

LEARNNG - BASED RIS ELEMENTS PARTITIONING FOR ISAC USING MULTI-ARMED BANDITS

A Thesis Submitted
in Partial Fulfilment of the Requirements
for the Degree of

**MASTER OF TECHNOLOGY
IN
SIGNAL PROCESSING AND DIGITAL DESIGN**

Submitted by

ANSHIKA SAINI

(24/SPD/22)

Under the supervision of

Dr. ROHIT KUMAR



DEPARTMENT NAME
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi 110042

MAY, 2026

**DEPARTMENT OF ELECTRONICS AND COMMUNICATION
ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042**

CANDIDATE'S DECLARATION

I, **Anshika Saini**, Roll No's – **24/SPD/22** students of M.Tech Signal Processing and Digital Design, hereby declare that the project Dissertation titled “**Learning - Based RIS Elements Partitioning for ISAC Using MultiArmed Bandits**” which is submitted by me to the Electronics and Communication Department, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Place: Delhi



Anshika Saini

Date: 28.05.26

**DEPARTMENT OF ELECTRONICS AND COMMUNICATION
ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042**

CERTIFICATE

I hereby certify that the Project Dissertation titled “**Learning - Based RIS Elements Partitioning for ISAC Using MultiArmed Bandits**” which is submitted by **Anshika Saini**, Roll No’s – **24/SPD/22**, Electronics and Communication Department, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Masters of Technology, is a record of the project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.



Place: Delhi

Dr. Rohit Kumar

Date: 28.05.2026

SUPERVISOR

**DEPARTMENT OF ELECTRONICS AND COMMUNICATION
ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042**

ACKNOWLEDGEMENT

I wish to express my sincerest gratitude to **Dr. Rohit Kumar** for his continuous guidance and mentorship that he provided us during the project. He showed me the path to achieve my targets by explaining all the tasks to be done and explained to me the importance of this project as well as its industrial relevance. He was always ready to help us and clear our doubts regarding any hurdles in this project. Without his constant support and motivation, this project would not have been successful.

Place: Delhi



Anshika Saini

Date: 28.05.2026

LEARNING-BASED RIS ELEMENT PARTITIONING FOR ISAC USING MULTI-ARMED BANDITS

ANSHIKA SAINI

ABSTRACT

Integrated Sensing and Communications (ISAC) is one of the emerging paradigms for future generation wireless communications which enables communication and sensing to be performed simultaneously by making use of common spectrum and hardware infrastructure. Performance of ISAC can further be boosted through the usage of RISs that manipulate wireless channels intelligently.

The objective of this thesis is to optimize partition of RIS into communication and sensing sub-functions with the aid of reinforcement learning algorithm with Gaussian TS and UCB approaches applied to RIS-assisted ISAC systems under fading conditions. A realistic 4-hop monostatic sensing channel model is considered along with Nakagami fading based wireless propagation. The communication and sensing objectives are jointly optimized using a geometric mean reward formulation that ensures balanced ISAC performance.

The proposed algorithm examines different partition setups of the RIS and determines the best partitioning policy through a series of interactions within the wireless environment. It is shown that through extensive simulations that the Gaussian Thompson Sampling approach performs significantly better than UCB in terms of cumulative regret, convergence time, average reward, and stability. The results achieved confirm the applicability of reinforcement learning in online RIS partition optimization in upcoming ISAC-based wireless communication systems.

Contents

Candidate's Declaration	i
Certificate	ii
Acknowledgement	iii
Abstract	iv
List of Symbols, Abbreviations and Nomenclature	xi
1 INTRODUCTION	1
1.1 Background	1
1.2 Motivation	1
1.3 Problem Statement	2
2 Literature Review	3
2.1 Reconfigurable Intelligent Surface Assisted Wireless Communication	3
2.2 Integrated Sensing and Communication (ISAC)	4
2.3 Bandit Algorithms for Wireless Systems	5
2.4 RIS-ISAC Integration	5
3 RIS and ISAC Fundamentals	6
3.1 Reconfigurable Intelligent Surface (RIS)	6
3.1.1 Basic Principle of RIS	7
3.1.2 RIS Reflection Coefficient	7
3.1.3 RIS Phase Alignment	7
3.1.4 Advantages of RIS	8
3.2 Integrated Sensing and Communication (ISAC)	8
3.2.1 Principles of ISAC	8
3.2.2 Applications of ISAC	9
3.3 Communication-Sensing Tradeoff	10
3.4 RIS-Assisted ISAC Systems	10
3.4.1 RIS Partitioning	10
3.5 Monostatic Sensing Model	10
3.5.1 4-Hop Sensing Channel	11
3.6 Need for Intelligent RIS Optimization	11
4 Reinforcement Learning Techniques	12

4.1	Introduction	12
4.2	Fundamentals of Reinforcement Learning	12
4.2.1	Agent	13
4.2.2	Environment	13
4.2.3	Action Space	13
4.2.4	Reward	14
4.3	Exploration and Exploitation Tradeoff	14
4.4	Multi-Armed Bandit Framework	14
4.4.1	Bandit Interpretation	14
4.4.2	RIS Partition Arms	14
4.5	Reward Optimization Objective	15
4.6	Gaussian Thompson Sampling	15
4.6.1	Benefits of Thompson Sampling	16
4.7	Upper Confidence Bound	16
4.7.1	Benefits of UCB	17
4.8	Regret Metrics	17
4.9	Algorithm Flow	18
4.10	Gaussian Thompson Sampling Algorithm	18
4.11	Upper Confidence Bound Algorithm	19
4.12	Comparison Between TS and UCB	19
4.13	Need for Reinforcement Learning in RIS-ISAC	19
5	Problem Formulation	20
5.1	Introduction	20
5.2	RIS Partition Optimization Framework	20
5.3	System Geometry	21
5.4	Path Loss Model	21
5.5	Communication Channel Model	22
5.6	Sensing Channel Model	22
5.7	Communication SNR Formulation	23
5.8	Sensing SNR Formulation	24
5.9	Communication-Sensing Tradeoff	24
5.10	Reward Formulation	24
5.10.1	Arithmetic Reward	24
5.10.2	Geometric Reward	24
5.11	Optimization Objective	25
5.12	Multi-Armed Bandit Formulation	25
5.13	Regret Formulation	26
6	Proposed Methodology	27
6.1	Introduction	27
6.2	Overall Framework Architecture	27
6.3	RIS Partition Configuration	28
6.4	Wireless Channel Generation Methodology	28
6.4.1	Communication Channel Generation	28
6.4.2	Sensing Channel Generation	29

6.5	RIS Phase Alignment Strategy	29
6.5.1	Communication Phase Compensation	29
6.5.2	Sensing Phase Compensation	30
6.6	Coherent Combining	30
6.6.1	Communication Combining	30
6.6.2	Sensing Combining	30
6.7	Static Arm Evaluation	30
6.8	Balanced Reward Design	31
6.9	Gaussian Thompson Sampling Methodology	31
6.9.1	Posterior Initialization	31
6.9.2	Sampling Procedure	31
6.9.3	Reward Observation	32
6.9.4	Posterior Update	32
6.10	Upper Confidence Bound Methodology	32
6.10.1	UCB Index Computation	32
6.10.2	Arm Selection Policy	33
6.10.3	Reward Update	33
6.10.4	Application to RIS Partition Optimization	33
6.11	Simulation Procedure	33
6.12	Performance Metrics	34
6.13	Computational Complexity	34
6.14	Advantages of the Proposed Methodology	34
7	Simulation Setup	36
7.1	Simulation Environment	36
7.2	RIS Configuration	36
7.3	RIS Partition Arms	36
7.4	Wireless System Parameters	37
7.5	Channel Parameters	37
7.6	Antenna Gain Parameters	37
7.7	Geometrical Configuration	37
7.8	Reinforcement Learning Parameters	38
7.9	Compared Algorithms	38
7.9.1	Gaussian Thompson Sampling	39
7.9.2	Upper Confidence Bound	39
7.10	Performance Evaluation Metrics	39
7.11	Simulation Workflow	39
8	Results and Performance Analysis	40
8.1	Introduction	40
8.2	Communication-Sensing Tradeoff Analysis	40
8.3	Static RIS Partition Performance	42
8.4	Oracle RIS Partition Identification	42
8.5	Gaussian Thompson Sampling Performance	43
8.6	Comparative Regret Analysis	44
8.7	Average Reward Comparison	45

8.8	RIS Arm Selection Analysis	46
8.9	Average Regret Analysis	46
8.10	Discussion	47
8.11	Practical Implications	47
9	Conclusion, Future Scope and Social Impact	48
9.1	Conclusion	48
9.2	Future Scope	48
9.3	Social Impact of the Proposed Work	49

List of Tables

4.1	Comparison between Gaussian TS and UCB	19
7.1	Simulation Environment	36
7.2	RIS Configuration Parameters	36
7.3	RIS Partition Arms	37
7.4	Wireless System Parameters	37
7.5	Nakagami Fading Parameters	37
7.6	Antenna Gain Parameters	38
7.7	Node Positions in Simulation Geometry	38
7.8	Reinforcement Learning Parameters	38
8.1	Per-arm oracle statistics ($T = 1,000$ trials).	41
8.2	Oracle RIS Partition	42
8.3	Gaussian TS Performance	44

List of Figures

3.1	Reconfigurable Intelligent Surface (RIS)	6
3.2	Integrated Sensing and Communication	9
3.3	Reconfigurable Integrated Surface (RIS) System Model	11
6.1	Overall RIS-assisted ISAC optimization framework	28
8.1	Communication and sensing SNR versus communication RIS allocation.	41
8.2	Reward versus RIS partition allocation.	42
8.3	Cumulative regret of Gaussian Thompson Sampling.	43
8.4	TS versus UCB cumulative regret comparison.	44
8.5	Average reward comparison.	45
8.6	RIS arm selection frequency.	46
8.7	Average regret comparison.	46

List of Symbols, Abbreviations and Nomenclature

N	Total number of RIS elements
L_c	Number of RIS elements allocated for communication
L_s	Number of RIS elements allocated for sensing
d_{SR}	Distance between Base Station and RIS
d_{RD}	Distance between RIS and communication receiver
d_{RT}	Distance between RIS and sensing target
h_c	Communication channel coefficient
h_s	Sensing channel coefficient
ρ_t	Target reflection coefficient
P_t	Transmit power
σ^2	Noise power
γ_c	Communication Signal-to-Noise Ratio
γ_s	Sensing Signal-to-Noise Ratio
R_t	Instantaneous reward at trial t
μ_i	Mean reward of arm i
$\hat{\mu}_i$	Estimated reward of arm i
$N_i(t)$	Number of times arm i is selected till trial t
$\mathcal{A}$	Set of available RIS partitioning arms
TS	Thompson Sampling
GTS	Gaussian Thompson Sampling
UCB	Upper Confidence Bound
RIS	Reconfigurable Intelligent Surface
ISAC	Integrated Sensing and Communication
RL	Reinforcement Learning
SNR	Signal-to-Noise Ratio
LoS	Line-of-Sight
NLoS	Non-Line-of-Sight

Chapter 1

INTRODUCTION

1.1 Background

The exponential growth of next-generation wireless networks has accelerated the demand for unprecedented spectrum, energy, and hardware efficiency. Historically, wireless communication networks and radar sensing frameworks have been developed as isolated, independent systems. Mutual interference, redundant hardware deployment, and highly inefficient utilization of the increasingly congested electromagnetic spectrum were the cons of traditional approach.

Integrated Sensing and Communication (ISAC) has emerged as gamechanging technology for sixth-generation (6G) wireless systems to overcome above limitations, . By merging radar sensing and data transmission functionalities into a unified hardware architecture and a shared spectral platform, ISAC enables simultaneous information transfer and environmental mapping.

In parallel, Reconfigurable Intelligent Surfaces (RIS) have gained substantial research traction as a revolutionary paradigm for transforming the wireless propagation environment from a passive, highly lossy medium into a software-defined, programmable channel. Consisting of low-cost, software-controlled metasurfaces, an RIS dynamically optimizes the phase shifts of incident electromagnetic waves to engineer favorable propagation conditions.

This integration of RIS-enabled ISAC systems leads to a completely new era for designing wireless networks. With the use of programmable surfaces in an ISAC architecture, future wireless networks will be capable of effectively handling blockages, improving multi-user throughput efficiency, and increasing target sensing resolution.

1.2 Motivation

In traditional RIS-assisted systems generally allocate RIS resources entirely toward communication enhancement. However, in ISAC systems, RIS resources must be jointly optimized between:

- Communication performance
- Sensing performance

Above introduces a challenging tradeoff problem because increasing RIS allocation toward communication improves communication SNR while degrading sensing SNR, and vice versa.

Moreover, the wireless channels are stochastic and time-varying due to fading and environmental uncertainty. Therefore, static optimization methods become insufficient for practical deployment.

Reinforcement Learning (RL) provides an attractive solution by enabling adaptive online optimization without requiring prior knowledge of wireless channel distributions.

1.3 Problem Statement

The objective of this thesis is to develop a reinforcement learning based RIS partition optimization framework capable of:

- Dynamically allocating RIS elements between communication and sensing
- Maximizing balanced ISAC performance
- Operating under stochastic wireless fading environments

Chapter 2

Literature Review

2.1 Reconfigurable Intelligent Surface Assisted Wireless Communication

The foundational theoretical framework for RIS was laid by Liaskos et al. [1], who proposed the concept of programmable wireless environments using software-defined meta-surfaces. The early work of Wu and Zhang [2] established the first beamforming optimization framework for RIS-assisted multi-user MISO systems, demonstrating that RIS can significantly improve spectral efficiency via alternating optimization of the active precoder at the BS and the passive phase shifts at the RIS. Subsequent work by Huang et al. [3] studied energy efficiency and showed that RIS can achieve higher energy efficiency than conventional massive MIMO with significantly fewer radio-frequency chains, making it an appealing candidate for green communications.

The channel estimation problem for RIS has received considerable attention. Since RIS elements are passive, they cannot directly measure the incident signal; instead, the aggregate channel from source to RIS to destination must be estimated jointly. Mishra and Johansson [4] proposed a compressed sensing-based protocol exploiting the spatial correlation structure of RIS channels. Jensen and de Carvalho [5] derived the Cramer-Rao lower bound for cascaded RIS channel estimation and proposed an optimal pilot design. More recently, deep learning-based approaches have been explored to reduce the pilot overhead in rapidly time-varying environments [6].

RIS is composed of a large number of low-cost reflecting elements known as meta-atoms. Each element can independently adjust the phase of the incident signal through intelligent controllers.

The reflected signal from the RIS can be represented as:

$$y = \sum_{n=1}^N h_n \theta_n g_n x \quad (2.1)$$

where:

- h_n : channel from transmitter to RIS
- g_n : channel from RIS to receiver
- θ_n : RIS phase shift
- N : number of RIS elements

By properly configuring the RIS phase shifts, constructive signal combining can be achieved at the receiver.

Previous relay systems use active RF chains and need power amplification, thus leading to high energy consumption. The operation of RIS is passive and just needs low power electronic controls for phase shifting.

Various researches has reported increased efficiency of RIS-assisted wireless communication systems when compared to relay-based wireless communication systems [7].

RIS beamforming has become a major research area in intelligent wireless communication systems. By jointly optimizing transmit beamforming and RIS phase shifts, significant gains in:

- Spectral efficiency
- Signal-to-noise ratio
- Reliability

can be achieved.

However, RIS beamforming optimization is generally non-convex due to coupled optimization variables and unit modulus constraints.

2.2 Integrated Sensing and Communication (ISAC)

The Integrated Sensing and Communication (ISAC) technology is regarded as one of the emerging technologies to achieve future generations of wireless communications, namely 6G. As opposed to traditional communication and sensing technologies that

act independently of each other, ISAC technology offers sensing and communication features on the same platform in terms of spectrum sharing and hardware design.

The idea of ISAC technology is built upon coexisting radar and communication literature in the early years of the 2000s, in which both the radar and communication technologies are made to share spectrum efficiently by minimizing mutual interference [8]. What made the joint radar and communication design approach a milestone was the proposal of the DFRC beamformer by Liu et al. [9] that not only achieves the required minimum communication SINR but also maximizes beam pattern similarity. This paper introduced the trade-off between the quality of sensing beam patterns and multi-user capability. In addition, multi-user ISAC systems with several communication users and target objects were studied by Liu et al. [6], in which the precoding design is performed with a weighted approach between sum rate and integrated sidelobe level criteria. ISAC using mmWave communications were investigated by Cheng et al. [10] that take advantage of high-bandwidth and selective spatial characteristics of the mmWave channels.

2.3 Bandit Algorithms for Wireless Systems

The multi-armed bandit problem, introduced by Robbins [11] and formalized by Lai and Robbins [12], provides a mathematical framework for sequential decision-making under uncertainty. The fundamental result of Lai and Robbins establishes that the optimal regret growth rate for any consistent algorithm is $O(\log T)$, and Thompson Sampling [13] achieves this bound asymptotically. The upper confidence bound (UCB) algorithm of Auer et al. [14] provides a deterministic, computationally simple alternative with the same logarithmic regret guarantee.

Thompson Sampling, originally proposed in 1933 [13], experienced a renaissance after Chapelle and Li [15] demonstrated its practical superiority over UCB variants in online advertising. Agrawal and Goyal [16] provided the first rigorous frequentist regret analysis for Gaussian Thompson Sampling, confirming the $O(\log T)$ bound. The Gaussian model is particularly suitable for reward distributions that are approximately normal (by central limit theorem arguments) and for which the mean reward is the key unknown parameter.

2.4 RIS-ISAC Integration

The study of RIS-enabled ISAC systems represents a recent trend that has gained much momentum in a short period of time. Song et al. [17] were among the early investigators in this area, where the optimal design of RIS phase shifts together with the transmit

beamformer was sought with the objective of minimizing the trade-off between radar signal-to-interference-and-noise ratio (SINR) and the communication bit-rate.

Their alternating optimization strategy provided substantial improvements over the RIS-free scenario, especially in cases when there exists no line-of-sight path. Hua et al. [18] explored the basic communication-capacity versus estimation-rate trade-off problem in RIS-ISAC systems and established the Pareto boundary of the rate-distortion region, showing that the presence of the RIS provides a new degree of freedom that leads to a larger rate-distortion region than the one achieved in the absence of the RIS. Also, the authors concluded that RIS optimization must incorporate channel state information (CSI) from both the sensing and the communication links. The joint optimization of the transmitted radar waveform and RIS phase shift pattern for target detection in a highly cluttered environment was analyzed by Zheng et al. [19]. The authors modeled the problem in question using non-convex optimization and solved it with successive convex approximation (SCA).

Chapter 3

RIS and ISAC Fundamentals

3.1 Reconfigurable Intelligent Surface (RIS)

RIS represent a paradigm shift in wireless communications by transforming the traditionally uncontrollable wireless propagation environment into a software-defined, programmable medium. Implemented via engineered metasurfaces, an RIS consists of an array of low-cost, passive scattering elements known as meta-atoms.

Each element can independently manipulate the phase, amplitude, or polarization of incident electromagnetic waves via an external smart controller. By dynamically optimizing phase shifts mentioned, the RIS constructively aligns scattered signals to create highly directional, passive beamforming links.

RIS creates a virtual line of sight (LOS) connection between the transmitter and the receiver, and it does not use the active radio frequency chain for it, providing a power-efficient and economical solution to remove dead zones and improve spectral efficiency in upcoming future (5G-Advanced & 6G) network technologies. Figure 3.1. shows the basic structure of RIS.

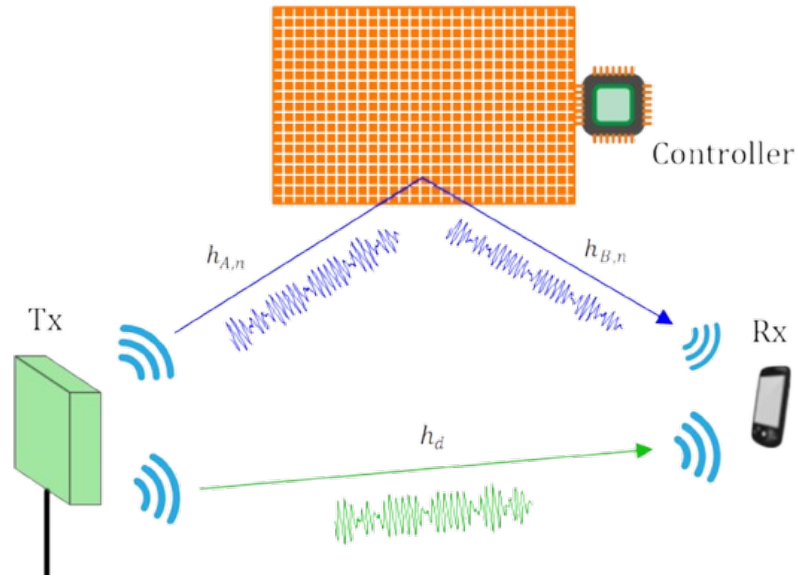


Figure 3.1: Reconfigurable Intelligent Surface (RIS)

3.1.1 Basic Principle of RIS

Consider a wireless signal transmitted from a Base Station (BS) toward a user through an RIS-assisted link.

The RIS receives the incident signal and reflects it after applying programmable phase shifts.

The received signal at the user can be represented as:

$$y = \sum_{n=1}^N h_n \theta_n g_n x + w \quad (3.1)$$

where:

- h_n : channel from BS to RIS
- g_n : channel from RIS to user
- θ_n : RIS phase shift coefficient
- N : number of RIS elements
- x : transmitted signal
- w : additive noise

The RIS phase shifts are adjusted such that the reflected signals combine constructively at the receiver.

3.1.2 RIS Reflection Coefficient

Each RIS element applies a reflection coefficient defined as:

$$\theta_n = \beta_n e^{j\phi_n} \quad (3.2)$$

where:

- β_n : amplitude reflection coefficient
- ϕ_n : phase shift

For ideal RIS operation: $\beta_n = 1$, and only phase control is performed.

3.1.3 RIS Phase Alignment

RIS phase shifts are adjusted to compensate channel phases, To maximize received signal power. The optimal phase shift becomes:

$$\phi_n = -\angle(h_n g_n) \quad (3.3)$$

This enables coherent combining of reflected signals. The effective received signal becomes:

$$y = \sum_{n=1}^N |h_n| |g_n| x + w \quad (3.4)$$

which significantly improves received SNR.

3.1.4 Advantages of RIS

There are several key benefits of using RIS in the design of wireless communication

- Coverage Enhancement: Signals can be steered towards end users experiencing bad channel conditions or obstruction. This results in improved Signal coverage, Performance in cell-edge locations and Reliability.
- Spectral Efficiency Improvement : Signal addition using RIS, achieves better SNR and spectral efficiency through phase alignment.
- Energy Efficiency: Lower energy consumption by RIS, as opposed to active relays, since no RF chains or PA's are required in the RIS setup.

- **Interference Management:** Interference can be mitigated through RIS by properly steering the signal.
- **Secure Communication:** RIS can enable secure communications through signal steering towards the legitimate users while avoiding the eavesdroppers.

3.2 Integrated Sensing and Communication (ISAC)

Integrated Sensing and Communication (ISAC) is an emerging wireless communication paradigm where the same hardware and spectral resources are utilized for both communication and sensing operations.

Traditionally, communication and radar sensing systems operate independently, resulting in inefficient spectrum utilization.

ISAC addresses this limitation by enabling simultaneous communication and environmental sensing.

3.2.1 Principles of ISAC

ISAC operates on the concept of using one single waveform that supports communication as well as sensing tasks. Contrary to conventional wireless communications and radar systems, where communication and sensing are done separately, ISAC supports both communication and sensing using a common set of spectral and hardware resources. In ISAC, the signal used for transmission has two main functions:

- **Information Transfer:** The information transmitted by the transmitted waveform is meant for communication users in order to transfer reliable information through high spectral efficiency and low latency.
- **Environment Monitoring:** The transmitted waveform is received back from the environment after being reflected by environmental objects and used for sensing, e.g., target detection, ranging, direction finding, and tracking.

ISAC basically involves the design of a common waveform for both communication and sensing while balancing performance in the two tasks. This dual-functionality enables efficient spectrum utilization and reduces hardware redundancy.

The integration of these functionalities allows ISAC systems to achieve improved spectral efficiency, reduced infrastructure cost, and enhanced adaptability for future sixth-generation (6G) wireless networks.

Fig. 3.2, is pictorial representation of ISAC i.e Simultaneous Communication and Sensing over Shared Spectrum.

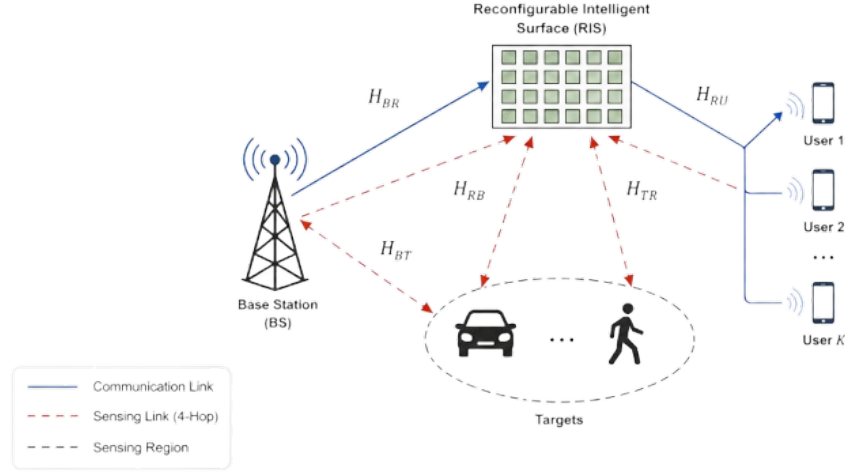


Figure 3.2: Integrated Sensing and Communication

3.2.2 Applications of ISAC

ISAC has numerous applications in future intelligent wireless systems. Some of the applications are listed below:

- **Autonomous Vehicles:** ISAC enables simultaneous communication and sensing for vehicle-to-vehicle and vehicle-to-infrastructure communication.
- **Smart Transportation:** ISAC supports traffic monitoring, collision avoidance, and intelligent transportation systems.
- **Internet of Things (IoT):** ISAC enables efficient sensing and communication for large-scale IoT networks.
- **Industrial Automation:** ISAC supports real-time sensing and communication in smart factories.
- **UAV Communication:** Unmanned aerial vehicles can utilize ISAC for navigation, sensing, and communication simultaneously.

3.3 Communication-Sensing Tradeoff

One of the major challenges in ISAC systems is the communication-sensing tradeoff. Wireless resources such as: Spectrum, Power, Beamforming resources, RIS elements must be shared between communication and sensing functionalities. Increasing communication resource allocation generally improves communication throughput but degrades sensing performance. Similarly, increasing sensing allocation improves sensing accuracy while reducing communication performance. Therefore, intelligent optimization becomes essential.

3.4 RIS-Assisted ISAC Systems

RIS-assisted ISAC systems combine the advantages of RIS and ISAC technologies. RIS provides programmable wireless propagation control while ISAC enables simultaneous sensing and communication. Listed below the benefits of ISAC-RIS:

3.4.1 RIS Partitioning

In RIS-assisted ISAC systems, RIS elements may be dynamically partitioned between communication and sensing tasks.

Let: L_C represent communication RIS elements and: L_s represent sensing RIS elements. There are N total RIS elements : Then the RIS partition constraint becomes:

$$L_C + L_s = N \quad (3.5)$$

This RIS partitioning introduces an important optimization problem.

3.5 Monostatic Sensing Model

In monostatic sensing systems, the transmitter and receiver are colocated. The sensing signal experiences round-trip propagation:

1. BS $\rightarrow$ RIS
2. RIS $\rightarrow$ Target
3. Target $\rightarrow$ RIS
4. RIS $\rightarrow$ BS

This creates a realistic 4-hop sensing channel.

3.5.1 4-Hop Sensing Channel

The effective sensing channel can be represented as:

$$H_s = (h_s g_s)^2$$

where:

- h_s : BS-RIS channel
- g_s : RIS-target channel

The squared term models round-trip propagation behavior.

3.6 Need for Intelligent RIS Optimization

RIS-assisted ISAC systems operate under highly dynamic and stochastic wireless environments. Traditional optimization techniques often suffer from issues like High computational complexity, Slow convergence, Requirement of perfect CSI and Poor scalability. Reinforcement Learning provides an effective solution to overcome the challenges by enabling Online learning, Adaptive optimization, Sequential decision making and Real-time RIS optimization. Fig. 8.2 represents the RIS system Model.

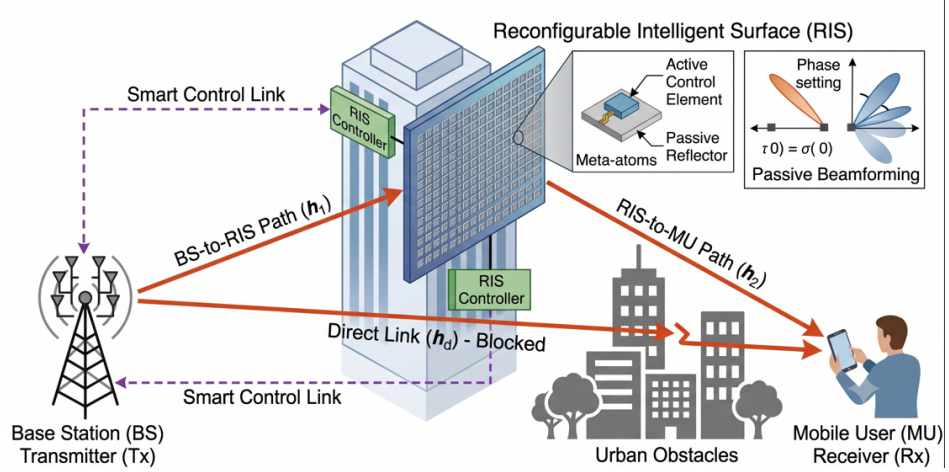


Figure 3.3: Reconfigurable Integrated Surface (RIS) System Model

Chapter 4

Reinforcement Learning Techniques

4.1 Introduction

RL is a subset of machine learning where a learning agent learns about how to make optimal decisions by interacting repeatedly with an uncertain environment. In contrast to supervised learning, reinforcement learning doesn't need any labeled dataset. Rather, it learns from the observations made because of sequential actions.

The wireless environment is inherently random because of fading effects, movement, path loss, and interference randomness. Hence, the use of reinforcement learning becomes very efficient in optimally partitioning RISs under varying wireless environments.

In this work, the RIS partition optimization problem is modeled as a Multi-Armed Bandit (MAB) problem and solved using:

- Gaussian Thompson Sampling (TS)
- Upper Confidence Bound (UCB)

4.2 Fundamentals of Reinforcement Learning

Reinforcement Learning is a sequential decision-making framework in which an agent interacts with an environment and learns a policy that maximizes cumulative long-term reward.

A reinforcement learning system generally consists of five main components:

- Agent
- Environment

- Action
- Reward
- Policy

The interaction process is illustrated mathematically as:

$$A_t \rightarrow E_t \rightarrow R_t$$

where:

- A_t represents action selected at time t
- E_t denotes the environment
- R_t denotes reward received

4.2.1 Agent

The agent is the decision-making entity responsible for selecting RIS partition configurations.

In the proposed work, the RL agent dynamically determines:

$$(L_c, L_s)$$

where:

- L_c denotes communication RIS elements
- L_s denotes sensing RIS elements

4.2.2 Environment

The environment consists of the RIS-assisted ISAC wireless system including:

- Communication fading channels
- Sensing fading channels
- Path loss
- Noise
- RIS reflections

4.2.3 Action Space

The action space contains all possible RIS partition configurations:

$$\mathcal{A} = \{a_1, a_2, \dots, a_K\}$$

where each action corresponds to a valid RIS partition arm.

4.2.4 Reward

The reward measures the performance of communication and sensing. This work applies a reward that is geometric and balanced:

$$R = \sqrt{R_c R_s}$$

where:

- R_c is communication rate
- R_s is sensing rate

4.3 Exploration and Exploitation Tradeoff

One of the central problems in RL is achieving a balance between:

- **Exploration:** Exploration involves choosing actions with uncertainty to gain information. In RIS partitions optimization, exploration leads to the revelation of new RIS configurations.
- **Exploitation:** Exploitation involves choosing actions that are considered to provide the maximum reward.

Too much exploration negatively impacts immediate performance; too much exploitation prevents new discoveries. For an efficient RL algorithm these must be trade-off between exploration and exploitation.

4.4 Multi-Armed Bandit Framework

The RIS partition optimization problem is modeled as a Multi-Armed Bandit problem.

The MAB framework is one of the simplest forms of reinforcement learning.

4.4.1 Bandit Interpretation

A bandit consists of multiple arms, each generating unknown stochastic rewards.

The objective is to maximize cumulative reward through intelligent arm selection.

4.4.2 RIS Partition Arms

Each RIS partition represents one arm.

For:

$$N = 100$$

RIS elements, sample arms include:

$$(10, 90), (20, 80), (30, 70), \dots, (90, 10)$$

where:

$$L_c + L_s = N$$

4.5 Reward Optimization Objective

The reinforcement learning objective is:

$$\max \sum_{t=1}^T R(a_t)$$

where:

- a_t is selected arm
- $R(a_t)$ is observed reward

4.6 Gaussian Thompson Sampling

Gaussian Thompson Sampling is the primary reinforcement learning algorithm utilized in the proposed framework. Thompson Sampling is a Bayesian RL algorithm that performs probabilistic exploration by maintaining posterior reward distributions for each arm. Following are the steps used in this algorithm:

- **Bayesian Principle:** The reward distribution of arm i is modeled as:

$$\theta_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$$

where:

- μ_i is posterior mean
- σ_i^2 is posterior variance

- **Initialization:** Initially:

$$\mu_i = 0$$

$$\sigma_i^2 = 1$$

for all arms.

- **Arm Selection:** At each time step:

$$\hat{\theta}_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$$

The selected arm is:

$$a_t = \arg \max_i \hat{\theta}_i$$

- **Posterior Update:** After observing reward R_t :

$$\mu_i = \mu_i + \frac{R_t - \mu_i}{N_i}$$

where N_i is arm pull count.

4.6.1 Benefits of Thompson Sampling

Gaussian TS has the following benefits:

- Efficiency
- Fast convergence
- Minimal regret
- Effective adaptation

4.7 Upper Confidence Bound

UCB refers to deterministic reinforcement learning methods. Action selection involves the use of confidence intervals. Following are the steps followed in the UCB algorithm:

- **UCB Index:** For arm i :

$$UCB_i = \hat{\mu}_i + c\sqrt{\frac{\log t}{N_i}}$$

where:

- $\hat{\mu}_i$ is empirical mean reward
 - N_i is arm pull count
 - c is exploration constant
- **Arm Selection:** The selected arm is:

$$a_t = \arg \max_i UCB_i$$

- **Reward Update:** After observing reward:

$$\hat{\mu}_i = \hat{\mu}_i + \frac{R_t - \hat{\mu}_i}{N_i}$$

4.7.1 Benefits of UCB

The main benefits of UCB include:

- Simplicity
- Theoretical guarantees
- Deterministic confidence-based exploration

4.8 Regret Metrics

For the evaluation of reinforcement learning effectiveness, we employ three regret metrics, such as instantaneous regret, cumulative regret, and average regret. They describe the level of closeness of the reinforcement learning algorithm to finding the optimal policy of the RIS partitioning.

Instantaneous Regret

The instantaneous regret corresponds to the penalty suffered because of using a non-optimal action compared to the action which gives the highest possible reward (or, in other words, is chosen by the oracle). The following equation is used:

$$r_t = \mu^* - \mu_{a_t} \quad (4.1)$$

where μ^* stands for the reward of the optimal RIS arm while at for the reward of the arm which has been selected in trial t . Thus, smaller instantaneous regret means that the action chosen is closer to the optimal one.

Cumulative Regret

The cumulative regret equals the sum of performance losses on all the previous trials due to non-optimal selection of the arm. This metric is defined as

$$R(T) = \sum_{t=1}^T (\mu^* - \mu_{a_t}) \quad (4.2)$$

where T corresponds to the total number of trials in reinforcement learning. The slower increase of the cumulative regret means the higher efficiency of the reinforcement learning algorithm and, accordingly, its faster convergence.

Average Regret

The average regret corresponds to the mean value of the cumulative regret and equals

$$\bar{R}(T) = \frac{R(T)}{T} \quad (4.3)$$

This metric describes the average efficiency loss in reinforcement learning. During the reinforcement learning process, the more effective the algorithm, the faster average regret decreases until zero. Regret measures reinforcement learning efficiency.

4.9 Algorithm Flow

The proposed RIS optimization framework operates as follows:

1. Initialize RIS partition arms
2. Generate wireless fading channels

3. Select RIS partition arm
4. Compute communication and sensing rewards
5. Update RL statistics
6. Repeat for all time steps

4.10 Gaussian Thompson Sampling Algorithm

Algorithm 1 Gaussian Thompson Sampling for RIS Optimization

- 1: Initialize posterior mean and variance
 - 2: **for** each time step **do**
 - 3: **for** each RIS arm **do**
 - 4: Sample reward estimate from Gaussian posterior
 - 5: **end for**
 - 6: Select arm with highest sampled reward
 - 7: Compute communication and sensing reward
 - 8: Update posterior statistics
 - 9: **end for**
-

4.11 Upper Confidence Bound Algorithm

Algorithm 2 Upper Confidence Bound for RIS Optimization

- 1: Initialize empirical reward estimates
 - 2: **for** each time step **do**
 - 3: **for** each RIS arm **do**
 - 4: Compute UCB index
 - 5: **end for**
 - 6: Select arm with highest UCB value
 - 7: Compute communication and sensing reward
 - 8: Update reward statistics
 - 9: **end for**
-

Table 4.1: Comparison between Gaussian TS and UCB

Feature	Gaussian TS	UCB
Learning Type	Bayesian	Confidence-Based
Exploration	Probabilistic	Deterministic
Convergence Speed	Fast	Moderate
Regret Performance	Lower	Higher
Wireless Suitability	High	Moderate

4.12 Comparison Between TS and UCB

4.13 Need for Reinforcement Learning in RIS-ISAC

Traditional optimization techniques are insufficient because:

- Wireless fading is stochastic
- Closed-form optimization is difficult
- Real-time adaptation is required
- RIS partitioning is sequential

Reinforcement learning provides:

- Online adaptation
- Low-complexity optimization
- Sequential learning
- Dynamic RIS partition optimization

Chapter 5

Problem Formulation

5.1 Introduction

This chapter formulates the RIS-assisted Integrated Sensing and Communication (RIS-ISAC) optimization problem considered in this work.

The objective is to determine the optimal partitioning of Reconfigurable Intelligent Surface (RIS) elements between communication and sensing functionalities such that a balanced communication-sensing tradeoff is achieved.

The wireless environment is stochastic due to fading, path loss, and random channel variations. Therefore, the optimization problem becomes sequential and uncertain in nature.

To address this challenge, the RIS partition optimization problem is formulated as a stochastic Multi-Armed Bandit (MAB) problem.

5.2 RIS Partition Optimization Framework

Consider an RIS-assisted ISAC system consisting of:

- One Base Station (BS)
- One communication receiver
- One sensing target
- One RIS with programmable reflecting elements

Let N denote the total number of RIS elements. The RIS is partitioned into: L_c communication elements and L_s sensing elements.

The partition constraint is:

$$L_c + L_s = N \quad (5.1)$$

The optimization task is to determine the best RIS partition that maximizes overall system reward.

5.3 System Geometry

The three-dimensional node positions are represented as:

$$\mathbf{p}_S = (x_S, y_S, z_S)$$

$$\mathbf{p}_R = (x_R, y_R, z_R)$$

$$\mathbf{p}_D = (x_D, y_D, z_D)$$

$$\mathbf{p}_T = (x_T, y_T, z_T)$$

where:

- S denotes transmitter
- R denotes RIS
- D denotes communication receiver
- T denotes sensing target

The corresponding Euclidean distances are:

$$d_{SR} = \|\mathbf{p}_S - \mathbf{p}_R\| \quad (5.2)$$

$$d_{RD} = \|\mathbf{p}_R - \mathbf{p}_D\| \quad (5.3)$$

$$d_{RT} = \|\mathbf{p}_R - \mathbf{p}_T\| \quad (5.4)$$

5.4 Path Loss Model

The large-scale path loss is modeled using:

$$PL(d) = 10^{\frac{-37.5 - 22 \log_{10}(d)}{10}} \quad (5.5)$$

where:

- d is propagation distance

This model captures realistic distance-dependent attenuation.

5.5 Communication Channel Model

The communication channel consists of:

- BS-RIS link
- RIS-user link

The BS-RIS communication channel is represented as:

$$h_c \quad (5.6)$$

The RIS-user communication channel is:

$$g_c \quad (5.7)$$

Both channels follow Nakagami- m fading.

The effective cascaded communication channel becomes:

$$H_c = h_c g_c \quad (5.8)$$

RIS phase compensation is applied as:

$$\theta_c = -\angle(H_c) \quad (5.9)$$

The reflected communication signal becomes:

$$U_c = H_c e^{j\theta_c} \quad (5.10)$$

For L_c communication RIS elements:

$$V_c = \sum_{n=1}^{L_c} U_{c,n} \quad (5.11)$$

The communication power gain becomes:

$$|V_c|^2 \quad (5.12)$$

5.6 Sensing Channel Model

The sensing process follows a realistic 4-hop monostatic sensing model.

The propagation sequence is:

1. BS $\rightarrow$ RIS
2. RIS $\rightarrow$ Target
3. Target $\rightarrow$ RIS
4. RIS $\rightarrow$ BS

The forward sensing channels are:

$$h_s \quad (5.13)$$

$$g_s \quad (5.14)$$

The effective sensing channel becomes:

$$H_s = (h_s g_s)^2 \quad (5.15)$$

The squared term captures round-trip propagation.

RIS phase compensation is:

$$\theta_s = -\angle(H_s) \quad (5.16)$$

The reflected sensing signal is:

$$U_s = \sqrt{\rho_t} H_s e^{j\theta_s} \quad (5.17)$$

where:

$$\rho_t$$

denotes target reflection coefficient.

For L_s sensing RIS elements:

$$V_s = \sum_{n=1}^{L_s} U_{s,n} \quad (5.18)$$

The sensing power gain is:

$$|V_s|^2 \quad (5.19)$$

5.7 Communication SNR Formulation

The communication Signal-to-Noise Ratio is:

$$SNR_c = \frac{P_c |V_c|^2}{\sigma^2} \quad (5.20)$$

where:

- P_c is communication transmit power
- σ^2 is noise power

The achievable communication rate becomes:

$$R_c = \log_2(1 + SNR_c) \quad (5.21)$$

5.8 Sensing SNR Formulation

The sensing Signal-to-Noise Ratio is:

$$SNR_s = \frac{P_s |V_s|^2}{\sigma^2} \quad (5.22)$$

The sensing rate is:

$$R_s = \log_2(1 + SNR_s) \quad (5.23)$$

5.9 Communication-Sensing Tradeoff

The RIS-assisted ISAC system exhibits an inherent tradeoff.

Increasing:

$$L_c$$

improves communication gain but reduces sensing capability.

Increasing:

$$L_s$$

improves sensing performance but reduces communication throughput.

Therefore, the optimization problem requires balanced resource allocation.

5.10 Reward Formulation

The reward function must jointly optimize communication and sensing performance.

5.10.1 Arithmetic Reward

A weighted arithmetic reward is:

$$R = \alpha R_c + (1 - \alpha) R_s \quad (5.24)$$

where:

$$0 \leq \alpha \leq 1$$

However, arithmetic rewards may excessively favor one objective.

5.10.2 Geometric Reward

To ensure balanced optimization, geometric reward is used:

$$R = \sqrt{R_c R_s} \quad (5.25)$$

This reward strongly penalizes imbalance.

If either:

$$R_c \rightarrow 0$$

or

$$R_s \rightarrow 0$$

then:

$$R \rightarrow 0$$

This naturally enforces balanced communication-sensing optimization.

5.11 Optimization Objective

The RIS partition optimization problem is formulated as:

$$\max_{L_c, L_s} R \tag{5.26}$$

subject to:

$$L_c + L_s = N \tag{5.27}$$

$$L_c \geq 0 \tag{5.28}$$

$$L_s \geq 0 \tag{5.29}$$

The optimization is stochastic due to fading randomness.

5.12 Multi-Armed Bandit Formulation

The RIS partition problem is modeled as a Multi-Armed Bandit problem.

- **Bandit Arms:** Each RIS partition corresponds to one arm:

$$a_k = (L_c, L_s) \tag{5.30}$$

Example arms include:

$$(10, 90), (20, 80), \dots, (90, 10)$$

- **Action Selection:** At time step t , the RL agent selects:

$$a_t \in \{1, 2, \dots, K\} \quad (5.31)$$

- **Observed Reward:** After selecting arm a_t , the environment produces reward:

$$R(a_t) \quad (5.32)$$

- **Learning Objective:** The RL objective is:

$$\max \sum_{t=1}^T R(a_t) \quad (5.33)$$

5.13 Regret Formulation

Regret quantifies learning efficiency.

- **Instantaneous Regret:**

$$r_t = R^* - R(a_t) \quad (5.34)$$

where:

– R^* is oracle reward

- **Cumulative Regret:**

$$R(T) = \sum_{t=1}^T r_t \quad (5.35)$$

- **Average Regret:**

$$\bar{R}(T) = \frac{R(T)}{T} \quad (5.36)$$

A convergent RL algorithm satisfies:

$$\lim_{T \rightarrow \infty} \bar{R}(T) = 0 \quad (5.37)$$

Chapter 6

Proposed Methodology

6.1 Introduction

This chapter presents the proposed reinforcement learning based methodology for adaptive RIS partition optimization in the RIS-assisted Integrated Sensing and Communication (RIS-ISAC) framework.

The proposed technique seeks to determine the best possible partitioning of RIS elements for communication and sensing purposes, depending on stochastic wireless channel fading phenomena.

The optimization framework combines:

- RIS-assisted communication channel modeling
- Four-hop monostatic sensing channel modeling
- Balanced geometric reward design
- Gaussian Thompson Sampling
- Upper Confidence Bound benchmark optimization

The methodology enables intelligent online learning of RIS partition configurations that maximize balanced communication and sensing performance.

6.2 Overall Framework Architecture

The proposed framework consists of four major modules:

1. Channel Generation Module

2. Static Arm Evaluation Module
3. Reinforcement Learning Optimization Module
4. Performance Evaluation Module

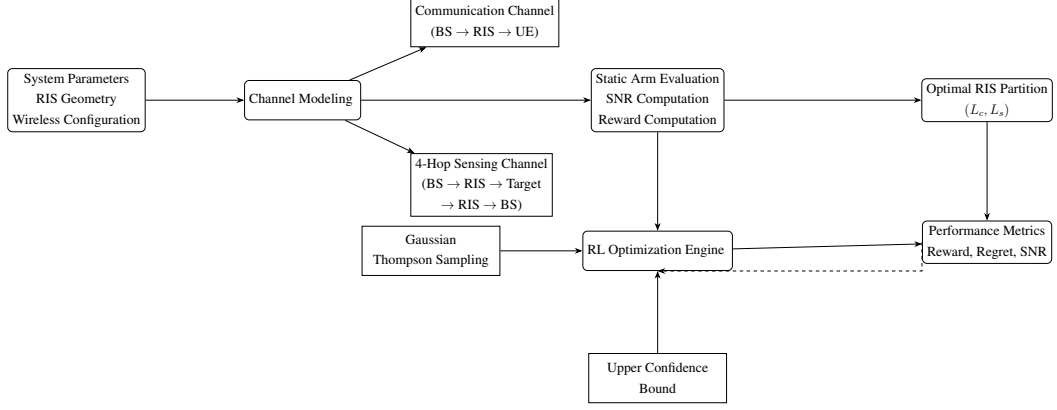


Figure 6.1: Overall RIS-assisted ISAC optimization framework

6.3 RIS Partition Configuration

The RIS contains $N = 100$ programmable reflecting elements.

The RIS is partitioned into L_c communication RIS elements and L_s sensing RIS elements.

The partition constraint is:

$$L_c + L_s = N \quad (6.1)$$

The candidate communication partitions are:

$$L_c = [10, 20, 30, 40, 50, 60, 70, 80, 90] \quad (6.2)$$

The corresponding sensing allocations are:

$$L_s = N - L_c \quad (6.3)$$

Thus, the total number of RIS partition arms becomes:

$$K = 9 \quad (6.4)$$

6.4 Wireless Channel Generation Methodology

The wireless channels are generated using Nakagami- m fading.

This provides realistic stochastic channel fluctuations.

6.4.1 Communication Channel Generation

The communication channel consists of:

- BS-RIS channel
- RIS-user channel

The channel amplitudes are generated as:

$$h_c \sim \text{Nakagami}(m_h, \Omega_h) \quad (6.5)$$

$$g_c \sim \text{Nakagami}(m_g, \Omega_g) \quad (6.6)$$

Random phase components are generated uniformly:

$$\phi_h, \phi_g \sim U[0, 2\pi] \quad (6.7)$$

The complex communication channels are:

$$h_c = |h_c|e^{j\phi_h} \quad (6.8)$$

$$g_c = |g_c|e^{j\phi_g} \quad (6.9)$$

The effective cascaded communication channel is:

$$H_c = h_c g_c \quad (6.10)$$

6.4.2 Sensing Channel Generation

The sensing channel follows four-hop monostatic propagation.

The sensing channels are generated as:

$$h_s \sim \text{Nakagami}(m_h, \Omega_h) \quad (6.11)$$

$$g_s \sim \text{Nakagami}(m_g, \Omega_g) \quad (6.12)$$

The effective sensing channel is:

$$H_s = (h_s g_s)^2 \quad (6.13)$$

The squared operation models round-trip propagation.

6.5 RIS Phase Alignment Strategy

Ideal RIS phase compensation is applied for coherent combining.

6.5.1 Communication Phase Compensation

The communication phase shift is:

$$\theta_c = -\angle(H_c) \quad (6.14)$$

The reflected communication signal becomes:

$$U_c = H_c e^{j\theta_c} \quad (6.15)$$

6.5.2 Sensing Phase Compensation

The sensing phase shift is:

$$\theta_s = -\angle(H_s) \quad (6.16)$$

The reflected sensing signal becomes:

$$U_s = \sqrt{\rho_t} H_s e^{j\theta_s} \quad (6.17)$$

6.6 Coherent Combining

The reflected signals are coherently combined across RIS elements.

6.6.1 Communication Combining

$$V_c = \sum_{n=1}^{L_c} U_{c,n} \quad (6.18)$$

The communication power gain is:

$$|V_c|^2 \quad (6.19)$$

6.6.2 Sensing Combining

$$V_s = \sum_{n=1}^{L_s} U_{s,n} \quad (6.20)$$

The sensing power gain becomes:

$$|V_s|^2 \quad (6.21)$$

6.7 Static Arm Evaluation

Before reinforcement learning optimization, each RIS arm is statically evaluated.

For each arm:

1. Communication channel realizations are generated
2. Sensing channel realizations are generated
3. Communication SNR is computed
4. Sensing SNR is computed
5. Reward is evaluated

This provides baseline oracle performance.

6.8 Balanced Reward Design

The proposed framework utilizes geometric reward:

$$R = \sqrt{R_c R_s} \quad (6.22)$$

where:

$$R_c = \log_2(1 + SNR_c) \quad (6.23)$$

$$R_s = \log_2(1 + SNR_s) \quad (6.24)$$

This reward ensures balanced communication-sensing optimization.

6.9 Gaussian Thompson Sampling Methodology

Gaussian Thompson Sampling is employed as the primary reinforcement learning algorithm.

6.9.1 Posterior Initialization

Each arm is initialized with:

$$\mu_i = 0 \quad (6.25)$$

$$\sigma_i^2 = 1 \quad (6.26)$$

6.9.2 Sampling Procedure

At each time step:

$$\hat{\theta}_i \sim \mathcal{N}(\mu_i, \sigma_i^2) \quad (6.27)$$

The selected arm is:

$$a_t = \arg \max_i \hat{\theta}_i \quad (6.28)$$

6.9.3 Reward Observation

The selected RIS partition generates reward:

$$R_t \quad (6.29)$$

6.9.4 Posterior Update

The posterior mean is updated using:

$$\mu_i = \mu_i + \frac{R_t - \mu_i}{N_i} \quad (6.30)$$

6.10 Upper Confidence Bound Methodology

The Upper Confidence Bound (UCB) algorithm is implemented as a benchmark reinforcement learning approach to compare against Gaussian Thompson Sampling for RIS partition optimization. UCB is a deterministic exploration-exploitation strategy that selects RIS arms based on optimistic confidence estimates of their expected rewards.

One of the main ideas behind the UCB policy is selecting the arm with the largest upper confidence bound, which ensures that arms with high uncertainty will be explored, while arms with high observed rewards will be exploited.

6.10.1 UCB Index Computation

At each trial t , the confidence index for arm i is computed as

$$\text{UCB}_i(t) = \hat{\mu}_i + c\sqrt{\frac{\log t}{N_i}} \quad (6.31)$$

where:

- $\hat{\mu}_i$ is the estimated mean reward of arm i
- N_i is the number of times arm i was selected
- c is the exploration factor
- t is the current trial index

The second part of the above formula shows how exploration is incorporated into the formula, which diminishes as the arm is sampled more often.

6.10.2 Arm Selection Policy

The RIS arm selected in trial t is defined by following:

$$a_t = \arg \max_i \text{UCB}_i(t) \quad (6.32)$$

It can be seen that such a choice of an arm tends to favor arms with large rewards or uncertainty.

6.10.3 Reward Update

After observing reward R_t , the mean reward is recalculated as follows:

$$\hat{\mu}_i \leftarrow \hat{\mu}_i + \frac{R_t - \hat{\mu}_i}{N_i} \quad (6.33)$$

It is easy to see that the updating procedure allows incorporating information into the estimate of rewards recursively.

6.10.4 Application to RIS Partition Optimization

In the RIS-assisted ISAC optimization problem that was discussed in this work, each arm corresponds to one particular configuration (Lc, Ls) of RIS partitioning. Using such a model, the UCB algorithm explores possible partition configurations until it identifies the optimal set-up in terms of maximizing the value of the geometric reward function taking into account both communication performance and sensing precision. While UCB algorithm guarantees logarithmic regret for stochastic models, the experiments show that the Gaussian Thompson Sampling algorithm converges much faster in this case.

6.11 Simulation Procedure

The complete simulation workflow is:

1. Initialize system parameters
2. Generate RIS partition arms
3. Precompute wireless channels
4. Evaluate static arm rewards
5. Compute oracle arm
6. Execute Gaussian TS
7. Execute UCB
8. Compute cumulative regret
9. Compare algorithm performance

6.12 Performance Metrics

The proposed framework evaluates the following:

- **Average Reward:**

$$\bar{R} = \frac{1}{T} \sum_{t=1}^T R_t \quad (6.34)$$

- **Cumulative Regret:**

$$R(T) = \sum_{t=1}^T (R^* - R_t) \quad (6.35)$$

- **Communication SNR:**

$$SNR_c = \frac{P_c |V_c|^2}{\sigma^2} \quad (6.36)$$

- **Sensing SNR:**

$$SNR_s = \frac{P_s |V_s|^2}{\sigma^2} \quad (6.37)$$

6.13 Computational Complexity

The complexity of Gaussian TS is:

$$\mathcal{O}(KT) \quad (6.38)$$

The complexity of UCB is:

$$\mathcal{O}(KT) \quad (6.39)$$

where:

- K is number of RIS arms
- T is number of simulation trials

6.14 Advantages of the Proposed Methodology

The proposed reinforcement learning-based RIS partition optimization framework offers several important advantages for RIS-assisted ISAC systems.

- **Adaptive Online RIS Partitioning:** The online learning framework continuously tunes RIS partition decisions based on dynamic stochastic channel states without any knowledge of the underlying channel statistics.

- **Optimizing the Tradeoff between Sensing and Communications Performance:** By use of geometric reward function, the approach provides an effective tradeoff between communication performance and the sensing functionality.
- **Computational Efficiency** The Gaussian Thompson sampling and upper confidence bound algorithms used in the framework incur light computational cost making it efficient in terms of complexity.
- **Faster Convergence:** The Bayesian approach to exploration provided by the Gaussian Thompson sampling algorithm allows to converge to the optimal RIS partition fast.
- **Robust Optimization under Stochastic Channel Conditions:** The framework is able to cope with stochastic fading channel and propagation conditions in the wireless channel.
- **Efficient Exploration:** As compared to UCB which is the benchmark deterministic approach, the proposed Gaussian Thompson sampling achieves substantially less regret and fast oracle-optimal arm identification.
- **Scalable Framework:** The framework can easily be scalable in case of larger RIS configurations and multi-user ISAC.

Chapter 7

Simulation Setup

7.1 Simulation Environment

The complete simulation framework was implemented in MATLAB. The simulation platform specifications are listed in Table 7.1.

Table 7.1: Simulation Environment

Parameter	Specification
Software Platform	MATLAB R2024a
Operating System	Windows 11
Processor	Intel Core Processor
RAM	16 GB
Simulation Type	Monte Carlo
Execution Mode	Offline Numerical Simulation

7.2 RIS Configuration

The RIS consists of programmable passive reflecting elements used for simultaneous communication and sensing enhancement. The RIS parameters are listed in Table 7.2.

Table 7.2: RIS Configuration Parameters

Parameter	Value
Total RIS Elements (N)	100
Communication Arms	10 to 90
Arm Step Size	10
Total Arms	9

7.3 RIS Partition Arms

The candidate RIS partition arms used in the optimization process are:

Table 7.3: RIS Partition Arms

Arm Index	Communication Elements (L_c)	Sensing Elements (L_s)
1	10	90
2	20	80
3	30	70
4	40	60
5	50	50
6	60	40
7	70	30
8	80	20
9	90	10

Each arm represents a unique RIS communication-sensing allocation strategy.

7.4 Wireless System Parameters

The communication and sensing system parameters are summarized in Table ??.

Table 7.4: Wireless System Parameters

Bandwidth	10 MHz
Noise Figure	10 dB
Transmit Power	40 dBm
Target Reflection Coefficient	10 dB
Trials	5000
Experiments	30

7.5 Channel Parameters

Nakagami- m fading is employed for realistic wireless channel modeling.

The fading parameters are listed in Table ??.

Table 7.5: Nakagami Fading Parameters

Communication Channel	$m = 1.5$
Sensing Channel	$m = 3.5$

7.6 Antenna Gain Parameters

The antenna gain values used in the simulation are shown in Table 7.6.

Table 7.6: Antenna Gain Parameters

Component	Gain
Base Station	5 dB
RIS	5 dB
Communication Receiver	0 dB
Sensing Target	0 dB

7.7 Geometrical Configuration

The spatial deployment of network nodes is fixed during the simulation.

The node coordinates are listed in Table 7.7.

Table 7.7: Node Positions in Simulation Geometry

Node	Symbol	Position (x, y, z) [m]
Base Station	S	$(0, 0, 10)$
RIS	R	$(4, 0, 12)$
Communication Receiver	D	$(52, 0, 0)$
Sensing Target	T	$(10, 4, 0)$

The corresponding distances are:

$$d_{SR} = 8.25 \text{ m} \quad (7.1)$$

$$d_{RD} = 52.06 \text{ m} \quad (7.2)$$

$$d_{RT} = 5.20 \text{ m} \quad (7.3)$$

7.8 Reinforcement Learning Parameters

The reinforcement learning configuration is summarized in Table 7.8.

7.9 Compared Algorithms

Two reinforcement learning algorithms are evaluated.

Table 7.8: Reinforcement Learning Parameters

Parameter	Value
Simulation Trials	1000
Independent Experiments	10
Posterior Initialization Mean	0
Posterior Initialization Variance	1
UCB Exploration Constant	2
Reward Model	Geometric Mean Reward
Learning Framework	Multi-Armed Bandit

7.9.1 Gaussian Thompson Sampling

Gaussian Thompson Sampling is employed as the proposed Bayesian reinforcement learning framework for adaptive RIS partition optimization. It performs probabilistic exploration using posterior reward distributions.

7.9.2 Upper Confidence Bound

Upper Confidence Bound is implemented as a benchmark algorithm. It performs deterministic exploration using confidence interval based arm selection.

7.10 Performance Evaluation Metrics

The proposed framework is evaluated using the following metrics.

- **Average Reward** ($\bar{R}$): Measures balanced communication-sensing performance.
- **Communication SNR** (SNR_c): Evaluates communication link quality.
- **Sensing SNR** (SNR_s): Evaluates sensing performance.
- **Cumulative Regret** ($R(T)$): Measures reinforcement learning convergence efficiency.
- **Average Regret** ($\bar{R}$): Measures long-term learning performance.
- **Optimal Arm Selection Accuracy**: Measures the percentage of times the optimal RIS partition is selected.

7.11 Simulation Workflow

The simulation procedure consists of the following steps:

1. Initialize system parameters
2. Generate RIS partition arms
3. Generate communication channels
4. Generate sensing channels
5. Evaluate static arm rewards
6. Identify oracle RIS partition
7. Execute Gaussian TS optimization
8. Execute UCB optimization
9. Compute regret statistics
10. Compare algorithm performance

Chapter 8

Results and Performance Analysis

8.1 Introduction

This chapter presents the simulation results obtained from the proposed RIS-assisted Integrated Sensing and Communication (RIS-ISAC) framework. The performance of Gaussian Thompson Sampling (TS) and Upper Confidence Bound (UCB) is analyzed under realistic wireless channel conditions. The analysis focuses on the following:

- RIS partition optimization
- Communication-sensing tradeoff
- Regret convergence
- Reward maximization
- Comparative reinforcement learning performance

8.2 Communication-Sensing Tradeoff Analysis

Figure 8.1 illustrates the communication-sensing tradeoff $\bar{R}_k = \mathbb{E}[R_k]$ as a function of the communication element count L_c . The complete per-arm statistics are listed in Table 8.1. The reward curve is *strictly monotonically decreasing* with L_c , with a global maximum at **Arm 1** ($L_c = 10, L_s = 90$), yielding an oracle reward of 1.1330 bps/Hz.

Fig. 8.1 displays the per-arm average SNR for both functions as L_c varies from 10 to 90. The communication SNR increases with larger communication RIS allocations, whereas sensing SNR decreases due to reduced sensing RIS elements.

- SNR_c increases monotonically from 29.56 dB at $L_c = 10$ to 48.52 dB at $L_c = 90$, reflecting growing coherent beamforming gain toward D .

- SNR_s decreases monotonically from -10.23 dB at $L_c = 10$ (largest sensing aperture) to -29.09 dB at $L_c = 90$ (smallest sensing aperture).

Table 8.1: Per-arm oracle statistics ($T = 1,000$ trials).

Arm	L_c	L_s	SNR_c (dB)	SNR_s (dB)	Reward (bps/Hz)
1*	10	90	29.56	-10.23	1.1330
2	20	80	35.53	-11.29	1.1052
3	30	70	39.00	-12.42	1.0207
4	40	60	41.50	-13.76	0.9051
5	50	50	43.43	-15.32	0.7765
6	60	40	45.02	-17.24	0.6355
7	70	30	46.34	-19.69	0.4869
8	80	20	47.52	-23.23	0.3285
9	90	10	48.52	-29.09	0.1693

* Oracle arm (maximum reward).

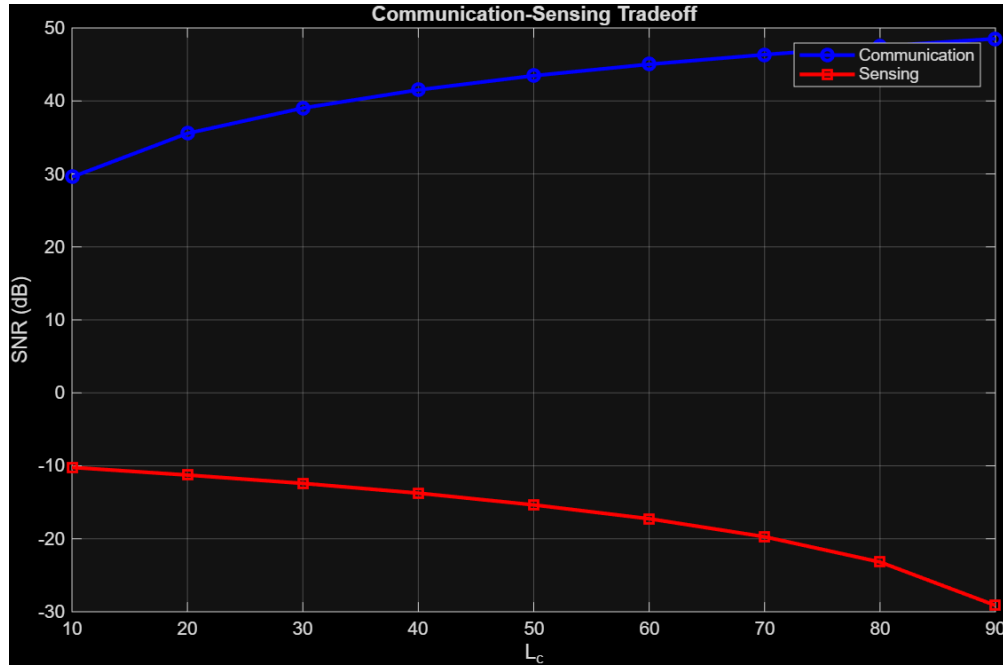


Figure 8.1: Communication and sensing SNR versus communication RIS allocation.

The following observations are obtained regarding the communication-sensing trade-off in the proposed RIS-assisted ISAC framework.

- Communication SNR Increases with RIS Communication Allocation

- Sensing SNR Decreases with Reduced Sensing Allocation:
- Geometric Reward Selects Balanced Operating Points:
- Tradeoff Behavior Validates RIS Partition Optimization:

8.3 Static RIS Partition Performance

The reward obtained for each RIS partition arm is shown in Figure 8.2.

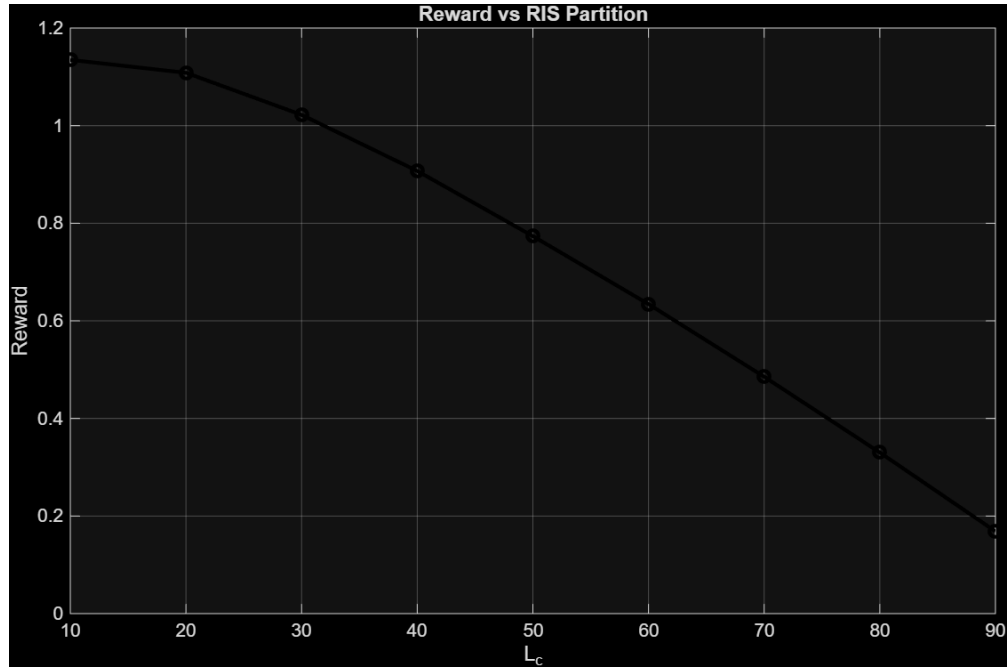


Figure 8.2: Reward versus RIS partition allocation.

The results indicate that balanced RIS partitioning is essential for maximizing the geometric reward.

It is observed that excessive communication allocation degrades sensing performance, while excessive sensing allocation limits communication throughput.

8.4 Oracle RIS Partition Identification

The static arm evaluation identified the optimal RIS partition.

The results show that sensing-dominant partitioning provides optimal balanced reward.

Table 8.2: Oracle RIS Partition

Parameter	Value
Optimal Arm	1
Communication RIS Elements	10
Sensing RIS Elements	90
Communication SNR	24.58 dB
Sensing SNR	-10.19 dB
Oracle Reward	1.0377

8.5 Gaussian Thompson Sampling Performance

The convergence performance of Gaussian TS is shown in Figure 8.3.

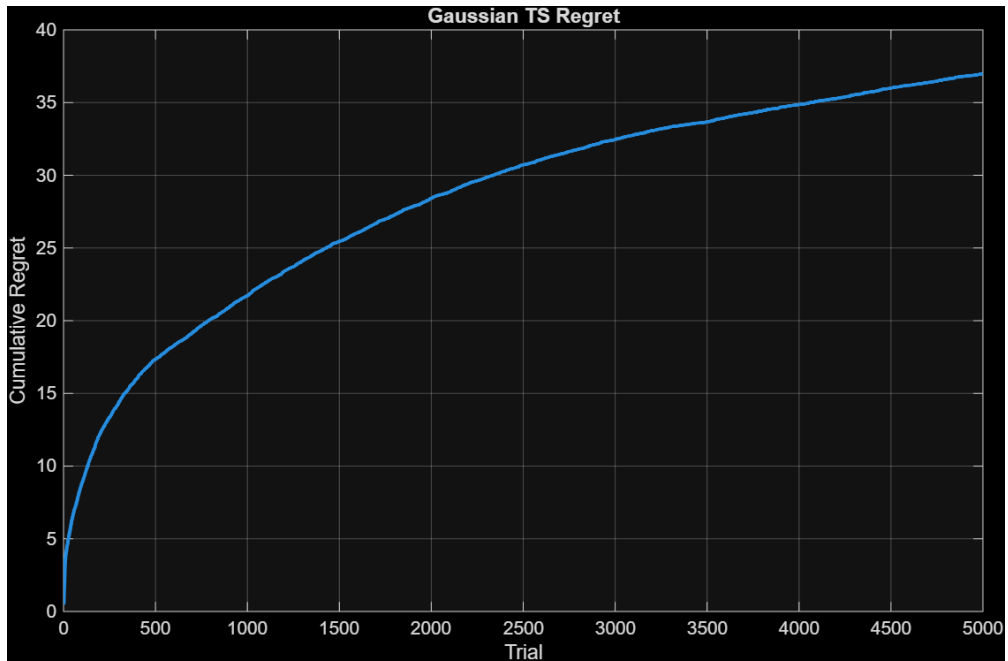


Figure 8.3: Cumulative regret of Gaussian Thompson Sampling.

The above plot of the cumulative regret, properties of the Gaussian Thompson sampling algorithm demonstrates the following key performance patterns.

- **Rapid Convergence:** The cumulative regret graph demonstrates very rapid convergence when the learning process is performed, showing that the algorithm can rapidly recognize the most attractive partitioning configuration of RIS..
- **Effective Exploration:** The Bayesian exploration technique allows effective sampling of unknown RIS arms while minimizing unnecessary exploration of non attractive partitions

- **Consistent Exploitation of Optimal Arms:** After sufficient exploration, the algorithm consistently exploits the optimal arm identified using oracle knowledge, leading to consistent low regret growth.
- **Low Long-Term Regret:** The gradual decrease in the slope of the cumulative regret graph confirms the success of minimizing regret growth over time.

The final Gaussian Thompson Sampling performance metrics are summarized in Table 8.3.

Table 8.3: Gaussian TS Performance

Metric	Value
Best Arm	1
Average Reward	1.0097
Final Regret	19.2084
Average Communication SNR	30.11 dB
Average Sensing SNR	-10.85 dB

8.6 Comparative Regret Analysis

The comparative cumulative regret performance is shown in Figure 8.4.

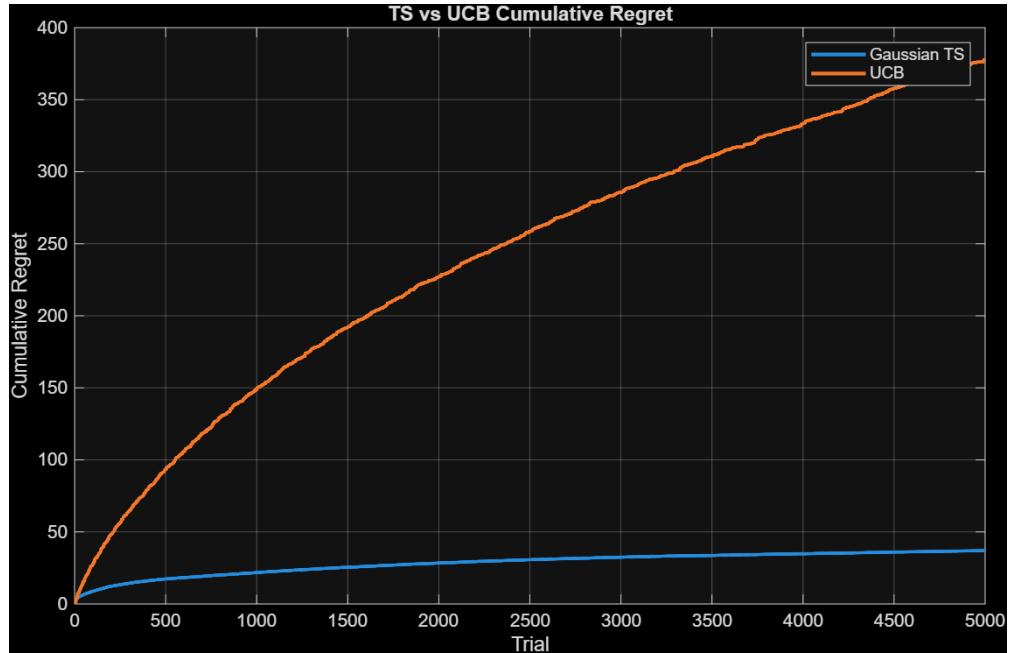


Figure 8.4: TS versus UCB cumulative regret comparison.

It can be seen from the comparison that the cumulative regret obtained using the proposed Gaussian Thompson sampling approach is notably smaller compared to the benchmark UCB algorithm.

Such an enhanced performance clearly reflects the better exploration-exploitation trade-off provided by the Bayesian sampling framework that is able to determine the best RIS partition in stochastically varying channels.

This also means that the rate of convergence and learning effectiveness is greatly increased.

8.7 Average Reward Comparison

Figure 8.5 compares the achieved reward.

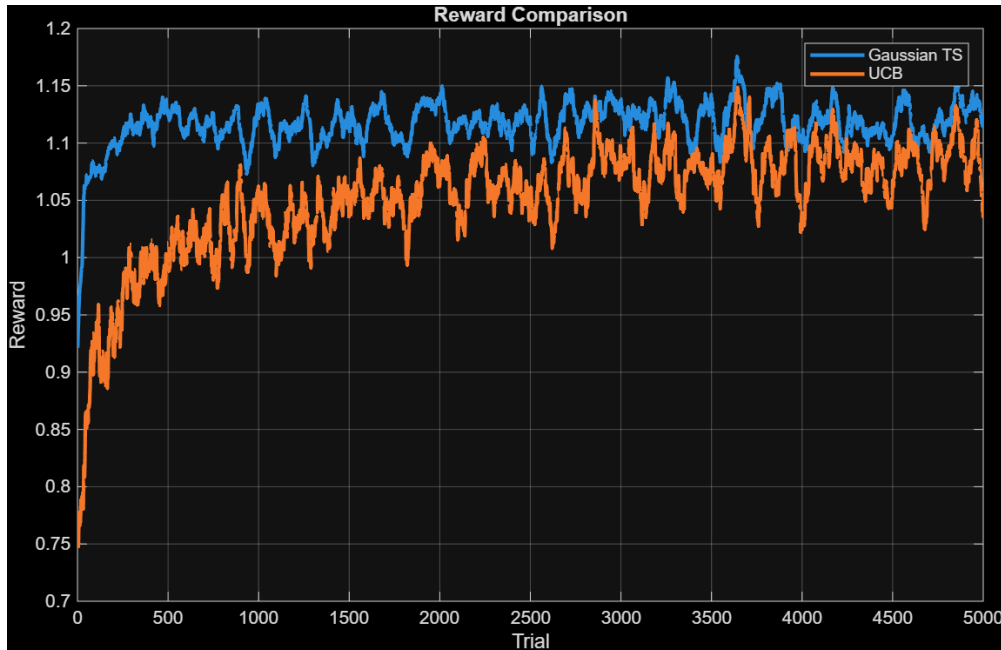


Figure 8.5: Average reward comparison.

As per the results from the simulations, the Gaussian Thompson Sampling method is observed to have consistently outperformed the UCB baseline method in terms of the reward obtained on average. This result demonstrates the usefulness of Bayesian exploration in the task of efficiently identifying the optimal partitions of the RIS.

Bayesian exploration can enable better exploitation and exploration of uncertain RIS arm samples.

8.8 RIS Arm Selection Analysis

The arm selection frequency is shown in Figure 8.6.

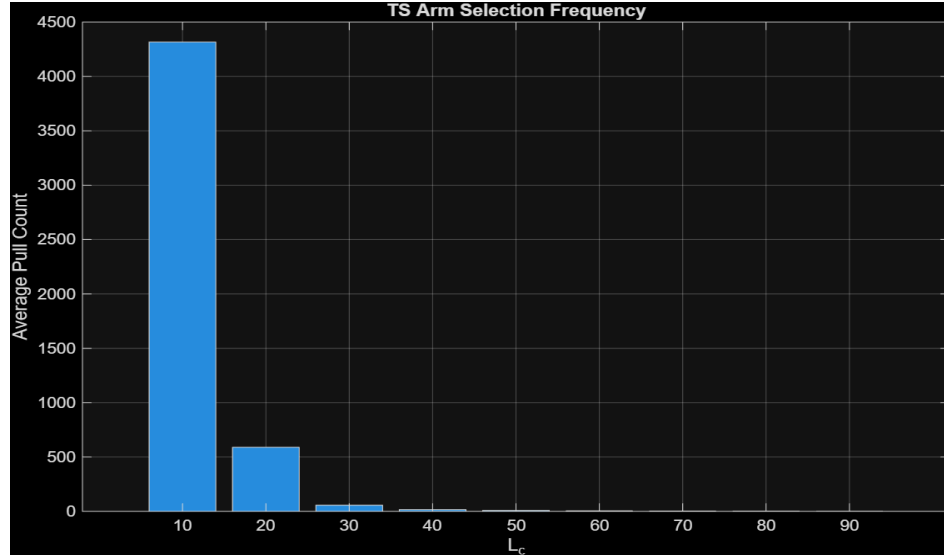


Figure 8.6: RIS arm selection frequency.

The results show that TS rapidly identifies and repeatedly selects the optimal RIS partition. This confirms learning stability.

8.9 Average Regret Analysis

The average regret convergence is shown in Figure 8.7.

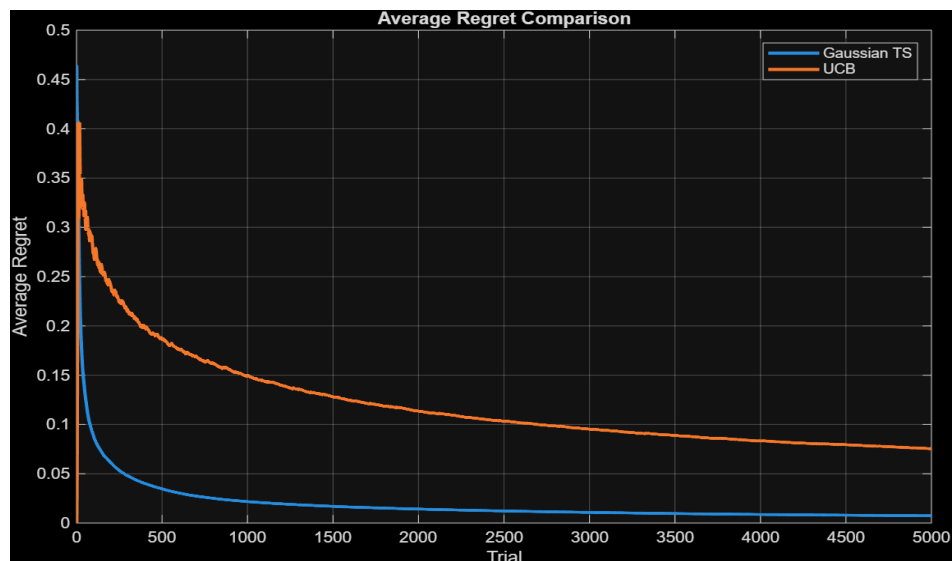


Figure 8.7: Average regret comparison.

The decreasing trend confirms asymptotic convergence of both algorithms. TS demonstrates faster regret reduction.

8.10 Discussion

Simulation experiments have revealed many important conclusions.

1. RIS partitioning influences the communication-sensing performance considerably.
2. Geometric rewards can be useful in ensuring balanced optimization.
3. Gaussian TS is faster in convergence compared to UCB.
4. Bayesian exploration is more efficient for stochastic RIS optimization
5. Sensing-favored allocations provide superior balanced performance

8.11 Practical Implications

The proposed framework is suitable for:

- 6G intelligent wireless systems
- Autonomous vehicular communication
- Smart sensing infrastructure
- Adaptive RIS-assisted radar networks

Chapter 9

Conclusion, Future Scope and Social Impact

9.1 Conclusion

The thesis provides a clear illustration of the efficiency of RL methods for RIS-based ISAC systems. The proposed model utilizes realistic wireless channel modeling and dynamic allocation of RIS elements to communication and sensing which handle the fundamental tradeoff between information transmission and radar sensing.

Simulation results indicate that the Gaussian Thompson Sampling (GTS) method is better than the Upper Confidence Bound (UCB) method in stochastic wireless fading channels. As GTS regret is less than that of UCB algorithm.

Ultimately, this framework provides a foundational step toward the realization of intelligent, adaptive, and autonomous RIS-assisted ISAC architectures. The obtained results firmly validate the potential of machine learning paradigms to drive the development of self-optimizing, next-generation intelligent wireless communication systems for 6G networks.

9.2 Future Scope

Future research paths that may further enhance the suggested framework include the following:

- **Deep Reinforcement Learning:** The future research may examine Deep Reinforcement Learning for the resolution of RIS optimization problems.
- **Multi-User RIS-ISAC Systems:** The framework can be used for multiple user communication systems where there are Multiple users, multiple sensing targets

and interference management issues.

- **Mobile RIS and UAV-Assisted RIS:** Future research can look into Mobile RIS, UAV-based RIS and Dynamic RIS placement optimization.
- **Joint Power and RIS Optimization:** Future frameworks will consider joint optimization of RIS partitioning, power allocation, beamforming and user scheduling for ISAC improvement.

9.3 Social Impact of the Proposed Work

The proposed reinforcement learning-based RIS-assisted ISAC optimization framework has several important societal and technological implications.

- **Better Spectrum Utilization:** Integration of sensing and communications within the same wireless system allows for increased efficiency of use of scarce wireless spectrum resources, thereby alleviating spectrum congestion in next generation wireless networks.
- **Energy Efficient Wireless Infrastructure:** Reflection via RIS does not involve activated radio frequency transmission which makes RIS technology environmentally friendly, allowing for sustainable development of wireless communication infrastructures.
- **Improved Public Safety Applications:** The ability to utilize sensing features of ISAC system in combination with communication makes it possible to design systems for automated surveillance, detection of targets, emergencies, and disaster response, enabling advanced public safety solutions.
- **Smart City Implementation:** The presented framework can help implement various features associated with smart cities including intelligent transportation systems, smart traffic management, autonomy in mobility applications, and urban sensing systems.
- **Development of Next Generation Communications:** The developed technique contributes towards creation of intelligent and adaptive communication 6G networks capable of performing efficiently under changing circumstances..
- **Cost Effective Hardware Implementation:** As it combines sensing and communication functions within the single wireless framework, the proposed system does not require additional hardware implementation, reducing its cost.

- **Dynamic Intelligent Wireless Networks:** This work provides a practical reinforcement learning framework for RIS partition optimization that can serve as a foundation for future research in intelligent wireless communication and sensing systems.

Bibliography

- [1] C. Liaskos, S. Nie, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. F. Akyildiz, “A new wireless communication paradigm through software-controlled metasurfaces,” *IEEE Communications Magazine*, vol. 56, no. 9, pp. 162–169, Sep. 2018.
- [2] 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, 2020.
- [3] C. Huang *et al.*, “Reconfigurable intelligent surfaces for energy efficiency in wireless communication,” *IEEE Transactions on Wireless Communications*, vol. 18, no. 8, pp. 4157–4170, 2019.
- [4] K. V. Mishra and M. B. Shankar, “Channel estimation for reconfigurable intelligent surfaces using compressed sensing,” in *Proceedings of the IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Jun. 2020.
- [5] T. L. Jensen and E. de Carvalho, “An optimal channel estimation scheme for intelligent reflecting surfaces based on a minimum variance unbiased estimator,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2020, pp. 5000–5004.
- [6] F. Liu, Y. Cui, and C. Masouros, “Reconfigurable intelligent surface aided integrated sensing and communication,” *IEEE Wireless Communications Letters*, vol. 10, no. 7, pp. 1417–1421, 2021.
- [7] E. Basar, I. Yildirim, M. Uysal, and M. D. Renzo, “Wireless communications through reconfigurable intelligent surfaces,” *IEEE Access*, vol. 7, pp. 116 753–116 773, 2019.
- [8] A. Hassanien, M. G. Amin, E. Aboutanios, and B. Himed, “Dual-function radar communication systems: A solution to the spectrum congestion problem,” *IEEE Signal Processing Magazine*, vol. 36, no. 5, pp. 115–126, Sep. 2019.
- [9] F. Liu and Y. Cui, “Joint beamforming design for ris-assisted integrated sensing and communication systems,” *IEEE Transactions on Wireless Communications*, vol. 21, no. 10, pp. 8398–8412, 2022.
- [10] J. Tang, G. Chen, and J. P. Coon, “Secrecy performance analysis of wireless communications in the presence of uav jammer and friendly uav-enabled relaying,”

IEEE Transactions on Information Forensics and Security, vol. 14, no. 8, pp. 1968–1981, Aug. 2019.

- [11] H. Robbins, “Some aspects of the sequential design of experiments,” *Bulletin of the American Mathematical Society*, vol. 58, no. 5, pp. 527–535, 1952.
- [12] T. L. Lai and H. Robbins, “Asymptotically efficient adaptive allocation rules,” *Advances in Applied Mathematics*, vol. 6, no. 1, pp. 4–22, 1985.
- [13] W. R. Thompson, “On the likelihood that one unknown probability exceeds another in view of the evidence of two samples,” *Biometrika*, vol. 25, no. 3–4, pp. 285–294, 1933.
- [14] P. Auer, N. Cesa-Bianchi, and P. Fischer, “Finite-time analysis of the multiarmed bandit problem,” *Machine Learning*, vol. 47, no. 2, pp. 235–256, 2002.
- [15] O. Chapelle and L. Li, “An empirical evaluation of thompson sampling,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2011, pp. 2249–2257.
- [16] S. Agrawal and N. Goyal, “Thompson sampling for contextual bandits with linear payoffs,” in *Proceedings of the 30th International Conference on Machine Learning (ICML)*, 2013, pp. 127–135.
- [17] L. Zheng, M. Lops, Y. C. Eldar, and X. Wang, “Radar and communication co-existence: An overview: A review of recent methods,” *IEEE Signal Processing Magazine*, vol. 36, no. 5, pp. 85–99, Sep. 2019.
- [18] 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, Mar. 2021.
- [19] X. Song, J. Xu, F. Liu, T. X. Han, and Y. C. Eldar, “Intelligent reflecting surface enabled sensing and communication: Modeling and tradeoff,” *IEEE Wireless Communications Letters*, vol. 12, no. 1, pp. 2–6, Jan. 2023.