Lemic, S. Abadal, and X. Costa-Perez, *Real-time Generation of 3-Dimensional Representations of Static Objects using Small Unmanned Aerial Vehicles*. New York, NY, USA: Association for Computing Machinery, October 2023. [Online]. Available: <https://doi.org/10.1145/3570361.3614078>

[262] A. A. Nasir, H. D. Tuan, T. Q. Duong, and H. V. Poor, "UAV-enabled communication using NOMA," *IEEE Transactions on Communications*, vol. 67, no. 7, pp. 5126–5138, July 2019.

[263] N. Zhao, X. Pang, Z. Li, Y. Chen, F. Li, Z. Ding, and M.-S. Alouini, "Joint trajectory and precoding optimization for UAV-assisted NOMA networks," *IEEE Transactions on Communications*, vol. 67, no. 5, pp. 3723–3735, May 2019.

[264] S. K. Zaidi, S. F. Hasan, X. Gui, N. Siddique, and S. Ahmad, "Exploiting UAV as NOMA based relay for coverage extension," in *2019 2nd International Conference on Computer Applications & Information Security (ICCAIS)*, 2019, pp. 1–5.

[265] Y. Yin, M. Liu, G. Gui, H. Gacanin, and H. Sari, "Minimizing delay for MIMO-NOMA resource allocation in UAV-assisted caching networks," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 4, pp. 4728–4732, April 2023.

[266] H. Peng and L.-C. Wang, "Energy harvesting reconfigurable intelligent surface for UAV based on robust deep reinforcement learning," *IEEE Transactions on Wireless Communications*, vol. 22, no. 10, pp. 6826–6838, October 2023.

[267] F. Huang, J. Chen, H. Wang, G. Ding, Z. Xue, Y. Yang, and F. Song, "UAV-assisted SWIPT in Internet of Things with power splitting: Trajectory design and power allocation," *IEEE Access*, vol. 7, pp. 68 260–68 270, May 2019.

[268] X. Hong, P. Liu, F. Zhou, S. Guo, and Z. Chu, "Resource allocation for secure UAV-assisted SWIPT systems," *IEEE Access*, vol. 7, pp. 24 248–24 257, February 2019.

[269] M. Hua, C. Li, Y. Huang, and L. Yang, "Throughput maximization for UAV-enabled wireless power transfer in relaying system," in *2017 9th International Conference on Wireless Communications and Signal Processing (WCSP)*, 2017, pp. 1–5.

[270] S. Yin, Y. Zhao, L. Li, and F. R. Yu, "UAV-assisted cooperative communications with power-splitting information and power transfer," *IEEE Transactions on Green Communications and Networking*, vol. 3, no. 4, pp. 1044–1057, December 2019.

[271] X. Sun, W. Yang, and Y. Cai, "Secure communication in NOMA-assisted millimeter-wave SWIPT UAV networks," *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 1884–1897, March 2020.

[272] C. Chaccour, M. N. Soorki, W. Saad, M. Bennis, P. Popovski, and M. Debbah, "Seven defining features of terahertz (THz) wireless systems: A fellowship of communication and sensing," *IEEE Communications Surveys and Tutorials*, vol. 24, no. 2, pp. 967–993, Secondquarter (Q2) 2022.

[273] 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 Communications Letters*, vol. 25, no. 6, pp. 1984–1988, June 2021.

[274] Q. Li, A. Nayak, Y. Zhang, and F. R. Yu, "A cooperative recharging-transmission strategy in powered UAV-aided terahertz downlink networks," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 4, pp. 5479–5484, April 2023.

[275] N. Iradukunda, Q.-V. Pham, Z. Ding, and W.-J. Hwang, "THz-enabled UAV communications using non-orthogonal multiple access," *IEEE Transactions on Vehicular Technology*, pp. 1–6, 2023.

[276] C. Han and I. F. Akyildiz, "Distance-aware bandwidth-adaptive resource allocation for wireless systems in the terahertz band," *IEEE Transactions on Terahertz Science and Technology*, vol. 6, no. 4, pp. 541–553, July 2016.

[277] D. P. Bertsekas, "Nonlinear programming," *Journal of the Operational Research Society*, vol. 48, no. 3, p. 334–334, 1997.

[278] J. M. Jornet and I. F. Akyildiz, "Channel modeling and capacity analysis for electromagnetic wireless nanonetworks in the terahertz band," *IEEE Transactions on Wireless Communications*, vol. 10, no. 10, pp. 3211–3221, October 2011.

[279] I. F. Akyildiz, C. Han, Z. Hu, S. Nie, and J. M. Jornet, "Terahertz band communication: An old problem revisited and research directions for the next decade," *IEEE Transactions on Communications*, vol. 70, no. 6, pp. 4250–4285, June 2022.

[280] R. Zhang, R. Tang, Y. Xu, and X. Shen, "Resource allocation for UAV-assisted NOMA systems with dual connectivity," *IEEE Wireless Communications Letters*, vol. 12, no. 2, pp. 341–345, February 2023.

[281] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Unmanned aerial vehicle with underlaid device-to-device communications: Performance and tradeoffs," *IEEE Transactions on Wireless Communications*, vol. 15, no. 6, pp. 3949–3963, June 2016.

[282] X. Liu, Z. Li, N. Zhao, W. Meng, G. Gui, Y. Chen, and F. Adachi, "Transceiver design and multihop D2D for UAV IoT coverage in disasters," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1803–1815, April 2019.

[283] H. Wang, J. Chen, G. Ding, and S. Wang, "D2D communications underlaying UAV-assisted access networks," *IEEE Access*, vol. 6, pp. 46 244–46 255, 2018.

[284] C.-H. Liu, M. A. Syed, and L. Wei, "Toward ubiquitous and flexible coverage of UAV-IRS-assisted NOMA networks," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, April 2022, pp. 1749–1754.

[285] S. Solanki, J. Park, and I. Lee, "On the performance of IRS-Aided UAV networks with NOMA," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 8, pp. 9038–9043, August 2022.

[286] X. Mu, Y. Liu, L. Guo, J. Lin, and H. V. Poor, "Intelligent reflecting surface enhanced multi-UAV NOMA networks," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 10, pp. 3051–3066, June 2021.

[287] O. Maraqa, A. S. Rajasekaran, S. Al-Ahmadi, H. Yanikomeroglu, and S. M. Sait, "A survey of rate-optimal power domain NOMA with enabling technologies of future wireless networks," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 4, pp. 2192–2235, August 2020.

[288] Y. Li, H. Zhang, K. Long, and A. Nallanathan, "Exploring sum rate maximization in UAV-based multi-IRS networks: IRS association, UAV altitude, and phase shift design," *IEEE Transactions on Communications*, vol. 70, no. 11, pp. 7764–7774, September 2022.

[289] C. K. Singh, P. K. Upadhyay, J. Lehtomäki, and M. Juntti, "Outage performance with deep learning analysis for UAV-Borne IRS relaying NOMA systems with hardware impairments," in *2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall)*, September 2022, pp. 1–7.

[290] J. Sipani, P. Sharda, and M. R. Bhatnagar, "Modeling and design of irs-assisted FSO system under random misalignment," *IEEE Photonics Journal*, vol. 15, no. 4, pp. 1–13, August 2023.

[291] C. Chen, S. Huang, H. Abumarsoud, I. Tavakkolnia, M. Safari, and H. Haas, "Frequency-domain channel characteristics of intelligent reflecting surface assisted visible light communication," *Journal of Lightwave Technology*, vol. 41, no. 24, pp. 7355–7369, December 2023.

[292] Q. Zhang, Z. Liu, F. Yang, J. Song, and Z. Han, "Simultaneous lightwave information and power transfer for OIRS-aided VLC system," *IEEE Wireless Communications Letters*, vol. 12, no. 12, pp. 2153–2157, December 2023.

[293] H. Matsuno, T. Ohto, T. Kanno, and T. Hayashi, "Practical evaluation method of large size irs: Synthesis of reflection pattern of sub-irs," *IEEE Access*, vol. 11, pp. 102 072–102 081, 2023.

[294] H. Huang, Y. Zhang, H. Zhang, Y. Cai, A. L. Swindlehurst, and Z. Han, "Disco intelligent reflecting surfaces: Active channel aging for fully-passive jamming attack," *IEEE Transactions on Wireless Communications*, vol. 23, no. 1, pp. 806–819, January 2024.

[295] H. Ajam, M. Najafi, V. Jamali, and R. Schober, "Optical IRSs: Power scaling law, optimal deployment, and comparison with relays," *IEEE Transactions on Communications*, vol. 72, no. 2, pp. 954–970, February 2024.

[296] E. A. Curry and D. K. Borah, "A multi-armed bandit algorithm for IRS-aided VLC system design with device-to-device relays," *IEEE Access*, vol. 12, pp. 15 764–15 777, 2024.

[297] Y. Ata, M. C. Gökçe, and Y. Baykal, "Intelligent reflecting surface aided vehicular optical wireless communication systems using higher-order mode in underwater channel," *IEEE Transactions on Vehicular Technology*, pp. 1–13, 2024.

[298] Y. Ata, H. Abumarsoud, L. Bariah, S. Muhaidat, and M. A. Imran, "Intelligent reflecting surfaces for underwater visible light communications," *IEEE Photonics Journal*, vol. 15, no. 1, pp. 1–10, February 2023.

[299] Z. Liu, F. Yang, S. Sun, J. Song, and Z. Han, "Sum rate maximization for noma-based VLC with optical intelligent reflecting surface," *IEEE Wireless Communications Letters*, vol. 12, no. 5, pp. 848–852, May 2023.

[300] T. Wang, C. You, F. Zhou, and C. Yin, "Base station beamforming design in near-field XL-IRS beam training," *IEEE Communications Letters*, pp. 1–1, 2024.

[301] M. Zhang and X. Yuan, "Intelligent reflecting surface aided MIMO with cascaded los links: Joint beamforming and array orientation optimization," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2024.

[302] F. Ahmed, M. U. Afzal, K. P. Esselle, and D. N. Thalakotuna, "Novel dual-band phase-gradient metascreen and dual-band near-field meta-steering antennas," *IEEE Transactions on Antennas and Propagation*, vol. 72, no. 3, pp. 2202–2216, March 2024.

[303] M. Rahim, T. L. Nguyen, G. Kaddoum, and T. N. Do, "Multi-IRS-aided terahertz networks: Channel modeling and user association with imperfect CSI," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 836–855, 2024.

[304] A. Nordio, L. Dossi, A. Tarable, and G. Virone, "Near-field IRS configuration techniques for wideband signals and THz communications," in *2023 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2023, pp. 1198–1203.

[305] W. Huang, B. Lei, S. He, C. Kai, and C. Li, "Condition number improvement of IRS-aided near-field MIMO channels," in *2023 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2023, pp. 1210–1215.

[306] C. Feng, H. Lu, Y. Zeng, T. Li, S. Jin, and R. Zhang, "Near-field modelling and performance analysis for extremely large-scale IRS communications," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2023.

[307] Y. Pan, C. Pan, S. Jin, and J. Wang, "RIS-aided near-field localization and channel estimation for the terahertz system," *IEEE Journal of Selected Topics in Signal Processing*, vol. 17, no. 4, pp. 878–892, July 2023.

[308] W. Hao, X. You, F. Zhou, Z. Chu, G. Sun, and P. Xiao, "The far-/near-field beam squint and solutions for THz intelligent reflecting surface communications," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 8, pp. 10 107–10 118, August 2023.

[309] N. T. T. Van, N. C. Luong, S. Feng, S. Gong, D. Niyato, and D. I. Kim, "SWIPT-enabled MISO Ad Hoc network underlay RSMA-based system with IRS," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2024.

[310] W. Jiang, B. Ai, M. Li, W. Wu, Y. Pei, and X. Shen, "Aerial IRSs assisted energy-efficient task offloading and computing," *IEEE Internet of Things Journal*, pp. 1–1, 2024.

[311] G. Hu, Q. Wu, D. Xu, K. Xu, J. Si, Y. Cai, and N. Al-Dahir, "Intelligent reflecting surface-aided wireless communication with movable elements," *IEEE Wireless Communications Letters*, pp. 1–1, 2024.

[312] F. Shu, R. Dong, Y. Lin, H. He, W. Shi, Y. Yao, L. Shi, Q. Cheng, J. Li, and J. Wang, "Beamforming and phase shift design for HR-IRS-aided directional modulation network with a malicious attacker," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2024.

[313] Q. Tao, T. Xie, X. Hu, S. Zhang, and D. Ding, "Channel estimation and detection for intelligent reflecting surface-assisted orthogonal time frequency space systems," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2024.

[314] C. Kai, L. Ma, S. He, J. Zhu, and W. Huang, "Delay pre-compensation scheme for single carrier IRS-aided wideband multi-user MISO systems," *IEEE Communications Letters*, vol. 28, no. 2, pp. 362–366, February 2024.

[315] S. Kumar, A. Jee, and S. Prakriya, "Performance analysis and optimization of partitioned-irs-assisted noma network," *IEEE Wireless Communications Letters*, vol. 13, no. 3, pp. 761–765, March 2024.

[316] Y. S. Ribeiro, A. L. F. de Almeida, Fazal-E-Asim, B. Makkii, and G. Fodor, "Low-complexity joint active and passive beamforming design for IRS-assisted MIMO," *IEEE Wireless Communications Letters*, vol. 13, no. 3, pp. 607–611, March 2024.

[317] V. B. Shukla, V. Bhatia, and K. Choi, "Cascaded channel estimator for IRS-aided mmwave hybrid MIMO system," *IEEE Wireless Communications Letters*, vol. 13, no. 3, pp. 622–626, March 2024.

[318] X. Liu, M. Lin, M. Tan, H. Guo, J. Ouyang, and T. Q. S. Quek, "Location-based downlink transmission scheme for IRS-aided integrated satellite-terrestrial networks," *IEEE Transactions on Communications*, vol. 72, no. 2, pp. 1090–1104, February 2024.

[319] Y. Wang, L. Fang, S. Cai, Z. Lian, Y. Su, and Z. Xie, "Low-complexity algorithm for maximizing the weighted sum-rate of intelligent reflecting surface-assisted wireless networks," *IEEE Internet of Things Journal*, vol. 11, no. 6, pp. 10 490–10 499, March 2024.

[320] R. Liang and J. Liang, "Two-hop IRS-assisted wireless communications exploiting statistical CSI," *IEEE Wireless Communications Letters*, pp. 1–1, 2023.

[321] J. Dai, F. Zhu, C. Pan, H. Ren, and K. Wang, "Statistical CSI-based transmission design for reconfigurable intelligent surface-aided massive MIMO systems with hardware impairments," *IEEE Wireless Communications Letters*, vol. 11, no. 1, pp. 38–42, January 2022.

[322] H. Ren, X. Liu, C. Pan, Z. Peng, and J. Wang, "Performance analysis for RIS-aided secure massive MIMO systems with statistical CSI," *IEEE Wireless Communications Letters*, vol. 12, no. 1, pp. 124–128, January 2023.

[323] W. Jiang, P. Xiong, J. Nie, Z. Ding, C. Pan, and Z. Xiong, "Robust design of IRS-aided multi-group multicast system with imperfect CSI," *IEEE Transactions on Wireless Communications*, vol. 22, no. 9, pp. 6314–6328, September 2023.

[324] G. Zhou, C. Pan, H. Ren, K. Wang, and A. Nallanathan, "A framework of robust transmission design for IRS-aided MISO communications with imperfect cascaded channels," *IEEE Transactions on Signal Processing*, vol. 68, pp. 5092–5106, August 2020.

[325] J. Bai, Q. Yan, H.-M. Wang, and Y. Liu, "Intelligent reflecting surface aided green communication with deployment optimization," *IEEE Transactions on Communications*, pp. 1–1, 2024.

[326] H. Huang, L. Dai, H. Zhang, Z. Tian, Y. Cai, C. Zhang, A. L. Swindlehurst, and Z. Han, "Anti-jamming precoding against disco intelligent reflecting surfaces based fully-passive jamming attacks," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2024.

[327] D. Perez-Adan, M. Joham, O. Fresnedo, J. P. Gonzalez-Coma, L. Castedo, and W. Utschick, "Alternating minimization for wideband multiuser IRS-aided MIMO systems under imperfect CSI," *IEEE Transactions on Signal Processing*, vol. 72, pp. 99–114, November 2024.

[328] S. Hong, C. Pan, G. Zhou, H. Ren, and K. Wang, "Outage constrained robust transmission design for IRS-aided secure communications with direct communication links," *IEEE Transactions on Communications*, vol. 72, no. 3, pp. 1548–1564, March 2023.

[329] A. Waraiet, K. Cumanan, Z. Ding, and O. A. Dobre, "Robust design for IRS-assisted MISO-NOMA systems: A DRL-based approach," *IEEE Wireless Communications Letters*, vol. 13, no. 3, pp. 592–596, March 2024.

[330] F. Xu, J. Yao, W. Lai, K. Shen, X. Li, X. Chen, and Z.-Q. Luo, "Coordinating multiple intelligent reflecting surfaces without channel information," *IEEE Transactions on Signal Processing*, vol. 72, pp. 31–46, November 2023.

[331] Y. Shen, C. Wang, W. Zang, L. Xue, B. Yang, and X. Guan, "Outage constrained Max–Min secrecy rate optimization for IRS-aided SWIPT systems with artificial noise," *IEEE Internet of Things Journal*, vol. 11, no. 6, pp. 9814–9828, March 2024.

[332] Y. Liao, J. Liu, X. Chen, Y. Han, Q. Ai, and G.-M. Muntean, "Energy minimization of inland waterway usvs for IRS-assisted hybrid UAV-terrestrial MEC network," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 3, pp. 4121–4135, March 2024.

[333] T. Wang, X. Pang, M. Liu, N. Zhao, A. Nallanathan, and X. Wang, "Offline-online design for energy-efficient IRS-aided UAV communications," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 2, pp. 2942–2947, February 2024.

[334] X. Xiong, B. Zheng, A. L. Swindlehurst, J. Tang, and W. Wu, "A new intelligent reflecting surface-aided electromagnetic stealth strategy," *IEEE Wireless Communications Letters*, pp. 1–1, 2024.

[335] C. Han, Y. Li, Y. Wang, and Z. Yu, "Still waters run deep: Extend THz coverage with non-intelligent reflecting surface," *IEEE Wireless Communications*, vol. 31, no. 1, pp. 41–47, February 2024.

[336] Z. Zhang, Y. Chen, Q. Yu, and L. Dai, *IRS Architecture and Hardware Design*. Wiley-IEEE Press, 2024, pp. 15–36.

[337] A. M. Said, M. Marot, H. Afifi, and H. Mounbla, "Optimal mobile IRS deployment for empowered 6G networks," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 540–552, 2024.

[338] P. Wang, W. Mei, J. Fang, and R. Zhang, "Target-mounted intelligent reflecting surface for joint location and orientation estimation," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 12, pp. 3768–3782, December 2023.

[339] H. Yang, Y. He, M. Tong, and Y. Zhang, "An ultrabroadband reconfigurable polarization conversion metasurface," in *2023 Photonics & Electromagnetics Research Symposium (PIERS)*, 2023, pp. 1905–1908.

[340] A. F. Vaquero, E. Martinez-De-Rioja, M. Arrebola, and J. A. Encinar, "Study on the effect of the wall in the performance of an intelligent reflective surface for providing coverage in mm-wave frequencies," in *2023 17th European Conference on Antennas and Propagation (EuCAP)*, 2023, pp. 1–5.

[341] C. Liaskos, G. G. Pyrialakos, A. Pitilakis, A. Tsoliaridou, M. Christodoulou, N. Kantartzis, S. Ioannidis, A. Pitsillides, and I. F. Akyildiz, *Towards the Internet of Meta-Material Things: Software Enablers for User-Customizable Electromagnetic Wave Propagation*. Wiley-IEEE Press, 2023, pp. 41–82.

[342] W. U. Khan, A. Mahmood, C. K. Sheemar, E. Lagunas, S. Chatzinotas, and B. Ottersten, "Reconfigurable intelligent surfaces for 6G non-terrestrial networks: Assisting connectivity from the sky," *IEEE Internet of Things Magazine*, vol. 7, no. 1, pp. 34–39, January 2024.

[343] W. U. Khan, A. Mahmood, E. Lagunas, M. A. Jamshed, S. Chatzinotas, and B. Ottersten, "Resource optimization for integrated terrestrial non-terrestrial networks involving IRS," in *2023 IEEE Globecom Workshops (GC Wkshps)*, 2023, pp. 1710–1715.

[344] A. A. Siddig, A. Al-Dweik, A. Al-Rimawi, Y. Iraqi, A. Pandey, and J.-P. Giacalone, "On the performance of IRS-assisted IoT-NTN with joint imperfect phase estimation and quantization," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 256–275, 2024.

[345] K. P. Rajput, L. Wu, M. R. Bhavani Shankar, and P. K. Varshney, "Joint transmit precoders and passive reflection beamformer design in IRS-aided IoT networks," in *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024, pp. 156–160.

[346] X. Guan, J. Xue, H. Jiang, and G. Tian, "Multi-objective optimization of wireless powered communication networks assisted by intelligent reflecting surface based on multi-agent reinforcement learning," *IEEE Transactions on Antennas and Propagation*, pp. 1–1, 2024.

[347] J. Yaswanth, M. Katwe, K. Singh, O. Taghizadeh, C. Pan, and A. Schmeink, "Towards green communication: Power-efficient beamforming for STAR-RIS-aided SWIPT," *IEEE Transactions on Green Communications and Networking*, pp. 1–1, 2024.

[348] N. H. Tu and K. Lee, "IRS-assisted coordinated direct and multiantenna relay transmission for MIMO SWIPT systems," *IEEE Systems Journal*, pp. 1–12, 2024.

[349] S. Li and Q. Song, "User-adaptive IRS beamforming design in hybrid wireless powered communication networks," *IEEE Transactions on Vehicular Technology*, pp. 1–6, 2024.

[350] A. Taneja and S. Rani, "A novel energy conservation scheme for IoT based wireless networks: A use case of E-commerce systems for consumer electronics," *IEEE Transactions on Consumer Electronics*, pp. 1–1, 2024.

[351] I. Hameed and I. Koo, "Max-Min throughput optimization in WPCNs: A hybrid active/passive IRS-assisted scheme," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 1123–1140, 2024.

[352] J. Wang and H. Yu, "Energy balance-oriented RF energy harvesting schemes for intelligent reflecting surface-assisted cognitive radio sensor networks," *IEEE Sensors Journal*, vol. 24, no. 6, pp. 8628–8641, March 2024.

[353] X. Zhu, Z. Liu, Y. Shi, H. Di, Z. Meng, and X. Tu, "Transmission design for intelligent reflecting surface-aided MIMO SWIPT systems with finite-alphabet inputs," *IEEE Transactions on Communications*, pp. 1–1, 2024.

[354] Y. Gao, Q. Wu, W. Chen, Y. Liu, M. Li, and D. B. Da Costa, "IRS-aided overloaded multi-antenna systems: Joint user grouping and resource allocation," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2024.

[355] Z. Chu, P. Xiao, D. Mi, C. Yin, W. Hao, W. Liu, and A. C. Sodre, "Jointly active and passive beamforming designs for irs-empowered WPCN," *IEEE Internet of Things Journal*, vol. 11, no. 7, pp. 11 579–11 592, April 2024.

[356] P. Chen, Y. Yang, J. Jiang, B. Lyu, Z. Yang, and A. Jamalipour, "Computational rate maximization for IRS-assisted multiantenna WP-MEC systems with finite edge computing capability," *IEEE Internet of Things Journal*, vol. 11, no. 4, pp. 6607–6621, February 2024.

[357] B. Li, J. Liao, W. Wu, and Y. Li, "Intelligent reflecting surface assisted secure computation of wireless powered MEC system," *IEEE Transactions on Mobile Computing*, vol. 23, no. 4, pp. 3048–3059, April 2024.

[358] H. Zhang and J. Zheng, "IRS-assisted secure radar communication systems with malicious target," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 1, pp. 591–604, January 2024.

[359] J. Chen, K. Zhu, K. Wu, J. Niu, and J. A. Zhang, "Introducing user grouping to counteract channel correlation in irs-assisted ISAC," *IEEE Communications Letters*, pp. 1–1, 2024.

[360] F. Zhao, W. Chen, Z. Liu, J. Li, and Q. Wu, "Deep reinforcement learning-based intelligent reflecting surface optimization for TDD multi-user MIMO systems," *IEEE Wireless Communications Letters*, vol. 12, no. 11, pp. 1951–1955, November 2023.

[361] Q. T. Ngo, K. T. Phan, A. Mahmood, and W. Xiang, "Hybrid IRS-assisted secure satellite downlink communications: A fast deep reinforcement learning approach," *IEEE Transactions on Emerging Topics in Computational Intelligence*, pp. 1–12, 2024.

[362] S. Zheng, S. Wu, H. Jia, C. Jiang, and L. Kuang, "Hybrid driven learning for joint activity detection and channel estimation in IRS-assisted massive connectivity," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2024.

[363] H. Shi, Y. Huang, S. Jin, Z. Wang, and L. Yang, "Automatic high-performance neural network construction for channel estimation in IRS-aided communications," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2024.

[364] W. Lai and K. Shen, "Blind beamforming for intelligent reflecting surface: A reinforcement learning approach," in *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024, pp. 8956–8960.

[365] D. Zhang, M. Xiao, Z. Pang, L. Wang, and H. V. Poor, "IRS assisted federated learning: A broadband Over-the-Air aggregation approach," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2024.

[366] H. M. Hariz, S. Sheikhzadeh, N. Mokari, M. R. Javan, B. Abbasi-Arand, and E. A. Jorswieck, "Ai-based radio resource management and trajectory design for IRS-UAV-assisted PD-NOMA communication," *IEEE Transactions on Network and Service Management*, pp. 1–1, 2024.

[367] T.-J. Yeh, W.-C. Tsai, C.-W. Chen, and A.-Y. Wu, "Enhanced-GNN with angular CSI for beamforming design in IRS-assisted mmwave communication systems," *IEEE Communications Letters*, pp. 1–1, 2024.

[368] S. Dominic and L. Jacob, "A distributed learning algorithm for power control in energy efficient IRS assisted SISO NOMA networks," *IEEE Transactions on Green Communications and Networking*, pp. 1–1, 2024.

[369] D. Pereira-Ruisánchez, O. Fresnedo, D. Pérez-Adan, and L. Castedo, "DRL-based sequential scheduling for IRS-assisted MIMO communications," *IEEE Transactions on Vehicular Technology*, pp. 1–15, 2024.

[370] J. Zhang, Z. Wang, J. Li, Q. Wu, W. Chen, F. Shu, and S. Jin, "How often channel estimation is required for adaptive IRS beamforming: A bilevel deep reinforcement learning approach," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2024.

[371] M. Shehab, M. Elsayed, A. Almohamad, A. Badawy, T. Khattab, N. Zorba, M. Hasna, and D. Trinchero, "Terahertz multiple access: A deep reinforcement learning controlled multihop IRS topology," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 1072–1087, 2024.

[372] X. Yu and D. Li, "Phase shift compression for control signaling reduction in IRS-aided wireless systems: Global attention and lightweight design," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2024.

[373] S. Idrees, S. Durrani, Z. Xu, X. Jia, and X. Zhou, "Joint active and passive beamforming for IRS-assisted monostatic backscatter systems: An unsupervised learning approach," *IEEE Transactions on Machine Learning in Communications and Networking*, pp. 1–1, 2024.

[374] M. Alishahi, P. Fortier, M. Zeng, Q.-V. Pham, and X. Li, "Energy minimization for IRS-aided wireless powered federated learning networks with NOMA," *IEEE Internet of Things Journal*, pp. 1–1, 2024.

[375] L. Dong, F. Jiang, M. Wang, Y. Peng, and X. Li, "Deep progressive reinforcement learning-based flexible resource scheduling framework for IRS and UAV-assisted MEC system," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–13, 2024.

[376] H. Hashmi, S. Pougkakiotis, and D. Kalogerias, "Model-free learning of two-stage beamformers for passive IRS-aided network design," *IEEE Transactions on Signal Processing*, vol. 72, pp. 652–669, 2023.

[377] Z. Li, H. Shen, W. Xu, D. Chen, and C. Zhao, "Deep learning-based adaptive phase shift compression and feedback in IRS-assisted communication systems," *IEEE Wireless Communications Letters*, vol. 13, no. 3, pp. 766–770, 2024.

[378] X. Hu, K. Wong, C. Masouros, and S. Jin, *IRS-Aided Mobile Edge Computing: From Optimization to Learning*. Wiley-IEEE Press, 2024, pp. 207–228.

[379] G. Zhu, J. Hu, K. Zhong, X. Cheng, and Z. Song, "Sum-path-gain maximization for IRS-aided MIMO communication system via riemannian gradient descent network," *IEEE Signal Processing Letters*, vol. 31, pp. 51–55, 2023.

[380] Z. U. A. Tariq, E. Baccour, A. Erbad, and M. Hamdi, "Reinforcement learning for resilient aerial-IRS assisted wireless communications networks in the presence of multiple jammers," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 15–37, 2024.

[381] W. Xu, J. Yu, Y. Wu, and D. H.-K. Tsang, "Joint channel estimation and reinforcement-learning-based resource allocation of intelligent-reflecting-surface-aided multicell mobile edge computing," *IEEE Internet of Things Journal*, vol. 11, no. 7, pp. 11 862–11 875, 2024.

[382] C. Chen, J. Zhang, T. Lu, M. Sandell, and L. Chen, "Secret key generation for irs-assisted multi-antenna systems: A machine learning-based approach," *IEEE Transactions on Information Forensics and Security*, vol. 19, pp. 1086–1098, 2023.

[383] A. A. Hammadi, L. Bariah, S. Muhamadat, M. Al-Qutayri, P. C. Sofotasios, and M. Debbah, "Deep Q-learning-based resource management in IRS-assisted VLC systems," *IEEE Transactions on Machine Learning in Communications and Networking*, vol. 2, pp. 34–48, 2023.

[384] T. Ohyama, Y. Kawamoto, and N. Kato, "Resource allocation optimization by quantum computing for shared use of standalone IRS," *IEEE Transactions on Emerging Topics in Computing*, vol. 11, no. 4, pp. 950–961, 2023.

[385] ——, "Intelligent reflecting surface (IRS) allocation scheduling method using combinatorial optimization by quantum computing," *IEEE Transactions on Emerging Topics in Computing*, vol. 10, no. 3, pp. 1633–1644, 2022.

[386] ——, "Quantum computing based optimization for intelligent reflecting surface (IRS)-aided cell-free network," *IEEE Transactions on Emerging Topics in Computing*, vol. 11, no. 1, pp. 18–29, 2023.

[387] N. K. Kundu, M. R. McKay, R. Murch, and R. K. Mallik, "Intelligent reflecting surface-assisted free space optical quantum communications," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2023.

[388] S. Zhang, H. Gao, Y. Su, J. Cheng, and M. Jo, "Intelligent mixed reflecting/relay-ing surface-aided secure wireless communications," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 1, pp. 532–543, 2024.

[389] Z. Yang, F. Li, and D. Zhang, "A joint model extraction and data detection framework for IRS-NOMA system," *IEEE Transactions on Signal Processing*, vol. 71, pp. 164–177, 2023.

[390] P. Staat, H. Elders-Boll, M. Heinrichs, R. Kronberger, C. Zenger, and C. Paar, "Intelli-gent reflecting surface-assisted wireless key generation for low-entropy environments," in *2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2021, pp. 745–751.

[391] R. T. Rockafellar, *Convex Analysis*, ser. Princeton Mathematical Series. Princeton, N. J.: Princeton University Press, 1970.

[392] C. Goh, *Duality in optimization and variational inequalities*. CRC Press, 2002.

[393] R. T. Rockafellar, *Conjugate Duality and Optimization*, ser. Regional Conference Series in Applied Mathematics. Philadelphia: Society for Industrial and Applied Mathematics, 1974, vol. 16, bibliography: p. 73-74. "Essentially to be regarded as supplement to the book Convex Analysis.". [Online]. Available: https://openlibrary.org/works/OL5120075M/Conjugate_duality_and_optimization

[394] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "An augmented la-grangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 681–695, March 2011.

[395] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*. John Wiley & Sons, Inc., 2005.

[396] D. G. Luenberger, *Optimization by Vector Space Methods*. 605 Third Ave. New York, NY United States: John Wiley & Sons, Inc., January 1997, published 01 January 1997.

[397] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed., ser. Springer Series in Operations Research and Financial Engineering. New York, NY: Springer New York, NY, 2006. [Online]. Available: <https://doi.org/10.1007/978-0-387-40065-5>

[398] C. Roos, T. Terlaky, and J.-P. Vial, *Interior Point Methods for Linear Optimization*, ser. International Series in Operations Research & Management Science. New York, NY: Springer, 2005, vol. 95.

[399] L. J. Hong, Y. Yang, and L. Zhang, "Sequential convex approximations to joint chance constrained programs: A monte carlo approach," *Operations Research*, vol. 59, no. 3, pp. 617–630, 2011.

[400] A. L. Yuille and A. Rangarajan, "The concave-convex procedure," *Neural computation*, vol. 15, no. 4, pp. 915–936, 2003.

[401] M. Razaviyayn, "Successive convex approximation: analysis and applications," Ph.D. dissertation, University of Minnesota, 2014.

[402] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Operations Research*, vol. 26, no. 4, pp. 681–683, 1978.

[403] Q. T. Dinh and M. Diehl, "Local convergence of sequential convex programming for nonconvex optimization," in *Recent Advances in Optimization and its Applications in Engineering*. Springer, 2010, pp. 93–102.

[404] A. Makhorin, "GNU Linear Programming Kit Reference Manual," December 2020, department for Applied Informatics, Moscow Aviation Institute, Moscow, Russia. [Online]. Available: <https://www.gnu.org/>

[405] Gurobi Optimization, LLC, "Gurobi Optimizer Reference Manual," 2023. [Online]. Available: <https://www.gurobi.com>

[406] E. D. Andersen and K. D. Andersen, *The Mosek Interior Point Optimizer for Linear Programming: An Implementation of the Homogeneous Algorithm*. Boston, MA: Springer US, 2000, pp. 197–232. [Online]. Available: https://doi.org/10.1007/978-1-4757-3216-0_8

[407] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization methods and software*, vol. 11, no. 1-4, pp. 625–653, 1999.

[408] R. H. Tütüncü, K.-C. Toh, and M. J. Todd, "Solving semidefinite-quadratic-linear programs using SDPT3," *Mathematical Programming, Series B*, vol. 95, pp. 189–217, 2003. [Online]. Available: <https://doi.org/10.1007/s10107-002-0347-5>

[409] K.-C. Toh, M. J. Todd, and R. H. Tütüncü, "SDPT3 — A matlab software package for semidefinite programming, Version 1.3," *Optimization Methods and Software*, vol. 11, no. 1-4, pp. 545–581, 1999. [Online]. Available: <https://doi.org/10.1080/10556789908805762>

- [410] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110, http://stanford.edu/~boyd/graph_dcp.html.
- [411] S. Diamond and S. Boyd, "CVXPY: A Python-embedded modeling language for convex optimization," *Journal of Machine Learning Research*, vol. 17, no. 83, pp. 1–5, 2016.
- [412] J. Lofberg, "YALMIP : a toolbox for modeling and optimization in MATLAB," in *2004 IEEE International Conference on Robotics and Automation (IEEE Cat. No.04CH37508)*, September 2004, pp. 284–289.
- [413] B. Stellato, G. Banjac, P. Goulart, A. Bemporad, and S. Boyd, "OSQP: An operator splitting solver for quadratic programs," in *2018 UKACC 12th International Conference on Control (CONTROL)*, September 2018, pp. 339–339.
- [414] R. Lougee-Heimer, "The Common Optimization INterface for operations research: Promoting open-source software in the operations research community," *IBM Journal of Research and Development*, vol. 47, no. 1, pp. 57–66, January 2003.

