Distributed Learning for Reconfigurable Intelligent Surfaces

Submitted in partial fulfillment of the requirements
of the degree of

Doctor of Philosophy

by

Ishaan Sharma
(Roll No. 2K20/PHDEC/05)

Supervisors:

Dr. Rohit Kumar

(Assistant Professor, Department of ECE, DTU, India)

Dr. Sumit J Darak

(Professor, Department of ECE, IIIT-Delhi, India)

Electronics and Communication Engineering
DELHI TECHNOLOGICAL UNIVERSITY
2025

Acknowledgments

It gives me immense pleasure to express my heartily gratitude to everyone who supported and guided me in the completion of my thesis. Foremost, I would like to express my sincere gratitude to my advisors Dr. Rohit Kumar and Dr. Sumit J Darak. Without their excellent guidance, encouragement and support, I would never be able to finish my thesis work. They have been a great source of inspiration and I feel extremely fortunate to work with them.

I would like to thank Shannon lab and Cloud Labs technical staff at IIIT Delhi, Mr. Khagendra Joshi for providing me quick access to all instruments whenever I needed them.

I would like to acknowledge my parents, friends and Algorithms to Architecture (A2A) Lab mates for encouraging and supporting me. They have been a source of moral support to me and have extended their helping hands without fail.

DELHI TECHNOLOGICAL UNIVERSITY
formerly Delhi College of Engineering
Shahbad Daulatpur, Main Bawana Road,
Delhi–110042

Candidate's Declaration

I, **Ishaan Sharma**, hereby certify that the work which is being presented in the thesis titled **"Distributed Learning in Reconfigurable Intelligent Surfaces"**,in partial fulfillment of the requirements for the awardof the Degree of Doctor ofPhilosophy, submitted in the Department of **Electronics and CommunicationEngineering**, Delhi Technological University is an authentic record of my own work carried out during the period from **August 2020** to **July2025** under thesupervision of **Dr. Rohit Kumar and Dr. Sumit J. Darak**.

The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other Institute.

Candidate's Signature

This is to certify that the student has incorporated all the corrections suggested bythe examiners in the thesis and the statement made by the candidate is correct tothe best of our knowledge.

Signature of Supervisor

Signature of External Examiner

DELHI TECHNOLOGICAL UNIVERSITY
formerly Delhi College of Engineering
Shahbad Daulatpur, Main Bawana Road,
Delhi–110042

Certificate by the Supervisors

Certified that **Ishaan Sharma** (Enrollment No.: 2K20/PHDEC/05) has carried out their research work presented in this thesis titled **"Distributed Learning for Reconfigura--ble Intelligent Surfaces"** for the award of **Doctor of Philosophy** in the Department of Electronics and Communication Engineering,Delhi Technological University, under my supervision. The thesis embodies resultsof original work and studies are carried out by the student himself and thecontents of the thesis do not form the basis for the award of any other degree to the candidate or anybody else from this or any other University/Institution.

Place: Delhi
Date:

Dr. Rohit Kumar
Supervisor
Department of ECE
Delhi Technological University,
Delhi–110042, India

Place: Delhi
Date:

Dr. Sumit J Darak
Co-Supervisor
Department of ECE
Indraprastha Institute of Information and Technology
Delhi–110020 India

Abstract

Reconfigurable Intelligent Surfaces (RIS) have gained significant attention as a transformative technology to enhance wireless communication by intelligently manipulating the propagation environment. By dynamically adjusting the phase shifts of a large array of passive elements, RIS can actively influence signal propagation to improve link quality, spectral efficiency, and interference mitigation. One of the fundamental challenges in RIS-assisted wireless communication is maximizing the Signal-to-Noise Ratio (SNR) at the receiver. Achieving optimal SNR necessitates determining the best RIS phase shift configuration, a task complicated by the high-dimensional search space, dynamic channel conditions, and practical constraints such as discrete phase shifts and limited feedback from the receiver. Existing solutions, including exhaustive search and model-based optimization techniques, suffer from significant computational complexity and are often infeasible for real-time adaptation in practical systems.

To address these challenges, this thesis proposes novel online-learning-based Multi-Armed Bandit (MAB) algorithms tailored for RIS optimization. The RIS configuration problem is formulated as a sequential decision-making process where the system continuously learns the optimal phase shift arrangement while balancing exploration and exploitation. Unlike traditional reinforcement learning approaches, our MAB-based framework provides a lightweight, adaptive, and sample-efficient solution, making it particularly suitable for practical deployments where acquiring full channel state information is costly or impractical. We introduce multiple algorithmic variants that incorporate advanced exploration-exploitation trade-offs, ensuring rapid convergence to near-optimal configurations without excessive computational overhead.

The proposed algorithms are evaluated through extensive simulations under diverse channel conditions and system constraints. The results demonstrate that our approach significantly outperforms traditional heuristic and non-learning-based methods, achieving faster convergence, higher achievable SNR, and improved robustness against environmental variations. Additionally, we analyze the computational efficiency of our methods, demonstrating their suitability for

real-time RIS control in dynamic wireless environments. By leveraging learning-based strategies, our approach enables RIS to autonomously adapt to changing conditions, unlocking its full potential for next-generation wireless networks.

Beyond performance gains, this thesis explores the practical considerations for implementing MAB-based RIS optimization in real-world networks. We address key aspects such as feedback mechanisms and computational complexity, ensuring that our proposed methods align with practical hardware capabilities. Furthermore, we extend our analysis to multi-user scenarios, cooperative RIS control, and integration with emerging wireless technologies, such as millimeter-wave and terahertz communications. These findings highlight the potential of intelligent, adaptive, and scalable RIS-based communication systems that dynamically optimize their behavior based on real-time observations.

This work contributes to the growing body of research on RIS-aided wireless networks by introducing efficient learning-based strategies for optimizing RIS configurations. The findings presented in this thesis pave the way for future studies exploring distributed learning approaches, joint RIS and beamforming optimization, and energy-efficient RIS deployment strategies. Although ISAC and ISAC-RIS are not the core focus of this work, future research could explore the application of MAB frameworks in ISAC-RIS systems, where joint optimization of sensing and communication objectives may benefit from efficient online learning strategies. Our work not only enhances theoretical understanding but also provides a practical foundation for deploying RIS in next-generation wireless networks, including 6G and beyond.

Contents

List of Tables

List of Figures

List of Abbreviations

RIS	Reconfigurable Intelligent Surfaces
MAB	Multi Armed Bandit
B5G	Beyond 5G
SDR	Software Defined Radio
SCA	Successive Convex Approximation
DRL	Deep Reinforcement Learning
mmWave	Millimeter Wave
NLoS	Non-line-of-sight
CSI	Channel State Information
UCB	Upper Confidence Bound
IID	Independently and identically distributed
INID	Independent but not identically distributed
AWGN	Additive White Gaussian Noise
SNR	Signal to Noise Ratio
ERA	Exhaustive RIS-aided
ORA	Opportunistic RIS-aided
FEUCB	Focused Exploration Upper Confidence Bound
TCB	Threshold Confidence Bound
EFEUCB	Enhanced Focused Exploration Upper Confidence Bound
HSCD	Hardware-Software Co-Design
PS	Processing System

PL	Programmable Logic
FPGA	Field Programmable Gate Array
GPU	Graphic Processing Unit
SoC	System on Chip
mEFEUCB	Modified Enhanced Focused Exploration Upper Confidence Bound
SIMD	Single Instruction Multiple Data
WL	Word Length
LUT	Look Up Table
FF	Flip Flops
SPFP	Single Precision Float Point
DPFP	Double Precision Float Point
FP	Fixed Point
TX	Transmitter
RX	Receiver
USS	Unsupervised Sensor Selection
CPAUCB	Consumed Power Aware Upper Confidence Bound
CUSUM	Cumulative Sum
DUCB	Discounted Upper Confidence Bound
MUCB	Monitored Upper Confidence Bound
DDR-EFEUCB	Dynamic Discounting and Restart Enhanced Focused Exploration Upper Confidence Bound
ISAC	Integrated Sensing and Communication

List of Symbols

t	Current time slot
T	Total time horizon
L	Number of Sub-Blocks
Z	Total Number of Elements
R	Total Number of Sub-blocks(Arms), i.e. KL
M	Combination of L Sub-Blocks
K	Total Number of RIS
N	Total Number of Receivers
I_t	Selected Sub-Block
A	Amplitude of sub-block
x_s	Transmit Symbol
P_s	Transmit Power
α_{klnz}	Amplitude Reflection Coefficient of the z^{th} element of the m
θ_{klnz}	phase-shift
ω_n	Additive White Gaussian Noise (AWGN)
$\tilde{G}_{klz}$	Complex Channel Coefficient Between the Transmitter and z^{th} Element of the RIS sub-block r
G_{klz}	Magnitude of Complex Channel Coefficient Between the Transmitter and z^{th} Element of the RIS sub-block r
ψ_{klz}	Phase of Complex Channel Coefficient Between the Transmitter and z^{th} Element of the RIS sub-block r

$\tilde{H}_{klnz}$	Complex Channel Coefficient Between the z^{th} Element of the RIS sub-block r and n^{th} receiver.
H_{klnz}	Magnitude of Complex Channel Coefficient Between the z^{th} Element of the RIS sub-block r and n^{th} receiver
σ_n^2	Variance
1_{cond}	Indicator Function
X^{I_t}	Reward of selecting Sub-block
r_*	Optimal Sub-block which offers highest SNR
S^r	Total number of Selections of sub-block r.
$\hat{X}^r(t)$	Total accumulated reward of sub-block r till time slot t
$A_r(t)$	An event of good RIS block, r, being selected during the UCB phase when $\hat{R}$ blocks are sampled sufficiently
$B_r(t)$	An event of good RIS block, r, being selected during the UCB phase when $\hat{R}$ blocks are not sampled sufficiently
$C_r(t)$	An event of poor RIS block, r, being selected.
SNR_d	Desired Threshold SNR
P_C	Consumed Power
$\hat{K}$	Subset of Good Arms
$\hat{\mu}_k(t)$	Empirical Mean Reward for arm k up to Time t
$UCB^k(t)$	UCB Quality Factor for k arms till time t
Cp_{I_t}	Consumed Power of selected block
ϕ_{I_t}	Error rate of the selected block
$P_{1/0_1}$	Binary Feedback on selecting the block
γ'	Discounting Factor
P_{out}	Outage Probability
η	Energy Efficiency

W	Window length
b	Threshold to check change occured between two windows
δ	Maximum magnitude of change
τ_i	Detection time
α	Set of Blocks lined up before the optimal block
β	Set of Blocks lined up after the optimal block

Chapter 1

Introduction

1.1 Background

Wireless communication systems have witnessed remarkable progress over the past few decades, primarily driven by the need to support ever-increasing data rates, improve spectral and energy efficiency, and enable ultra-reliable low-latency communication (URLLC). The fifth-generation (5G) networks and the anticipated sixth-generation (6G) wireless systems have underscored the limitations of conventional communication paradigms. These paradigms, which focus primarily on optimizing transmission and reception techniques while treating the wireless propagation channel as an uncontrollable entity, are now being re-examined in light of emerging technologies that empower the environment to actively participate in signal shaping and propagation.

Among these technologies, Reconfigurable Intelligent Surfaces (RIS) have garnered substantial interest. RISs are planar metasurfaces comprising a large number of low-cost, passive reflecting elements capable of independently adjusting the phase and amplitude of incident electromagnetic waves. By controlling the phase shifts applied by each element, RIS can reshape

the wireless propagation environment to enhance signal strength, suppress interference, extend coverage, and improve overall network throughput and reliability.

RIS introduces a paradigm shift from conventional channel adaptation to environment reconfiguration, thereby offering new degrees of freedom in wireless system design. Its low power consumption, flexible deployment on surfaces such as building facades, and compatibility with millimeter-wave and terahertz bands make RIS a key enabler for sustainable and energy-efficient wireless communication systems.

Despite these promising attributes, realizing the full potential of RIS hinges on efficient configuration strategies that can adapt in real time to dynamic channel conditions. The phase configuration space is exponentially large, especially when multiple RIS panels and users are involved. Furthermore, practical constraints such as discrete phase resolution, hardware imperfections, and limited channel state information (CSI) at the transmitter pose additional challenges. These factors necessitate the development of low-complexity, real-time optimization frameworks capable of operating with limited feedback.

1.2 Motivation

The deployment of RIS in next-generation wireless systems introduces new challenges in terms of real-time configuration and control. Traditional optimization approaches, such as exhaustive search, convex optimization, and gradient descent methods, are computationally intensive and impractical for systems with a large number of RIS elements. These techniques often assume full or near-perfect CSI, which is difficult to obtain in real-time due to the passive nature of RIS and the complexity of estimating cascaded channels.

Reinforcement learning-based solutions have been proposed to address these limitations, offering a framework to learn optimal policies through interactions with the environment. However, these approaches typically require extensive training data, high computational overhead, and prolonged convergence times, making them unsuitable for latency-sensitive and resource-constrained deployments such as edge computing environments.

The motivation for this thesis stems from the need for lightweight, scalable, and adap-

tive algorithms that can operate efficiently under uncertainty and limited feedback. Online learning algorithms, particularly those based on the Multi-Armed Bandit (MAB) framework, present an appealing solution. MAB algorithms are well-suited for scenarios where decisions must be made sequentially with partial information and minimal overhead. By balancing exploration (trying less-known configurations) and exploitation (choosing known good configurations), MAB can learn optimal or near-optimal RIS configurations over time.

Moreover, the application of MAB in RIS-assisted communication systems allows for a model-free optimization framework, obviating the need for full CSI and enabling efficient adaptation to time-varying channel conditions. This thesis aims to harness the potential of MAB algorithms to address several key challenges in RIS configuration, including combinatorial action spaces, change detection, sensor selection, and edge deployment feasibility.

1.3 Objectives and Contributions

This thesis presents novel contributions to the field of intelligent wireless communications, with a particular emphasis on the use of online learning algorithms for real-time optimization of RIS. The objectives of the thesis are:

- Design and analysis of the algorithm for identification of the best RIS in the scenario of Single Transmitter and single as well as multiple receivers with multiple RIS.
- To design and analyze the algorithm for best RIS identification with the BER and transmitted power trade-off.
- Development of the algorithm for multiple RIS and multiple receivers for homogeneous networks
- Development of the algorithm for multiple RIS and multiple receivers for hetrogeneous networks

The key contributions are outlined below:

- **Modeling RIS Configuration as a MAB Problem:** The thesis introduces a fundamentally new perspective by modeling the RIS sub-block selection task as a stochastic MAB

problem. By abstracting each RIS configuration (i.e., a group of RIS elements or sub-blocks) as an arm in the bandit setting, the problem is framed as a sequential decision-making process. This formulation enables real-time learning of the best RIS configuration that maximizes the SNR at the receiver, even in the absence of full CSI. The approach is model-free, highly scalable, and requires only minimal feedback, making it ideally suited for practical deployment in next-generation wireless systems.

- **Design of Compute-Efficient and Scalable MAB Algorithms:** To tackle the challenges posed by large action spaces inherent in multi-RIS systems, the thesis proposes several algorithmic innovations. These include focused exploration techniques, sub-block reduction strategies, and hierarchical arm selection frameworks. Special attention is given to algorithms that are not only statistically efficient but also computationally lightweight—enabling their execution on resource-constrained platforms. The proposed algorithms are benchmarked against conventional UCB, Thompson Sampling, and heuristic-based methods, demonstrating superior performance in terms of convergence speed, regret minimization, and reliability under limited CSI.

- **Integration of Change Detection Mechanisms for Dynamic Environments:** A novel framework is proposed to handle time-varying wireless channels where the optimal RIS configuration changes due to mobility or environmental dynamics. The integration of passive and active change detection mechanisms into the MAB framework allows the system to adaptively reset and refine its learned decisions in response to abrupt or gradual changes in the channel. This hybrid approach ensures responsiveness to change without incurring unnecessary resets, thereby maintaining high SNR and throughput in dynamic scenarios.

- **Formulation of Energy-Efficient RIS Selection via Sensor Selection Framework:** The thesis draws parallels between RIS sub-block selection and classical sensor selection problems, particularly in contexts involving trade-offs between measurement cost and accuracy. Inspired by partial monitoring settings, a consumed-power-aware MAB algorithm is developed to identify RIS sub-blocks that achieve desired communication performance while minimizing power consumption. This contribution addresses a critical practical aspect of RIS deployment—sustainability and energy efficiency.

- **Theoretical and Simulation-Based Validation Across Diverse Scenarios:** The pro-

posed algorithms are rigorously analyzed through both theoretical regret bounds and extensive MATLAB simulations. Performance metrics include cumulative regret, average received SNR, ergodic capacity, outage probability, and energy efficiency. The simulations encompass single and multi-RIS systems, varied RIS element configurations, discrete phase shift constraints, and both stationary and dynamic environments. These results validate the robustness and adaptability of the algorithms under realistic system assumptions.

- **Real-Time Implementation and Edge Platform Feasibility:** A significant contribution of this thesis is the successful mapping of the proposed MAB algorithms to processor-based edge platforms (e.g., ARM Cortex-A9, Cortex-A53). Execution time, memory footprint, and instruction efficiency are measured and analyzed under different floating-point precisions and SIMD configurations. The results confirm the practical viability of deploying the algorithms in real-time communication systems, showing sub-10ms execution latency and reduced compute overhead.

- **Scalability to Multi-RIS and Multi-User Configurations:** The framework is designed to operate effectively in scenarios involving multiple RIS panels and multiple receivers. Through architectural design and algorithmic layering, the system maintains performance as the number of RIS elements, sub-blocks, or users increases. Cooperative and distributed RIS scenarios are also considered, showing the extensibility of the proposed learning-based approach.

- **Laying the Groundwork for Future Intelligent RIS Systems:** Although this thesis focuses on maximizing SNR using bandit learning, the underlying framework provides a foundation for future extensions such as: multi-objective optimization (e.g., joint radar and communication design in ISAC-RIS systems), contextual and adversarial bandit models, integration with deep reinforcement learning for trajectory or beam design, and feedback-aware resource allocation in mobile networks. These future directions are discussed in the final chapter.

1.4 Organization of the Thesis

The remainder of this thesis is structured into seven chapters, each addressing a specific aspect of RIS-aided wireless communication systems and the proposed learning-based solutions. The contributions span the development of theoretical models, algorithmic innovations, practical implementation strategies, and performance evaluations, as outlined below:

- **Chapter 2: Literature Survey** provides an extensive review of existing research on intelligent and reconfigurable wireless environments. It begins with the evolution of RIS and their role in modern wireless networks, emphasizing their architectural features and signal manipulation capabilities. The chapter discusses applications of Multi-RIS systems, their integration into 5G and beyond networks, and highlights how online learning—particularly MAB algorithms—have emerged as a practical alternative to conventional optimization methods. It also covers advanced topics such as sensor selection under partial monitoring, the importance of change detection in dynamic environments, and the challenges of deploying learning algorithms on hardware-constrained edge platforms.

- **Chapter 3: Online Learning-Based Multi-RIS-Aided Wireless Systems** formulates the problem of RIS block selection in large-scale, multi-RIS-assisted networks. It proposes a lightweight online learning framework based on MAB algorithms to sequentially identify the optimal RIS sub-blocks that maximize the SNR at multiple receivers. The proposed framework operates without full CSI, making it suitable for real-time deployment in practical systems with minimal overhead.

- **Chapter 4: High-Speed Compute-Efficient Bandit Learning for Many Arms** addresses the challenge of scalability when the number of arms (i.e., RIS configurations) becomes very large. This chapter presents a class of computationally efficient MAB algorithms tailored for low-latency processors. It includes algorithmic architectures that reduce execution time while maintaining learning accuracy, making these algorithms highly suitable for deployment on embedded and edge platforms. Theoretical regret analysis and experimental results are presented to support the proposed methods.

- **Chapter 5: Optimizing RIS Block Selection for Power Consumption** investigates the

trade-off between power consumption and SNR performance in RIS systems. Inspired by the sensor selection framework, the chapter introduces learning algorithms that intelligently choose subsets of RIS elements, thereby achieving energy-efficient communication. It formulates the optimization task as a partial monitoring problem where the learner must infer optimal decisions with limited reward feedback, striking a balance between performance and energy usage.

- **Chapter 6: Online Learning and Change Detection-Based Multi-RIS-Aided Wireless Systems for Dynamic Environments** extends the learning framework to dynamic scenarios where the optimal RIS configuration may change over time due to user mobility or varying channel conditions. This chapter integrates change detection techniques with the MAB-based learning structure to enable adaptive configuration updates without requiring prior knowledge of change intervals. Both abrupt and gradual changes are considered, and the proposed approach demonstrates robustness and adaptability in non-stationary environments.

- **Chapter 7: Conclusion and Future Work** summarizes the thesis contributions, providing a synthesis of the proposed algorithms and their practical implications. The chapter discusses the advantages of online learning for RIS configuration and outlines several potential future research directions, including the integration of deep learning, mobility-aware RIS control, and large-scale hardware implementation strategies.

This page was intentionally left blank.

Chapter 2

Literature Survey

2.1 RIS: Fundamentals and Applications

RIS have emerged as a revolutionary concept in the design of future wireless communication systems. Unlike conventional infrastructure, which treats the wireless channel as an uncontrollable entity, RIS enables programmable control over the propagation of electromagnetic waves. RIS typically comprises a two-dimensional array of low-cost, passive reflective elements, each capable of independently adjusting the phase—and in some designs, the amplitude—of incident signals. By properly configuring these elements, RIS can shape the wireless environment to enhance signal strength, direct the signals towards specific users, reduce interference, and improve the overall coverage and capacity of wireless networks (1).

RIS has garnered increasing attention for its potential to meet the growing demands of next-generation networks, such as 6G, which require highly adaptive, energy-efficient, and spectrum-efficient communication technologies. Applications of RIS include mmWave and THz communication, physical layer security, non-line-of-sight (NLoS) communication, 3D positioning, and vehicular-to-everything (V2X) systems (1; 2; 3). Moreover, RIS offers notable

advantages such as minimal energy consumption due to its passive nature, flexible deployment on surfaces like walls or building facades, and the ability to complement or replace active relays in certain scenarios (4). From a theoretical perspective, RIS represents a paradigm shift in the way wireless systems are modeled and optimized. Traditional communication systems treat the propagation environment as a passive medium. In contrast, RIS enables the wireless environment to become an active participant in signal transmission, offering a new dimension of controllability (5). This shift has prompted extensive research into the modeling of RIS-assisted channels, including studies on line-of-sight and non-line-of-sight propagation, as well as frequency-selective fading scenarios.

Recent research has focused on integrating RIS into various wireless architectures, exploring its performance benefits under different channel conditions and its potential to support intelligent and environment-aware communication systems (1). Some of the key technical challenges in RIS deployment include channel estimation, phase quantization limitations, and synchronization among multiple RIS elements (1). In response to these challenges, various signal processing and optimization techniques have been developed, such as iterative beamforming, convex optimization, and heuristic algorithms (5). These methods aim to approximate the optimal RIS configuration to enhance system performance.

One of the most studied use cases for RIS is its deployment in mmWave communication systems, where the high path loss and sensitivity to blockage pose significant challenges. RIS can overcome these issues by creating alternative reflective paths that bypass obstacles and extend communication range (5). In THz systems, RIS has been explored for its ability to provide beam focusing and reduce path loss in environments with limited scattering(2; 6). In addition to enhancing communication performance, RIS has also been applied in sensing and localization systems, enabling functionalities such as 3D imaging and user tracking through intelligent reflection control (7; 8).

The integration of RIS with multiple-input multiple-output (MIMO) systems has also been a focus of recent studies(9). Joint optimization of the transmit beamforming and RIS phase shifts can yield significant gains in achievable rate and energy efficiency(10). However, such joint optimization problems are often non-convex and computationally intensive, prompting the exploration of suboptimal yet efficient algorithms. Moreover, the assumption of perfect channel state information (CSI) at both the transmitter and RIS is often unrealistic in practical

scenarios. This has motivated research into robust RIS designs that operate effectively under partial or outdated CSI (11; 12).

RIS is also considered a key enabler for green communications, as it offers the potential to significantly reduce power consumption in wireless networks. Unlike traditional active relays or base stations, RIS elements consume minimal power, primarily limited to the circuitry used to adjust their phase shifts (13). This makes RIS particularly attractive for deployment in dense urban environments, where energy efficiency and spectrum reuse are critical. Several studies have proposed RIS-based solutions for energy harvesting, cooperative relaying, and interference management in such scenarios (14; 15).

Despite its potential, realizing the full benefits of RIS requires solving a number of practical and theoretical challenges. These include optimizing the placement and orientation of RIS panels, managing the trade-off between complexity and performance in phase shift design, and ensuring compatibility with existing communication standards (8). Furthermore, as the number of RIS elements increases, the configuration space becomes exponentially larger, making real-time optimization increasingly difficult. This has sparked growing interest in machine learning-based methods, including reinforcement learning and multi-armed bandits, which can adapt to dynamic environments and learn optimal configurations from interaction data.

In summary, RIS presents a highly promising solution to the challenges faced by modern wireless communication systems. It offers a flexible and energy-efficient means to enhance coverage, capacity, and reliability. However, the potential of RIS can only be fully realized through the development of intelligent algorithms capable of efficiently configuring its elements in real time (8). This motivates the research in this thesis, which explores the use of online learning, particularly multi-armed bandit algorithms, as a scalable and effective approach for RIS optimization in practical communication environments.

2.2 RIS Channel Modeling and System Architectures

RIS have emerged as a transformative technology to reshape wireless communication environments through the programmable control of electromagnetic waves. Accurate channel modeling and practical architectural considerations are critical for realizing the potential of RIS-assisted

systems. This section elaborates on the various channel modeling techniques, assumptions used in the literature, and RIS deployment architectures suited for modern wireless networks.(16; 17)

2.2.1 RIS-Assisted Wireless Channel Modeling

The accurate characterization of wireless channels in the presence of RIS is essential for performance analysis and system design. Unlike conventional wireless systems where signals undergo reflection and scattering passively through environmental objects, RIS allows for programmable control over the propagation medium(18). As such, traditional models must be revisited to incorporate the controllable reflection and phase shift introduced by RIS elements.

RIS-assisted channels are typically modeled as a concatenation of two or more individual links:

- Transmitter-to-RIS link.
- RIS-to-receiver link.
- Direct link between transmitter and receiver.

Let us denote the baseband equivalent complex channel between the transmitter and the receiver through the RIS as the product of two matrices/vectors:

$$\mathbf{h}_{\text{eff}} = \mathbf{H}_{\text{RIS, RX}} \boldsymbol{\Theta} \mathbf{H}_{\text{TX, RIS}}, \tag{2.1}$$

where $\mathbf{H}_{\text{TX, RIS}}$ and $\mathbf{H}_{\text{RIS, RX}}$ represent the channels between the transmitter-RIS and RIS-receiver respectively, and $\boldsymbol{\Theta} = \text{diag}(\alpha_1 e^{j\theta_1}, ..., \alpha_Z e^{j\theta_Z})$ is the RIS reflection coefficient matrix with amplitude and phase control per element.

The literature assumes different channel fading models based on the deployment environment:

- **Rician fading** for line-of-sight (LoS)-dominant channels(19; 20).
- **Rayleigh fading** for non-LoS links in rich scattering environments(21).

- **Nakagami-*m* fading** for modeling flexible fading severity, as adopted in (22; 23).

Moreover, some works consider independent and identically distributed (IID) fading across RIS elements, while others use spatially correlated models to capture practical scenarios more accurately.

2.2.2 Near-Field vs. Far-Field Modeling

RIS-assisted systems may operate under different field regions depending on the communication distance, RIS size, and wavelength(24). Traditionally, far-field assumptions are made where plane-wave propagation is valid(25). However, large RIS panels or proximity to users/transmitters may invalidate these assumptions.

- **Far-field model**: Assumes planar wavefronts; simplifies path loss and phase calculation.
- **Near-field model**: Requires spherical wave modeling; accounts for wave curvature, especially in large aperture RIS or indoor scenarios.

Recent works, such as (26), emphasize the importance of using near-field models for accurate simulation of RIS-aided short-range networks.

2.2.3 Element-Level Reflection Models

The reflection model of RIS elements plays a pivotal role in determining the composite channel. Each passive element in the RIS introduces a phase shift and, optionally, an amplitude attenuation. The simplified ideal model assumes unit-amplitude reflection with continuous phase control, but in practice, constraints such as:

- **Quantized phase shifts** (e.g., 2-bit or 4-bit phase control),
- **Element mutual coupling**,
- **Limited amplitude control**

must be accounted for.

These imperfections are modeled using realistic phase-shift constraints or hardware-induced error distributions, as detailed in (27).

2.2.4 RIS Deployment and System Architectures

RIS can be deployed in various topologies and roles, influencing the corresponding system model:

1. **Single-RIS Systems**: Most early works analyze systems with one RIS positioned to aid communication between a transmitter and a receiver. (28)
2. **Multi-RIS Systems**: More recent approaches, such as (7; 23; 29; 30; 31), explore scenarios with multiple spatially distributed RISs to extend coverage and enhance spatial diversity. Such setups introduce new challenges in resource selection, channel estimation, and coordination.
3. **User-Centric RIS**: RIS is moved closer to mobile users, offering personalized enhancement(32; 33).
4. **Cell-Free Architectures**: RISs are treated as part of a distributed MIMO network, coordinated through centralized or decentralized algorithms(34; 35).

These issues motivate data-driven and learning-based approaches for RIS configuration, paving the way for the subsequent sections of this thesis.

2.3 RIS Configuration Optimization: Traditional Approaches

The efficient configuration of RIS is a challenge in realizing their full potential in wireless communication systems. Since RIS can actively influence signal propagation by manipulating the phase and amplitude of incident signals(2), determining the optimal configuration of RIS elements is crucial to achieving desirable performance metrics such as SNR, data rate, outage

probability, and energy efficiency. Traditional optimization approaches have been extensively explored in the literature to solve this configuration problem, often under various assumptions regarding channel knowledge, RIS hardware constraints, and computational resources.

One of the most straightforward methods for RIS configuration is the exhaustive search approach, wherein all possible combinations of RIS phase shifts are evaluated to identify the configuration that maximizes a particular performance metric, typically the SNR or channel gain. While this approach guarantees optimal performance, it is computationally infeasible for even moderately sized RIS arrays. The search space grows exponentially with the number of RIS elements and the resolution of phase quantization(13; 36). For example, a RIS with 64 elements and 4-bit phase resolution results in more than 2^{256} possible configurations. Clearly, this method becomes impractical for real-time implementation and is generally used only for benchmarking purposes in simulations. To mitigate the complexity of exhaustive search, researchers have proposed heuristic and iterative algorithms. One such approach is coordinate descent, which sequentially optimizes the phase of each RIS element while holding the others fixed. This method reduces the computational burden significantly but may converge to local optima depending on the initialization and channel conditions. Another commonly used technique is the alternating optimization framework, where the beamforming at the transmitter and the RIS configuration are alternately optimized. This approach often assumes the availability of full channel state information (CSI) and relies on iterative procedures that may still be computationally expensive. Gradient-based methods have also been applied to RIS configuration problems, particularly when the optimization objective is differentiable with respect to the RIS phase shifts(37; 38; 39). These methods utilize gradient descent or projected gradient algorithms to navigate the solution space. However, the non-convex nature of the RIS optimization problem—due to the unit-modulus constraints on the RIS elements—poses significant challenges for gradient-based methods. The presence of multiple local minima can lead to suboptimal solutions, and the requirement of gradient information implies the need for either accurate channel models or substantial training overhead. Convex relaxation techniques, such as semidefinite relaxation (40; 41) and successive convex approximation (SCA)(42; 43), have also been employed to handle RIS optimization problems. SDR lifts the problem to a higher-dimensional space where it becomes convex, solves the relaxed version, and then extracts a feasible solution through randomization or rounding. While these methods provide better tractability, they often require high-dimensional matrix operations and are not scalable to

large RIS arrays. Moreover, the performance gap between the relaxed and original problems can be significant in some scenarios. In addition to these mathematical optimization techniques, metaheuristic algorithms such as genetic algorithms(44), particle swarm optimization(45), and simulated annealing have been explored for RIS configuration(46). These algorithms are inspired by natural processes and are capable of exploring a large solution space without relying on gradient information. Although they offer more flexibility in handling non-convex and discrete optimization problems, they typically involve many hyperparameters and can be sensitive to tuning. Furthermore, their convergence speed may not be sufficient for real-time applications, especially in rapidly changing channel conditions. Another line of research focuses on codebook-based approaches(47), where a finite set of RIS configurations is pre-designed based on offline optimization or heuristic rules. During operation, the system selects the best configuration from the codebook based on limited feedback or signal measurements(47; 48). This reduces the search complexity but often leads to suboptimal performance due to the limited diversity in the codebook. Furthermore, codebook design itself can be a complex task, requiring channel statistics and prior knowledge of the deployment scenario. Traditional optimization techniques also typically assume perfect or near-perfect knowledge of the wireless channel, which is difficult to obtain in practice, especially in RIS-assisted systems where the base station must estimate the cascaded channel involving the transmitter-to-RIS and RIS-to-receiver links. Channel estimation in such scenarios requires sophisticated protocols and incurs significant overhead, further reducing the practicality of these methods(12; 49). To address this issue, some researchers have proposed blind or semi-blind optimization techniques that rely on received signal strength measurements instead of full CSI. While these methods reduce the feedback burden, they tend to converge slowly and are often sensitive to noise and environmental variations. The majority of traditional RIS optimization methods are also designed for single-user scenarios or assume orthogonal access in multi-user environments. However, the increasing demand for spectral efficiency necessitates simultaneous optimization for multiple users sharing the same RIS infrastructure(12; 50). Extending traditional methods to multi-user setups introduces additional complexity, as the optimal RIS configuration must balance the performance across users with potentially conflicting objectives. Moreover, the problem becomes combinatorially harder when multiple RIS panels or distributed RIS elements are involved. In practical deployments, several hardware constraints must also be considered. These include the finite resolution of the phase shifters, mutual coupling between RIS elements, and latency in

reconfiguring the RIS states. Traditional optimization methods rarely account for these non-idealities explicitly, leading to performance degradation when implemented on real hardware. Additionally, the overhead involved in configuring RIS in real-time is often ignored, which can be substantial in scenarios where rapid adaptation is required, such as vehicular networks or mobile users. In summary, while traditional optimization techniques have provided valuable insights into the performance limits and design principles of RIS systems, they face several limitations in real-world deployments. These include:

- High computational complexity, making real-time adaptation infeasible.
- Dependence on full CSI, which is expensive to obtain and prone to errors.
- Lack of scalability to large RIS arrays or multi-user scenarios.
- Insufficient adaptability to dynamic and time-varying environments.
- Neglect of practical constraints, such as energy efficiency, hardware imperfections, and feedback limitations.

These limitations highlight the need for new approaches that can offer low-complexity, data-driven, and adaptive solutions to the RIS configuration problem. Machine learning, and in particular, online learning algorithms such as MAB, have recently emerged as promising alternatives(51; 52; 53). These approaches are capable of learning optimal RIS configurations from limited feedback and adapting to changing environments without requiring complete channel knowledge. In the next section, we explore how learning-based strategies, including MAB, have been applied to RIS optimization and how they overcome the shortcomings of traditional techniques.

2.4 Multi-User and Multi-RIS Systems

As wireless networks continue to densify and diversify in structure, the need to efficiently support multiple users across extended coverage areas has become paramount. RIS offer an innovative solution by enabling programmable wireless environments. However, extending RIS systems to multi-user and multi-RIS settings introduces a host of technical challenges and design

considerations. This section reviews the literature and underlying principles in such systems, with a focus on architecture, coordination, optimization, and learning-based adaptation.

2.4.1 Multi-User RIS-Aided Communication

In a multi-user RIS-aided system, a central base station (BS) communicates with multiple receivers via a passive RIS, which reflects the incident signal toward users by applying phase shifts at its elements. These systems aim to improve spectral efficiency, reduce interference, and provide user-specific beamforming without the need for active RF chains at the RIS(54; 55; 56).

The major challenges in this context include:

- **Inter-user Interference (IUI):** Optimizing the RIS to serve multiple users simultaneously may lead to signal degradation due to mutual interference among user channels.(57)
- **Coupled Optimization:** The BS transmit beamforming and RIS reflection coefficients must be optimized jointly, forming a highly non-convex and coupled problem. (58)
- **Fairness and QoS:** Ensuring that all users meet their quality-of-service (QoS) constraints, such as minimum SINR or throughput levels.
- **CSI Acquisition:** Obtaining accurate cascaded BS-RIS-user CSI is particularly challenging in multi-user scenarios. (59)

2.4.2 Multi-RIS Deployment Architectures

To enhance coverage and provide spatial diversity, multiple RISs can be deployed within a communication environment. These RISs may operate independently or be coordinated centrally depending on the network architecture. Multi-RIS systems offer multiple propagation paths, leading to improved link reliability and energy focusing capabilities. (23; 29; 60; 61)

The advantages of multi-RIS deployments include:

- **Extended Coverage:** RISs can be positioned to mitigate shadowing and blockage effects,

ensuring communication continuity.

- **Diversity Gains:** Signals reflected by different RISs undergo independent fading, improving the robustness of communication links.
- **Flexible Association:** Users can dynamically associate with different RISs depending on their location and channel conditions.

However, multi-RIS systems also pose new design challenges:

- **RIS Selection:** Determining which subset of RISs should be activated for each user or group of users.(53)
- **Coordination Complexity:** Managing phase shift design across distributed RISs requires sophisticated coordination protocols.
- **Synchronization:** Aligning multiple RISs to operate coherently in time and phase.
- **Inter-RIS Interference:** In dense deployments, reflections from multiple RISs can destructively interfere unless carefully aligned.

2.4.3 Joint Optimization in Multi-User Multi-RIS Systems

The joint optimization of BS beamforming, RIS phase shifts, and user-RIS association leads to large-scale, highly non-convex problems. Several solution approaches have been developed:

- **Block Coordinate Descent (BCD):** Alternately updates different sets of variables while fixing others. Commonly used for RIS phase shift and precoding design.(50)
- **Successive Convex Approximation (SCA):** Approximates non-convex constraints and objective functions with convex surrogates.(13; 62)
- **Semidefinite Relaxation (SDR):** Used to relax rank-one constraints in beamforming problems, though often leading to suboptimal performance.(41)
- **Heuristic and Metaheuristic Algorithms:** Includes evolutionary methods, particle swarm optimization, and simulated annealing for global search.(45)

Many of these methods are computationally intensive and rely on full or partial CSI, which may not be available in real-time systems. This has led to increased interest in learning-based strategies.

2.4.4 Learning-Based Approaches in Multi-RIS Systems

Online learning algorithms, including Multi-Armed Bandits (MABs) and reinforcement learning, have been proposed as scalable alternatives to conventional optimization. These methods treat RIS selection, configuration, and association as sequential decision problems.

MAB-Based Approaches: Each RIS or sub-block can be treated as an arm, and the transmitter learns over time which combinations maximize a reward function (e.g., SNR, throughput, energy efficiency)(53).

Reinforcement Learning: Q-learning and deep Q-networks (DQN) have been used to model RIS-user association and dynamic environment adaptation, as seen in (51) and (63).

Federated and Distributed Learning: Emerging frameworks allow each RIS to learn and adapt independently while sharing minimal information with a central controller, preserving privacy and reducing overhead.(64)

2.5 RIS in Millimeter-Wave and Terahertz (THz) Bands

RIS are increasingly recognized as critical enablers for high-frequency communication systems, particularly in the mmWave and THz bands. These bands offer large spectral resources to support ultra-high data rates for emerging 6G applications, including immersive extended reality (XR), ultra-HD video streaming, and wireless backhaul. However, mmWave and THz systems suffer from severe path loss, high penetration losses, and limited diffraction, all of which restrict coverage and link reliability(5; 65; 66). RIS provides a passive, energy-efficient means to overcome these limitations by introducing programmable reflections that can compensate for propagation losses and maintain connectivity in non-line-of-sight (NLoS) scenarios.

2.5.1 RIS Design Considerations in High-Frequency Bands

The implementation of RIS in mmWave and THz bands requires specialized design to accommodate unique propagation and hardware constraints:

- **Element Size and Spacing:** At higher frequencies, the wavelength is smaller, enabling more RIS elements to be packed in a given aperture, which increases the angular resolution of beamforming.
- **Phase Shift Precision:** Fine phase control becomes more critical due to the short wavelengths, demanding higher-resolution or continuous-phase tuners.
- **Material and Substrate Limitations:** RIS elements must be designed with materials compatible with high-frequency operations, such as graphene-based or plasmonic metasurfaces for THz bands.
- **Integration with Beam Training:** Due to narrow beamwidths at mmWave/THz, RIS design must be tightly integrated with beam discovery and alignment mechanisms.

Several hardware prototypes have been demonstrated using PIN diodes, varactors, or micro-electromechanical systems (MEMS) to provide the reconfigurability required at high frequencies.

2.5.2 RIS-Aided mmWave/THz System Architectures

In RIS-enhanced mmWave and THz systems, the RIS can serve different roles depending on deployment strategy:

- **RIS as a Reflective Relay:** Positioned between the transmitter and receiver to establish virtual line-of-sight links. (67)
- **RIS-Assisted Beam Tracking:** Supports fast alignment of narrow beams in mobile scenarios, where conventional beam sweeping is too slow.(68; 69)

- **RIS for Cell-Free Massive MIMO:** Distributed RISs act as passive access points to emulate a dense, energy-efficient MIMO environment.(70)

- **RIS-Enabled Wireless Backhaul:** Facilitates high-capacity, low-cost wireless backhaul in ultra-dense networks.(71)

Works such as (72) and (73) have demonstrated through simulations and prototypes that RISs can significantly reduce outage probability and enhance spectral efficiency at mmWave frequencies. THz studies, such as (68), explore RIS-based reflectarrays to overcome alignment and mobility issues in short-range ultra-high-speed links.

2.6 Multi-Armed Bandit (MAB) Algorithms in Wireless Communication Systems

The MAB framework offers a robust mathematical model for decision-making under uncertainty. Widely applied in various domains, the MAB framework has proven particularly useful in wireless communication systems, where efficient resource allocation decisions must be made dynamically, often with limited or incomplete knowledge of the system's environment.

2.6.1 Introduction to MAB

In the canonical stochastic MAB setting, an agent has K arms, each associated with an unknown reward distribution. At each time step t, the agent selects an arm $I_t \in \{1, \ldots, K\}$ and receives a stochastic reward $r_{i_t}(t)$ drawn from the corresponding distribution. The goal is to maximize the cumulative expected reward over a time horizon T, or equivalently, minimize the *regret*, defined as the difference between the reward of the optimal arm and the accumulated reward of the chosen sequence.

Formally, the expected regret is given by:

$$\mathscr{R}(T) = T\mu^* - \sum_{t=1}^{T} \mathbb{E}[\mu_{i_t}] \tag{2.2}$$

where $\mu^* = \max_i \mu_i$ is the expected reward of the optimal arm. The key challenge in MAB is balancing *exploration* (trying out uncertain arms to learn their rewards) and *exploitation* (selecting the currently best-known arm).

Variants such as the adversarial bandit, contextual bandit, and non-stationary bandits have been developed to address more complex dynamics in practical systems.

2.6.2 MAB Algorithms

Several foundational algorithms have been developed to address different MAB settings:

- **ε-Greedy**: With probability $1-\varepsilon$, the agent selects the best-known arm; otherwise, it explores randomly.
- **Upper Confidence Bound (UCB)**: A deterministic strategy that selects the arm with the highest upper confidence index. UCB1 achieves logarithmic regret under stationary assumptions.(74)
- **Thompson Sampling (TS)**: A Bayesian approach that maintains a posterior distribution over expected rewards and samples from it to make decisions.(75; 76)
- **EXP3**: Suitable for adversarial settings where rewards may be chosen by an adversary.
- **Sliding Window UCB, Discounted UCB**: Designed for non-stationary environments where reward distributions may change over time.
- **Combinatorial MAB (CMAB)**: Where the agent selects a subset of arms at each time step; relevant for scenarios like RIS sub-block or beam combinations.

These methods provide formal performance guarantees and have shown strong empirical results in various domains including wireless networks.

2.6.3 Applications in Classical Wireless Systems

MAB algorithms have been successfully applied in many wireless communication problems, particularly those requiring adaptive, low-overhead decision-making:

- **Channel Selection in Cognitive Radio Networks**: Secondary users choose frequency bands for transmission based on limited knowledge of the interference levels in those bands, aiming to maximize communication quality while avoiding interference.
- **Power Control and Scheduling**: In distributed networks, MAB can guide transmitters in choosing power levels or user subsets.
- **User Association in HetNets**: Base stations can model user association as a MAB problem with uncertain or time-varying load statistics.
- **Beam Selection in mmWave and Massive MIMO Systems**: Sparse channel characteristics and beam misalignment can be addressed via adaptive beam probing using bandit models.

2.6.4 Comparison with Other Learning Techniques

While supervised learning and deep reinforcement learning (DRL) offer rich modeling capacity, they also require significant training data, computational resources, and suffer from poor interpretability. In contrast, MABs:

- Require no offline training or labeled data.
- Are robust to noise and non-stationarity.
- Offer theoretical performance guarantees.
- Have lower complexity and better explainability — critical for real-time embedded implementations.

These advantages make MABs attractive for dynamic RIS control, especially in energy-constrained and latency-sensitive environments.

This page was intentionally left blank.

Chapter 3

Online Learning Based Multi-RIS-Aided Wireless Systems

3.1 Overview

The evolution of software-defined radios and RIS has enabled on-the-fly control and reconfigurability at the physical layer parameters and radio propagation environment. In multi-RIS-aided communication, the RIS block, comprising a certain number of RIS elements from one or more RIS, is selected to achieve high throughput reliable communication between transmitter and receiver. However, selecting an RIS block when there are multiple RIS and receivers is not trivial due to the large number of candidate blocks. In this chapter, a novel multi-armed bandit (MAB) framework, which can learn and select the optimal RIS block using focused exploration, is proposed. We provide the theoretical regret bound for the proposed algorithm and demonstrate the gain in performance over existing state-of-the-art statistical and MAB approaches via detailed simulation results in terms of rate, Ergodic capacity, outage probability, energy efficiency, and received SNR.

3.2 Introduction

After an initial breakthrough in the evolution of analog radio to digital radio, there have been exciting contributions to making digital radio software-controlled and software-defined. This is followed by intelligent radio evolution in which the radio parameters such as carrier frequency, bandwidth, number of antennas, beam directions, and various physical layer parameters can be controlled on the fly (77; 78). One well-known example of such intelligent radio is cognitive radio, through which dynamic spectrum access has become a reality (78). Still, there is limited control over the radio propagation environment, and it limits the performance of cellular networks, resulting in poor throughput and a large number of call drops. In the last few years, we have seen significant interest in making the radio environment smart and controllable due to its potential to improve network performance and cellular coverage (79). Specifically, the aim is to exploit smart radio's on-the-fly configuration capabilities by controlling the transmitted signal's reflection, scattering, and refraction to improve the network performance. One potential solution is reconfigurable intelligent surfaces (RIS) comprised of multiple passive metamaterial reflecting elements (80). By controlling the amplitude and phase of these elements via the controller, we can redirect the transmitted signal in the direction of the intended receivers (81; 2). Various studies showed that the RIS offers higher reliability, improved throughput, and low-cost solutions by reducing the number of RF chains in massive antenna systems and enabling full duplex communication (82; 83). Other RIS applications include improved security (51) and target localization (7). The challenges related to channel estimation due to many RIS elements and robust beamforming due to imperfect channel state information were also considered in the literature (11; 12). In beyond 5G applications, RIS was explored for improving the coverage of mmWave systems where propagation loss is significant (84). Recent works discussed channel modeling for RIS-based communication in the Terahertz (THz) spectrum (73). In (85; 86; 51), machine learning and deep learning algorithms were explored to improve the performance of RIS-based communication.

Recently, few works have considered multiple RIS to improve the throughput and reliability of communication links. In (87), cooperative beamforming design for multi-RIS-aided systems was discussed along with channel modeling. They demonstrated the superiority of distributed RIS over centralized RIS. In (88), an efficient pilot transmission method was proposed

to enable the receiver to separate signals arriving from different RIS, which further helps in estimating the channel state and localization. In (31), the challenge of timing synchronization between multiple RIS was addressed while the accurate configuration of multiple RIS, especially in mmWave multi-antenna system, was discussed in (30). In (89), the application of stochastic geometry aided robust beamforming approach showed improved performance of the multi-RIS-aided system in mmWave. The accurate channel estimation in an RIS-aided system is challenging. A robust transceiver design for the multi-RIS assisted multi-antenna system with erroneous channel estimation was discussed in (61). In (90), a multiple RIS-assisted multi-hop communication system, in which multiple RISs work in a domino pattern, was explored along with closed-form expressions for ergodic capacity and the outage probability. In (60), the need for backhaul support for the feasibility of a multi-RIS system was highlighted. Recent works demonstrated the usefulness of a multi-RIS-aided system for indoor THz wireless communication in the presence of mobile human blockage (91) and for vehicular communication (92). In (23), authors considered a multi-RIS-aided communication system and proposed two schemes for RIS selection. In the first scheme, Exhaustive RIS-aided (ERA), all RIS were used, while in the second scheme, Opportunistic RIS-aided scheme (ORA), single optimal RIS was used. In (29), authors considered relay-aided communication systems using multiple RIS and compared the gain in outage probability. However, most of these works assume prior knowledge of perfect channel state information (CSI).

To address scalability issues, prior knowledge of CSI and RIS selection, the multi-armed bandit (MAB) framework was explored in the literature (93; 94). MAB framework based online learning algorithms are a type of reinforcement learning in which the player needs to learn and select the optimal arm as many times as possible. Since the arm statistics are unknown, the algorithm must optimally satisfy the exploration-exploitation trade-off. MAB algorithms are being used extensively to develop learning algorithms for decision-making due to their analytical traceability. Further, they are compute and memory efficient and do not need prior training like machine learning and deep learning algorithms (95). Popular wireless applications of MAB-based decision-making include cognitive radio networks (96; 97), Energy harvesting networks (98), Millimeter Wave communications (99), massive multi-antenna system (100; 101) and 5G cellular networks (102). We focus on the stochastic setting where classical bandit algorithms like UCB (103), KL-UCB (104), and Thompson sampling (105) are applied to networks by suitably mapping actions (channels, modulations schemes, power level, RIS) to arms and quantity

interests (success rate, throughput) to rewards to obtain learning algorithms. Recently, MAB algorithms were used for the selection of RIS in cellular Internet-of-Things (C-IoT), unmanned aerial vehicle (UAV), and millimeter waves (mmWave) applications in (93), (106) and (94), respectively.As discussed later in Section 3.4.2, direct application of MAB algorithm for RIS selection in multi-RIS-aided system results in poor performance. To the best of our knowledge, the design of MAB algorithms for multi-RIS-aided wireless systems has not been discussed yet in the literature. The MAB problem setup for multiple receivers is not trivial. This work aims to develop a new efficient MAB algorithm for wireless systems comprised of multiple RIS and receivers. We analyze the performance of the proposed algorithm via theoretical, simulation, and experimental results on the edge platform. In Table 3.1, we compare various state-of-the-art works in the literature, and our contributions are summarized as follows:

1. In this chapter, we set up the RIS selection problem in a multi-RIS-aided system with multiple receivers as a MAB framework where each arm corresponds to one or more RIS with a certain number of passive elements. The aim is to learn the optimal combination of RIS and their elements that offer higher average SNR/throughput. Since the number of RIS blocks can be significantly large, resulting in higher exploration time of the conventional MAB algorithm, we exploit the sparsity to significantly reduce the number of candidate blocks by pre-identifying the blocks whose average throughput is above a given threshold. We demonstrate the gain in performance over existing state-of-the-art approaches via theoretical and simulation results.

2. We propose two MAB algorithms: focused exploration UCB (FEUCB) and Enhanced focused exploration UCB (EFEUCB) algorithms. We compare the performance with existing learning and non-learning-based approaches in terms of various performance metrics such as regret, rate, outage probability, ergodic capacity, energy efficiency, transmit power, and consumed power.

3. We compare the performance of the centralized RIS and distributed RIS systems and demonstrate the superiority of the distributed multi-RIS-aided system for multiple-user cases.

4. In time-slotted communication, faster RIS selection allows more time for data communication, resulting in higher throughput. To analyze this, we design and implement the

Table 3.1: Comparison of State-of-the-art Works

Papers	**No. of RIS**	**No. of Recievers**	**Active RIS**	**Learning used for Selection**	**Hardware Results**
(51)	Single	Multiple	One	Yes	No
(106)	Single	Multiple	One	Yes	No
(87)	Multiple	Multiple	Multiple	No	No
(90)	Multiple	Multiple	Multiple	No	No
(23)	Multiple	Single	All (ERA), One(ORA)	No	No
(93)	Multiple	Multiple	One	Yes	No
(94)	Multiple	Single	One	Yes	No
Proposed Work	Multiple	Multiple	Multiple	Yes	Yes

proposed and existing MAB-based RIS selection algorithms on various embedded edge platforms such as ARM Cortex A9 and ARM Cortex A53 processors augmented with single instruction multiple data (SIMD) co-processors. We study the effect of word length on the latency, i.e., execution time, of the algorithms. We demonstrate the lower execution time of the proposed algorithms compared to the state-of-the-art MAB algorithms.

3.3 Network Model

In this chapter we consider the multi-RIS-aided system comprising one transmitter, K RIS, and N receivers. Each RIS is divided into L sub-blocks, and each sub-block consists of Z sub-λ sized passive RIS elements (In Fig. 3.1). We consider the time-slotted communication where the transmitter must select M out of KL sub-blocks in each time slot. Thus, there are $R=\binom{KL}{M}$ combinations of M sub-blocks. The combination of any M sub-blocks is referred to as a block. For the chosen block, the amplitude, $A \in [0,1]$, and phase, $\theta \in [0,2\pi]$, of all elements are optimally configured to establish the communication with N receivers. We denote the block selected in the time slot, t, as $I_t \in [R] := \{1,2,..R\}$. Note that the size of each sub-block and the number of sub-blocks per RIS are not fixed. Without loss of generality, our framework allows the transmitter to select more than one RIS in a given slot by having the sub-blocks in I_t from the same or multiple RIS.

We consider the channel model in (23), where channels associated with elements of the

same RIS are assumed to be independent and identically distributed (IID). In contrast, channels associated with different RISs are assumed to be independent but not identically distributed (INID), and the system undergoes Nakagami-m fading. The complex channel coefficient between the transmitter and z^{th} element of the RIS sub-block, r, is defined in the polar form as $\tilde{G}_{klz} = G_{klz}e^{j\psi_{klz}}$ where G_{klz} and ψ_{klz} denote the magnitude and phase, respectively. Similarly, the channel between the z^{th} element of the RIS sub-block, r, and n^{th} receiver is defined in the polar form as $\tilde{H}_{klnz} = H_{klnz}e^{j\phi_{klnz}}$ where H_{klnz} and ϕ_{klnz} denote the magnitude and phase, respectively. The channel coefficient along the direct path between the transmitter and n^{th} receiver is $\tilde{G}_n = G_n e^{j\psi_n}$ where G_n and ψ_n denote the magnitude and phase, respectively.

For transmit symbol, x_s with power, P_s in dBm, the received signal at the n^{th} receiver using the RIS sub-block, m, is given by (23)

$$y_n^m = \sqrt{P_s}\left(\tilde{G}_n + \sum_{z=1}^{z=Z} \tilde{G}_{klz}\alpha_{klz}e^{j\theta_{klz}}\tilde{H}_{klnz}\right)x_s + \omega_n \tag{3.1}$$

where α_{klnz} and θ_{klnz} are the amplitude reflection coefficient and phase-shift of the z^{th} element of the m. ω_n denotes the additive white Gaussian noise (AWGN) with zero mean and variance, σ_n^2. Then, the SNR at the n^{th} receiver, assuming optimal RIS configuration with zero phase error, is given as

$$\mathrm{SNR}_n^m = \frac{P_s}{\sigma_n^2}\left|G_n + \sum_{z=1}^{z=Z} G_{klz}\alpha_{klz}H_{klnz}\right|^2 \tag{3.2}$$

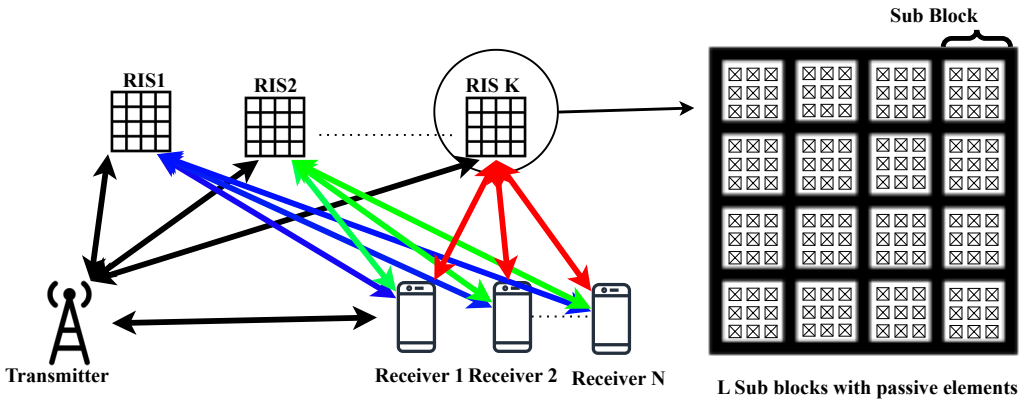

Figure 3.1: Illustrations of network model for multi-RIS-aided wireless system.

For reliable communication between the transmitter and all receivers, we need to select the block, i.e., M RIS sub-blocks, which cumulatively offer an average high cumulative SNR and guarantee that the SNR at each receiver is above a certain threshold, Δ.

$$I_t = \underset{|I_t|=M}{I_t \in [R]} \left[\sum_{\substack{v=i_1 \\ v \in [I_t]}}^{i_M} \left(\prod_{n=1}^{N} 1_{\mathrm{SNR}_n^r > \Delta} \sum_{n=1}^{N} \mathrm{SNR}_n^r \right) \right] \tag{3.3}$$

where 1_{cond} is an indicator function. Selection of all RIS results in good throughput performance but extremely high power consumption. Since SNR over each sub-block is unknown, the selection of RIS block is not trivial; hence, we need a learning framework. Due to a large number of sub-blocks and multiple receivers, one-shot learning via deep learning is not scalable (63). The MAB algorithm is well suited for scenarios where decisions have to be made in the face of uncertainty, and the objective is to maximize accumulated rewards over time. Though such sequential learning is a promising approach, it is time-consuming, and hence, efficient algorithms are needed.

3.4 Proposed Work

The selection of the RIS block in a multi-RIS-aided system with multiple receivers can be set up as an MAB problem where the aim is to select the optimal block as often as possible via exploration-exploitation trade-off. In this section, we discuss the selection of optimal block for multiple receiver scenarios and the drawbacks of the existing MAB algorithms. We address these challenges using the proposed algorithms.

3.4.1 MAB Framework

The standard stochastic multi-play MAB consists of a set of arms and a single player. Here, we refer to RIS sub-blocks as arms and a player as a transmitter. In each time slot, a transmitter selects a block of M RIS sub-blocks and receives a reward equal to SNR at each receiver. For each receiver, the reward is assumed to be drawn independently across time from distributions that

are stationary and independent across sub-blocks. We denote the mean reward of n^{th} receiver for block, r as SNR_n^r and hence, total reward in time slot, t is given as

$$X^{I_t} = \sum_{\substack{v=i_1 \\ v\in[I_t]}}^{i_M} \left(\prod_{n=1}^{N} 1_{\text{SNR}_n^r > \Delta} \sum_{n=1}^{N} \text{SNR}_n^r \right) \tag{3.4}$$

The performance metric for the MAB algorithm is regret, which compares the SNR degradation due to sub-optimal block selection, and it is given as

$$R = TX^{r_*} - \mathbb{E}\left[\sum_{\substack{r=1 \\ r\in[R]}}^{R} X^r S^r \right] \tag{3.5}$$

where T is the total number of time slots, S^r is the number of times the r^{th} block is selected by Base Station (BS) and r_* is the optimal block which offers highest SNR. The expectation is with respect to the random number of selections of the block S^r. Thus, the regret can be minimized by selecting r_*, i.e., the combination of M sub-blocks with the highest SNR as often as possible in a given horizon of size T.

3.4.2 Limitations of Existing MAB Framework

In this chapter, we limit our discussion to the upper confidence bound-based (UCB) MAB algorithm (74) and provide regret bounds. The proposed idea can be easily extended to other MAB algorithms, such as UCB variants and Thompson Sampling. In the UCB algorithm, each block is selected once in the beginning. After that, in each time slot ($t > R$), the block that offers the highest UCB factor is selected. The UCB factor of the block, r, in time slot t is given as

$$UCB^r(t) = \frac{\hat{X}^r(t)}{S^r(t)} + 2\sqrt{\frac{\log(t)}{S^r(t)}} \tag{3.6}$$

where $\hat{X}^r(t)$ denotes the accumulated SNR obtained over $S^r(t)$ time slots during which UCB selects the block, r. The expected regret of the UCB scales as $\mathcal{O}(\sum_{r\in[R]\setminus r_*} \frac{\log T}{\Delta_r})$ (107) where $\Delta_r = X^{r_*} - X^r$ for all $r \neq r_*$. For fixed T, the distribution-independent bounds are of the order

$\sqrt{RT}$ (107). It is evident that the UCB suffers from high exploration time, especially when R is large.

3.4.3 Proposed MAB Algorithms

In the Multi-RIS-aided system with N receivers, the number of RIS blocks that offer reasonably good SNRs for all receivers simultaneously is expected to be limited, and hence, we can safely assume that $\hat{R} \leq R$ where $\hat{R}$ are good RIS blocks. In such a case, we can reduce the distribution-independent upper bound to $\sqrt{\hat{R}T}$ and develop the algorithms that satisfy this bound. In the proposed algorithm, inspired from (108), we focus on quickly identifying and exploring these $\hat{R}$ blocks, and it is referred to as the focused exploration UCB (FEUCB) algorithm. As shown in Algorithm 1, all the blocks are selected once initially (Lines 4-5). In the rest of the time slots, we check whether at least $\hat{R}$ blocks have been explored sufficiently till time t (Line 9). If not, we will use the conventional UCB algorithm with R blocks (Line 12). Otherwise, we restrict the UCB only to the selected $\hat{R}$ blocks that have been explored sufficiently. The set of blocks that are sufficiently explored is given as

$$\mathscr{E}(t) := \left\{ r \in [R] \middle| \hat{X}^r(t) \geq 2\sqrt{\frac{\log(t)}{S^r(t)}} \right\} \tag{3.7}$$

The FEUCB algorithm aims to avoid selecting RIS blocks with poor SNR into the $\hat{R}$ blocks. This is done by ensuring that each of the selected $\hat{R}$ blocks has been sampled a sufficient number of times, as shown in Eq. 4.3. Using Theorem 1, we show that the proposed FEUCB algorithm satisfies logarithm regret, and the scaling factor depends on $\hat{R}$ instead of R, thereby outperforming the UCB algorithm.

Theorem 1 *The FEUCB algorithm guarantees that the expected regret is upper-bounded as*

$$\mathbb{E}[R(T)] \leq \log(T) \sum_{r \in [\hat{R}] \setminus r_*} \frac{1}{\Delta_r} \tag{3.8}$$

Algorithm 1 FEUCB

1: **Input:** $R, \hat{R}, T$
2: **Initialize:** $\hat{X}^r(t) \leftarrow 0$ and $S^r \leftarrow 0$ for all r
3: **for** $t = 1, 2 \ldots T$, **do**
4: **if** $t \leq R$ **then**
5: Select block, $I_t = t$.
6: **else**
7: $UCB^r(t) \leftarrow 0$ for all r
8: Compute $\mathscr{E}(t)$ using Eq. (4.3).
9: **if** $|\mathscr{E}(t)| \geq \hat{R}$ **then**
10: $\forall r \in [\mathscr{E}(t)]$: Update $UCB^r(t)$ as given in Eq. (4.1)
11: **else**
12: $\forall r \in [R]$ Update $UCB^r(t)$ as given in Eq. (4.1)
13: **end if**
14: Select block, $I_t = _{r \in [R]}\, UCB^r(t)$
15: **end if**
16: Transmitter configures RIS block, I_t and transmits a data frame.
17: Each receiver observes instantaneous normalized SNR, and communicates to the transmitter, which calculates X^{I_t} using Eq. 3.4.
18: $S^{I_t} \leftarrow S^{I_t} + 1$ and $\hat{X}^r(t) \leftarrow \hat{X}^r(t) + X^{I_t}$.
19: **end for**

Proof: We define three events: 1) $A_r(t)$: An event of good RIS block, r, being selected during the UCB phase when $\hat{R}$ blocks are sampled sufficiently (Line 10), 2) $B_r(t)$: An event of good RIS block, r, being selected during the UCB phase when $\hat{R}$ blocks are not sampled sufficiently (Line 12), and 3) $C_r(t)$: An event of poor RIS block, r, being selected. Then,

$$\mathbb{E}[R(T)] = 1 + \sum_{r \in [\hat{R}] \setminus r_*} \Delta_r \mathbb{E}\left[\sum_{t=1}^{T} 1_{A_r(t)}\right] + \sum_{r \in [\hat{R}] \setminus r_*} \Delta_r \mathbb{E}\left[\sum_{t=1}^{T} 1_{B_r(t)}\right] + \sum_{r \notin [\hat{R}]} \Delta_r \mathbb{E}\left[\sum_{t=1}^{T} 1_{C_r(t)}\right] \quad (3.9)$$

The first term is during the initialization phase when each block is selected once. Using Lemma 1-3, we can independently upper-bound each of the three terms.

$$\mathbb{E}[R(T)] = 1 + \sum_{r \in [\hat{R}] \setminus r_*} \left(\frac{16\log(T) + 8}{\Delta_r} + 3 \right) + \frac{R\Delta_{\hat{R}}\pi^2}{6} + \sum_{r \notin [\hat{R}]} \frac{\pi^2 \Delta_r}{6} \quad (3.10)$$

Since all except the second terms are independent of T, the second term dominates the regret. This concludes the proof.

Lemma 1 *The regret due to event $A_r(t)$ is given as:*

$$\mathbb{E}\left[\sum_{t=1}^{T} 1_{A_r(t)} \right] \leq \sum_{r \in [\hat{R}] \setminus r_*} \left(\frac{16\log(T) + 8}{\Delta_r^2} + 3 \right) \quad (3.11)$$

Proof: The proof is directly based on the UCB algorithm regret analysis in (74).

Lemma 2 *The regret due to event $B_r(t)$ is given as:*

$$\sum_{r \in [\hat{R}] \setminus r_*} \Delta_r \, \mathbb{E}\left[\sum_{t=1}^{T} 1_{B_r(t)} \right] \leq \frac{R\Delta_{\hat{R}}\pi^2}{6} \quad (3.12)$$

Proof: The FEUCB algorithm uses the UCB algorithm over all RIS blocks when good $\hat{R}$ blocks are not sampled sufficiently. Since UCB is based on an exploration-exploitation trade-off, the probability of this happening for the entire horizon is finite. Thus, the probability that a good RIS block is selected when good $\hat{R}$ blocks are not sufficiently sampled is finite.Assuming $\Delta_{\hat{R}} = max_{r \in [\hat{R}] \setminus r_*} \Delta_r$, we have

$$\sum_{r \in [\hat{R}] \setminus r_*} \Delta_r \, \mathbb{E}\left[\sum_{t=1}^{T} 1_{B_r(t)} \right] \leq \Delta_{\hat{R}} \sum_{t=1}^{T} \mathbb{E}\left[\sum_{r \in [\hat{R}] \setminus r_*} 1_{B_r(t)} \right]$$

Then,

$$\mathbb{E}\left[\sum_{r\in[\hat{R}]} 1_{B_r(t)}\right] \leq \sum_{r\notin\hat{R}} \mathbb{P}\left[\hat{X}^r(S^r(t)) \geq 2\sqrt{\frac{\log(t)}{S^r(t)}}\right]$$

Using the Chernoff bound, we have

$$\begin{aligned}\mathbb{E}\left[\sum_{r\in[\hat{R}]} 1_{B_r(t)}\right] \leq \sum_{r\notin\hat{R}} \mathbb{P}\left[\hat{X}^r(S^r(t)) - X^r \geq 2\sqrt{\frac{\log(t)}{S^r(t)}}\right] \\ \leq \sum_{r\notin\hat{R}} t^{-2}\end{aligned}$$

Thus,

$$\sum_{r\in[\hat{R}]\setminus r_*} \Delta_r \mathbb{E}\left[\sum_{t=1}^{T} 1_{B_r(t)}\right] \leq \Delta_{\hat{R}} \sum_{t=1}^{T}\sum_{r\notin\hat{R}} t^{-2} \leq \frac{R\Delta_{\hat{R}}\pi^2}{6}$$

Lemma 3 *The regret due to event $C_r(t)$ is given as:*

$$\mathbb{E}\left[\sum_{t=1}^{T} 1_{C_r(t)}\right] \leq \frac{\pi^2}{6} \tag{3.13}$$

Proof: Since the poor RIS, $r \notin \hat{R}$, are not sufficiently explored by the algorithm, their regret is bounded. Let $r \notin \hat{R}$. Then, $1_{C_r(t)} = 0$ for $t \leq R$. Then,

$$\mathbb{E}\left[\sum_{t=1}^{T} 1_{C_r(t)}\right] = \mathbb{E}\left[\sum_{t=R}^{T} 1_{C_r(t)}\right] \leq \mathbb{E}\left[\sum_{u=1}^{\infty} 1_{\hat{X}^r(S^r(u))\geq 2\sqrt{\frac{\log(u)}{S^r(u)}}}\right]$$

Using the Chernoff bound, we have

$$\begin{aligned}
\mathbb{E}\left[\sum_{u=1}^{\infty} 1_{\hat{X}^r(S^r(u))\geq 2\sqrt{\frac{\log(u)}{S^r(u)}}}\right] &= \sum_{u=1}^{\infty}\mathbb{P}\left[\hat{X}^r(S^r(u)) \geq 2\sqrt{\frac{\log(u)}{S^r(u)}}\right] \\
&\leq \sum_{u=1}^{\infty}\mathbb{P}\left[\hat{X}^r(S^r(u)) - X^r \geq 2\sqrt{\frac{\log(u)}{S^r(u)}}\right] \\
&\leq \sum_{u=1}^{\infty} e^{-2\log(u)} \\
&\leq \sum_{u=1}^{\infty} u^{-2} = \frac{\pi^2}{6}
\end{aligned}$$

3.4.4 Enhanced FEUCB Algorithm

In the FEUCB algorithm, the exploration is done over all R blocks until we have $\hat{R}$ sufficiently sampled blocks. Since R is large and few RIS blocks are good satisfying sparsity requirement, the performance of the FEUCB algorithm can be improved empirically by increasing the selection of good RIS blocks during the period where UCB is exploring all R blocks. The proposed empirical approach is based on a thresholding approach where the algorithm aims to identify all the RIS blocks that offer SNR above a certain desired threshold, SNR_d. Once such $\hat{R}$ blocks are identified, conventional UCB is used. As shown in Algorithm 2, the EFEUCB algorithm differs from the FEUCB algorithm on Lines 12-13, where we identify the blocks that offer SNR above a certain desired threshold, SNR_d. Such blocks are identified based on the threshold confidence bound (TCB), and it is given as

$$TCB^r(t) = \sqrt{S^r(t)}\left|\frac{\hat{X}^r(t)}{S^r(t)}\right| \tag{3.14}$$

In the FEUCB algorithm, we must select $\hat{R}$ judiciously. When the selection of $\hat{R}$ is not trivial or $\hat{R}$ is large, then the performance of the FEUCB algorithm may degrade. The EFEUCB addresses this problem to a certain extent by identifying good RIS blocks using the thresholding approach, and threshold parameters can be selected based on prior knowledge of wireless system deployment, the number of RIS and their locations, wireless physical parameters, etc.

Algorithm 2 EFEUCB

Input: $R, \hat{R}, T$

Initialize: $\hat{X}^r(t) \leftarrow 0$ and $S^r \leftarrow 0$ for all r

for $t = 1, 2 \ldots T$, **do**

  if $t \leq R$ **then**

    Select block, $I_t = t$.

  else

    $UCB^r(t) \leftarrow 0$ for all r

    Compute $\mathscr{E}(t)$ using Eq. (4.3).

    if $|\mathscr{E}(t)| \geq \hat{R}$ **then**

      $\forall r \in [\mathscr{E}(t)]$: Update $UCB^r(t)$ as given in Eq. (4.1)

    else

      $\forall r \in [R]$ Update $TCB^r(t)$ as given in Eq. (5.2)

      $\forall r \in [R] s.t. TCB^r > SNR_d$, Update $UCB^r(t)$ as given in Eq. (4.1)

    end if

    Select block, $I_t = _{r \in [R]} UCB^r(t)$

  end if

  Transmitter configures RIS block, I_t and transmits a data frame.

  Each receiver observes instantaneous normalized SNR, and communicates to the transmitter, which calculates X^{I_t} using Eq. 3.4.

  $S^{I_t} \leftarrow S^{I_t} + 1$ and $\hat{X}^r(t) \leftarrow \hat{X}^r(t) + X^{I_t}$.

end for

In both algorithms, M is fixed. Future works will use an online learning approach to select the M depending on channel conditions.

Sparsity-based approaches are useful in large-scale applications where only a small fraction of options yield significant rewards. The empirical thresholding approach in EFEUCB ensures that only good RIS sub-blocks that offer SNR above a certain threshold are selected often. Due to fewer blocks, the EFEUCB algorithm is expected to converge to the optimal block faster than the FEUCB algorithm. However, the close-form expression for EFEUCB regret is challenging due to the non-linear thresholding operation.

3.5 Performance Analysis

We consider time-slotted communication, where the transmitter selects the RIS block in each time slot and transmits a certain fixed number of data packets over the selected RIS block. The aim is to select the optimal RIS block as often as possible. We compare the performance of the proposed FEUCB and EFEUCB algorithms with existing learning and non-learning-based algorithms. In the learning-based approach, we consider UCB-based RIS selection approaches. In non-learning-based algorithms, we consider Exhaustive RIS-aided (ERA) and Opportunistic RIS-aided scheme(ORA) algorithms in (23). ORA needs prior knowledge of RIS statistics since it selects the optimal RIS block in each time slot, while ERA selects all RIS simultaneously in each time slot. In addition, we consider a simple random selection approach.

We consider the number of receivers, $N = \{1,2,3,4\}$, and the number of RIS, $K = \{5,20\}$. Each RIS is divided into $L = 5$ sub-blocks. The number of elements in each sub-block is $Z = [25,35,45,55,65]$. In addition, we consider two more cases by varying the size of the sub-blocks of each RIS. We consider $K = \{5,10\}$, with $L = \{4,3,5,3,4\}$ and $L = \{4,3,5,3,4,2,6,5,2,3\}$ respectively. The number of elements in this case is also randomly chosen between 25 and 75. The horizon size, T, is between 10000 and 500000. Each result is averaged over 15 independent experiments over the selected horizon size. In each experiment, the positions of the transmitter, RIS, and receivers are selected randomly. Unless otherwise specified, the simulation parameters are set to the values mentioned in Table 5.1. The equivalent noise power at the receiver is given as:$\sigma_n^2 = N_0 + 10log(BW) + NoiseFigure[dBm]$ where N_0 is the thermal noise power density.

The performance metrics used for comparison are regret (Eq. 4.2), outage probability, ergodic capacity, and energy efficiency. The lower regret indicates a higher average SNR at the receivers, guaranteeing reliable communication. The outage probability is defined as the probability that the SNR of the system falls below a certain threshold. The achievable SNR can be affected by various factors, including RIS selection, channel fading, and receiver noise. Mathematically it can be expressed as

$$\mathrm{P}_{out} = Pr(\mathrm{SNR}_n^r < \mathrm{SNR}_{th}) \tag{3.15}$$

where SNR_{th}, is the minimum threshold SNR. The outage probability decreases with increased transmit power and the number of RIS elements.

Ergodic capacity refers to the maximum average rate at which the data can be transmitted reliably over a selected channel. Essentially, it represents the average capacity of a channel when both the sender and the recipient possess information on the channel state, and the channel varies randomly over time. Mathematically it can be expressed as

$$\text{EC} = \mathbb{E}[\log_2(1 + \text{SNR}_n^r)] \tag{3.16}$$

The energy efficiency gauges the amount of data that can be sent for every unit of power used and can be expressed as:

$$\eta = \frac{BW \times \text{Average Achievable Rate}}{\text{P}_C} \tag{3.17}$$

where the consumed power, P_C, is the sum of the circuit dissipated power at the transmitter, N receivers, and each element of the selected RIS block.

3.5.1 Regret Performance Analysis

We begin by comparing various algorithms for $K = 5$ (fewer number of RIS) and $K = 20$ (larger number of RIS), respectively, with $L = 5$ and $N = 4$. In Fig. 3.2(a) and Fig. 3.2(b), we compare the cumulative regret for the horizon size of 10000. As expected, random RIS selection incurs high regret due to the frequent selection of the non-optimal RIS blocks. The regret of the ORA scheme is zero as it has prior knowledge of the optimal RIS block. We have not included the ERA scheme for regret comparison since ERA selects all RIS blocks, while regret calculation requires selecting a single RIS block in each slot. Among learning-based UCB, FEUCB, and EFEUCB algorithms, the regret of the EFEUCB algorithm is lowest due to improved exploration, resulting in fewer selections of non-optimal RIS blocks. It can be observed that the regret increases with the increase in K due to a large number of candidate RIS blocks, resulting in higher exploration time. However, the difference between the regret of the

Table 3.2: Parameters

Parameters	**Values**
Location of Transmitter	(0,0)
Location of Receivers	Random
Location of RISs	Random
Transmit Power [dBm]	[-40,40]
Amplitude reflection coef., A	1(109)
Number of RIS	5 and 20
Number of Sub-blocks per RIS	5
Number of Elements in sub-blocks	[25,35,45,55,65]
Threshold SNR (SNR_{th})	7dBm
Bandwidth	10 MHz (109)
Noise Figure	10 (109)
Thermal noise power density,(N_0)	-174 (109)
Antenna gain[dB]	5 (109)
Carrier Frequency [GHz]	3 (109)
Circuit Dissipated Power in RIS [mW]	7.8 (110)
Circuit Dissipated Power in TX and RX[mW]	10 (110)

FEUCB/EFEUCB and UCB increases with the increase in K, which validates the superiority of the proposed focused exploration approach. The difference between FEUCB and EFEUCB increases as K increases, highlighting the impact of empirical thresholding-based enhancement in focused exploration. We have also compared the experimental regret and its desired upper bound, i.e., worst case performance. It is evident that the experimental results follow a similar trend as the theoretical results, and as expected, experimental results offer lower regret than the corresponding upper bound.

Next, we compare the regret performance for the case when all the RIS units are not identical in terms of element size and the number of sub-blocks. As shown in Fig. 3.3 (a) and (b), the performance of the proposed algorithms is superior to the existing UCB algorithm, and EFEUCB offers improved performance as K increases.

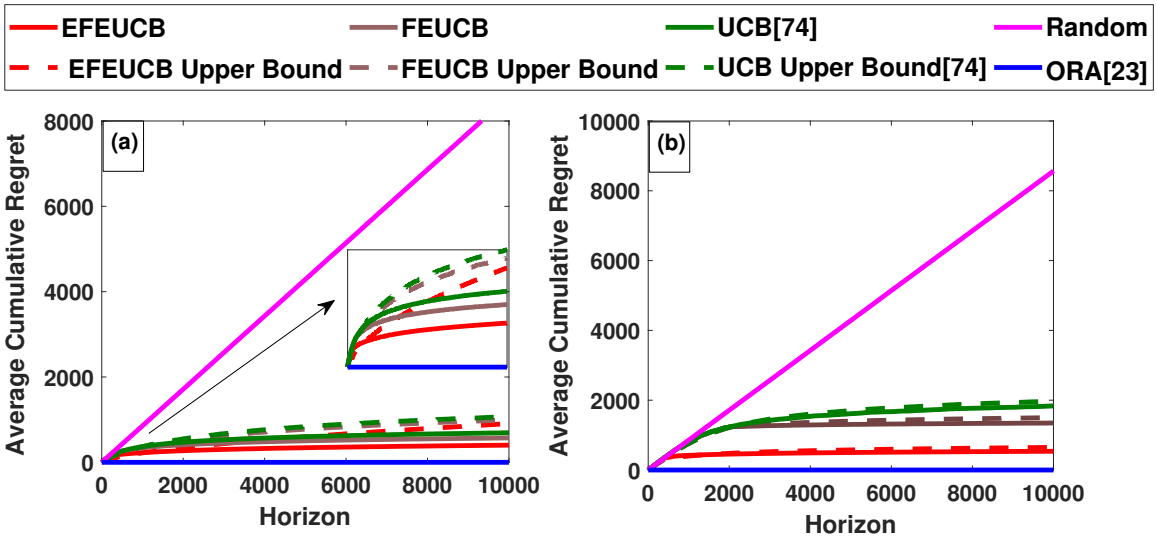

Figure 3.2: Comparison of Average Cumulative Regret for different algorithms for (a) $K = 5, L = 5$ and (b) $K = 20, L = 5$

3.5.2 Comparison of Achievable Rate, Outage Probability, and Ergodic Capacity

Next, we study the effect of the transmit power and total consumed power on the achievable rate in bits per second per Hertz (b/s/Hz). As shown in Fig. 3.4 and Fig. 3.5, different algorithms need different transmit power to achieve a given rate. For example, ERA (23) demands the low-

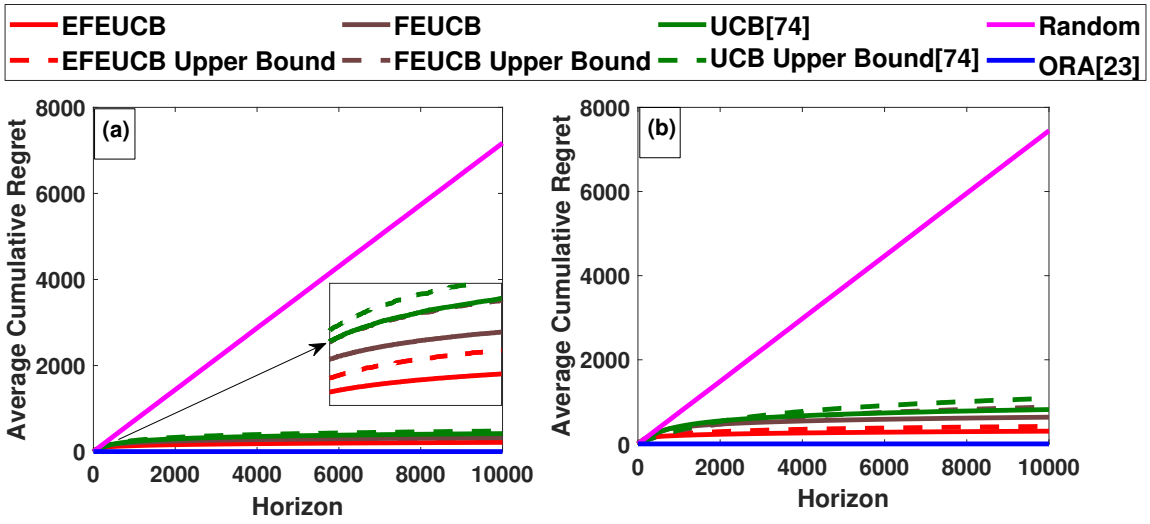

Figure 3.3: Comparison of Average Cumulative Regret for different algorithms for (a) $K = 5, L = [4,3,5,3,4]$ and (b) $K = 10, L = [4,3,5,3,4,2,6,5,2,3]$

est transmit power due to the use of all RIS, which in turn helps to improve the average SNR at the receiver. The random approach needs the highest transmit power due to the frequent selection of sub-optimal RIS. All learning-based approaches need a similar transmit power as that of ORA (23), validating the successful learning and frequent selection of optimal RIS. EFEUCB offers better performance than FEUCB, which in turn offers better performance than UCB (74). As K increases, ERA performance improves due to increased SNR at the receiver. Furthermore, the difference between the performance of ORA and learning approaches is significant for large K, which is the penalty paid to learn optimal RIS block. As expected, focused exploration leads to a smaller penalty for EFEUCB and FEUCB than UCB.

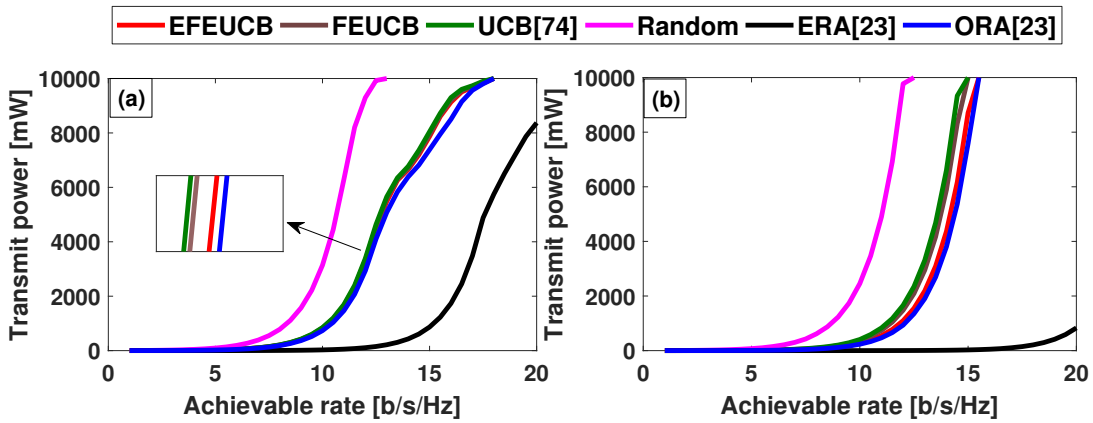

Figure 3.4: Comparison of Transmit Power and Achievable Rate for different algorithms for (a) $K = 5, L = 5$ and (b) $K = 20, L = 5$

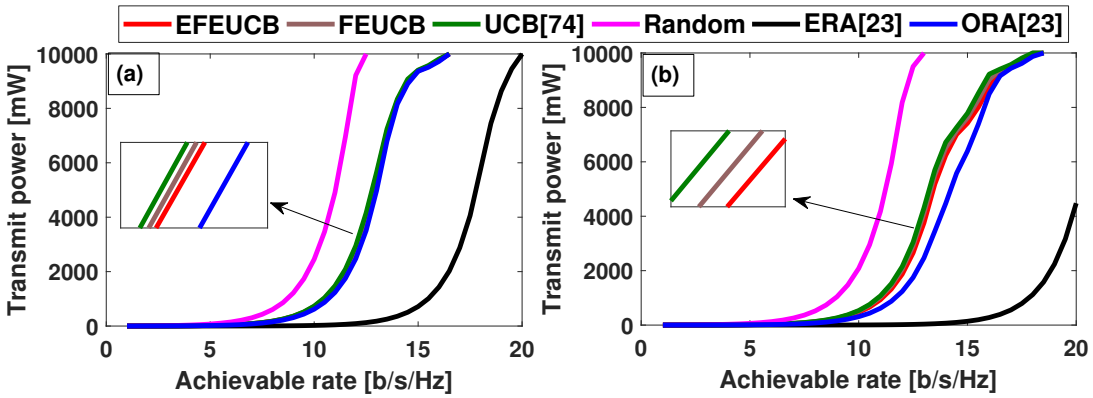

Figure 3.5: Comparison of Transmit Power and Achievable Rate for different algorithms for (a) $K = 5, L = [4,3,5,3,4]$ and (b) $K = 10, L = [4,3,5,3,4,2,6,5,2,3]$

Next, we consider total consumed power, P_C, instead of transmit power in Fig. 3.6 and Fig. 3.7. As expected, ERA offers poor performance, and the performance degrades substantially with the increase in the number of RIS. This confirms the practical challenges in the widely used ERA scheme, especially when K is large. Using the proposed learning approaches, consumed power can be reduced significantly by the appropriate selection of M.

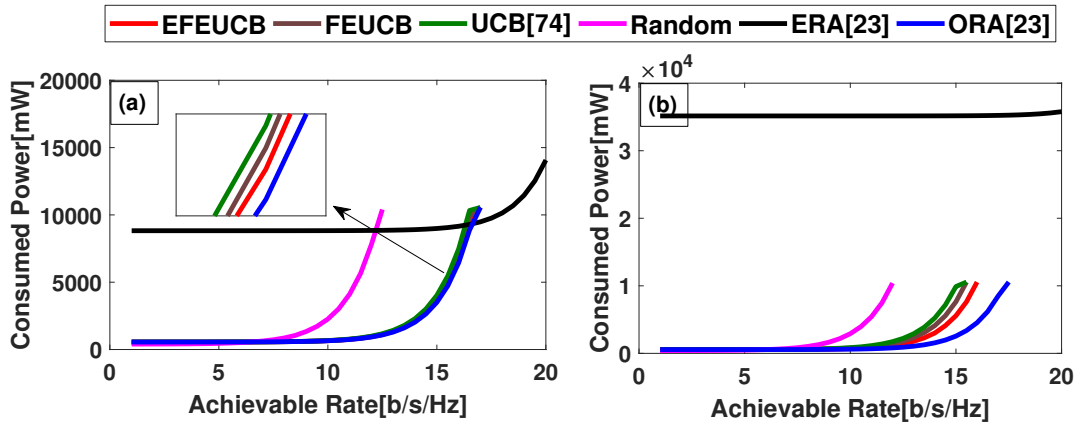

Figure 3.6: Comparison of Total Consumed Power and Achievable Rate for different algorithms for (a) $K = 5, L = 5$ and (b) $K = 20, L = 5$

Next, we compare the outage probability in Fig. 3.8(a) and (c) for different values of the transmit power. To avoid repetition of plots and discussion, the discussion is limited to $K = \{5,20\}$ with $L = 5$. As expected, the outage probability decreases with the increase in the transmit power. The ERA scheme outperform others since it selects all RIS blocks, improving SNR at all receivers. This is because the number of sub-blocks L in the ERA scheme is more as compared to the ORA scheme; hence, the total number of elements Z are more in the case of

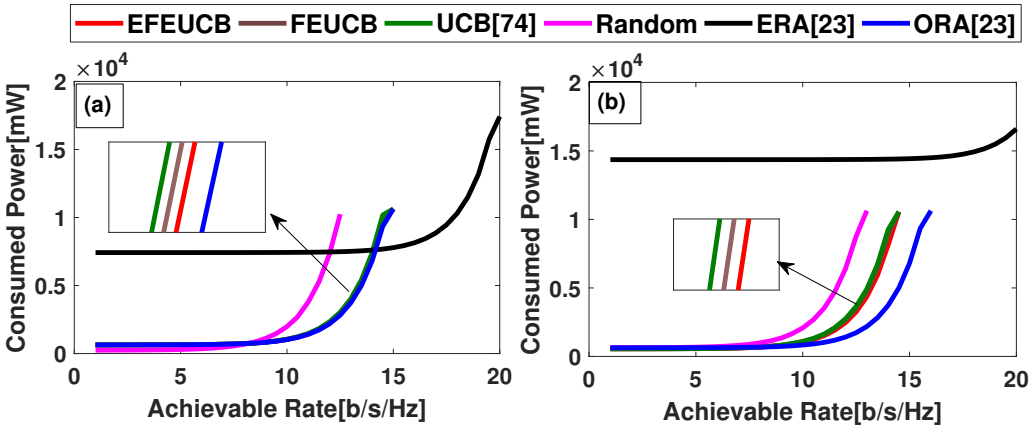

Figure 3.7: Comparison of Total Consumed Power and Achievable Rate for different algorithms for (a) $K = 5, L = [4,3,5,3,4]$ and (b) $K = 10, L = [4,3,5,3,4,2,6,5,2,3]$

ERA scheme (23), because of which the outage probability of ERA decreases faster than that of ORA with a slight increase in transmit power. As expected, the outage probability of learning-based approaches is close to that of ORA since learning-based approaches eventually select the optimal RIS block after the initial exploration-exploitation trade-off. The divergence at $P_{out} = 10^{-2}$ is because of exploring other blocks while learning the optimal block. In Fig. 3.8(c), with a total of 20 RIS, ERA offers further improvement in performance as its outage probability starts decreasing at lower transmit power than in Fig. 3.8(a) with 5 RIS.

Similarly, in Fig. 3.8(b) and (d), we compare the Ergodic capacity for different values of the transmit power. As expected, the ERA scheme offers the highest ergodic capacity due to the selection of all RIS elements, which allows the highest throughput for multiple users. The proposed EFEUCB offers a better outage probability among learning-based approaches. As K increases, exploration time increases; hence, the difference between the performance of ORA and learning-based approaches increases. This penalty is paid due to no prior knowledge of channel conditions and positions of RIS and receivers in the learning-based approach. Note that the difference between the performance of ERA and ORA increases with the increase in K, which validates the need for multi-RIS-aided wireless networks. In Fig. 3.9, we compare the outage probability and ergodic capacity for different values of consumed power. As expected, the performance of the ERA scheme is significantly poor due to the selection of all RIS, which results in high consumed power. Thus, the proposed approach addresses ERA's high consumed power drawback by learning and selecting the optimal RIS block instead of all RIS. It also outperforms UCB via focused exploration, especially when K is large.

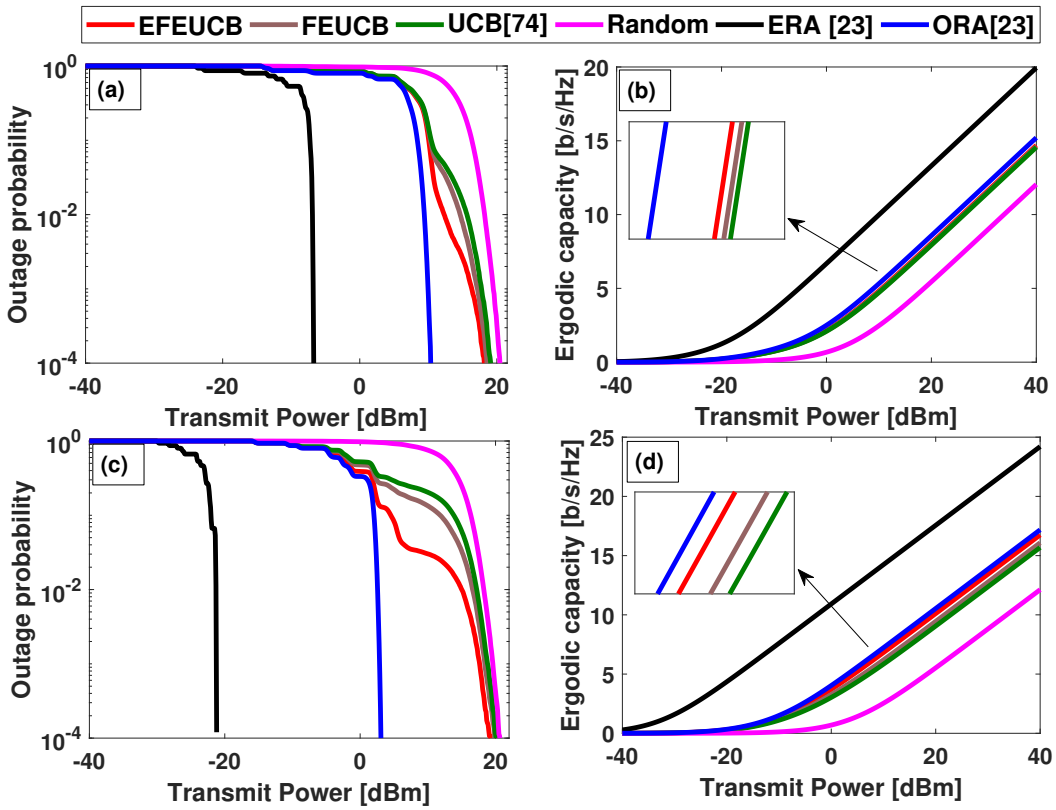

Figure 3.8: Comparison of Outage probability for (a) $K = 5$, and (c) $K = 20$, and Ergodic capacity for (b) $K = 5$ and (d) $K = 20$ with $L = 5$ for different values of transmit power.

3.5.3 Energy Efficiency Comparison

Next, we compare the energy efficiency of various algorithms for different achievable rates in Fig. 3.10(a) and (b). Here, we increase the transmit power to achieve the desired rate. As expected, the energy efficiency of the ERA scheme is abysmal due to the selection of all RIS elements. In contrast, the energy efficiency of ORA is highest due to the selection of optimal RIS, which results in lower consumed power. The energy efficiency of the learning-based scheme is significantly better than the ERA scheme and close to that of the ORA scheme. The difference between the energy efficiency of learning-based approaches and ORA is due to the higher transmit power needed to meet the desired achievable rate whenever sub-optimal RIS is selected due to the exploration-exploitation trade-off. As the number of RIS increases, the energy efficiency of the proposed FEUCB and EFEUCB is better than that of UCB and random approaches, validating the superiority of the proposed focused exploration approach. The energy efficiency of the ORA and proposed learning-based algorithms decreases when the desired achievable rate goes beyond a certain value, as shown in Fig. 3.10. This is due to a significant increase in the

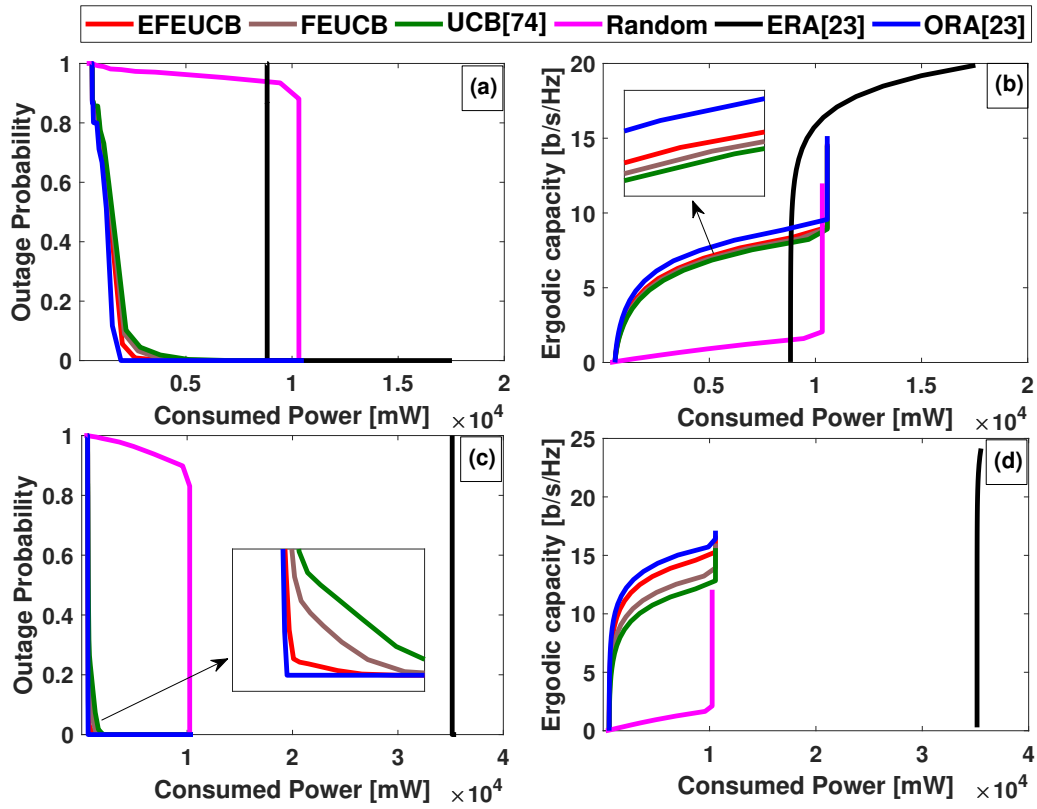

Figure 3.9: Comparison of Outage probability for (a) $K = 5$, and (c) $K = 20$, and Ergodic capacity for (b) $K = 5$ and (d) $K = 20$ with $L = 5$ for different values of consumed power.

desired transmit power, as shown in Fig. 3.4.

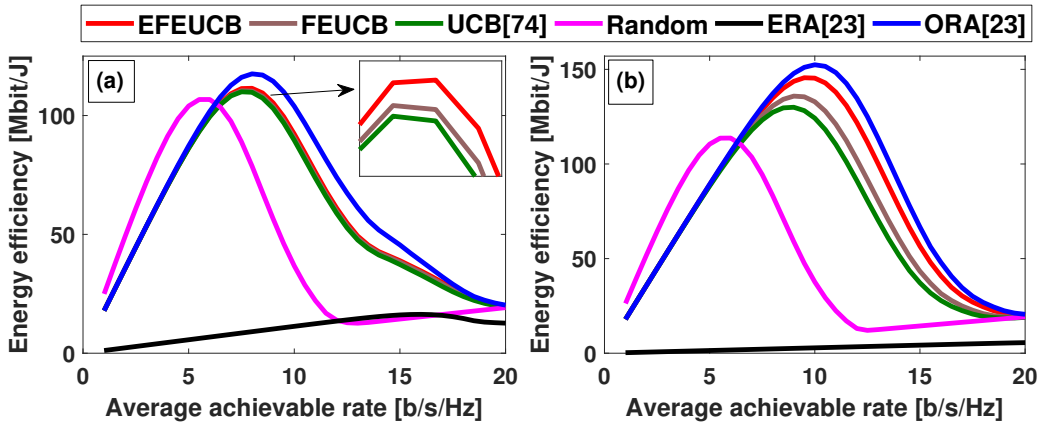

Figure 3.10: Comparison of energy efficiency and average achievable rate for different algorithms for (a) $K = 5, L = 5$ and (b) $K = 20, L = 5$

3.5.4 Effect of Horizon Size, T

Compared to ORA schemes, learning-based approaches do not have prior knowledge of the optimal RIS; hence, they need exploration time to learn and identify the optimal RIS block. In Fig. 3.11, we compare the effect of horizon size on the performance of learning algorithms. It can be observed that the difference between the learning and ORA schemes decreases with the increase in horizon time for a given K and L. This is because the penalty for selecting non-optimal RIS during exploration time becomes insignificant as horizon time increases. This also validates the functional accuracy of the proposed learning-based approaches to identify the optimal RIS accurately and select it as many times as possible over the horizon.

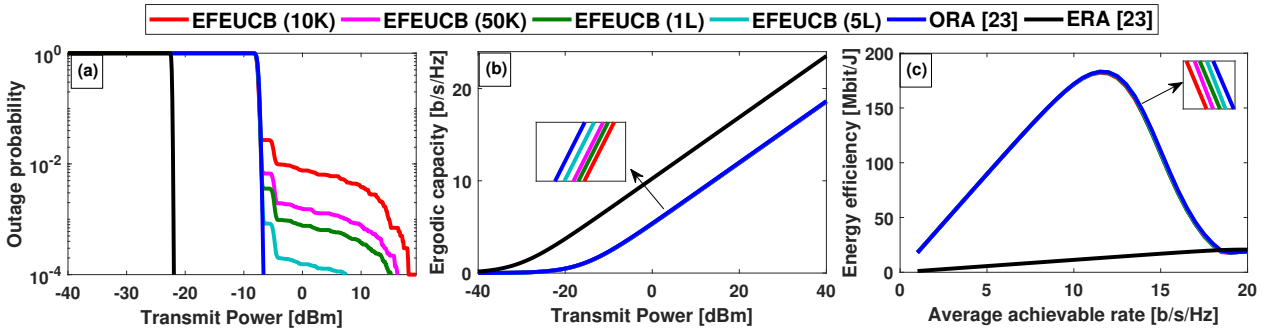

Figure 3.11: Comparison of Outage Probability, Ergodic Capacity, and Energy Efficiency for different horizon time.

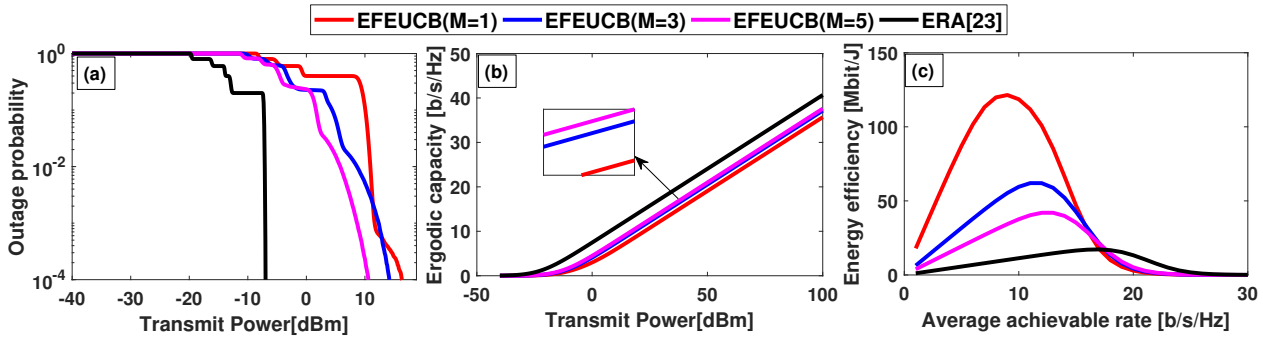

Figure 3.12: Comparison of Outage Probability, Ergodic Capacity, and Energy Efficiency when different block sizes, M.

3.5.5 Effect of Block Size, M

Each RIS block consists of M number of sub-blocks, and $M = 1$ corresponds to a single-RIS-aided system. We compare the effect of M on the performance of wireless systems in Fig. 3.12 for $M = \{1, 3, 5\}$. It can be observed that the outage probability and ergodic capacity of the EFEUCB improved with the increase in M and approaches that of ERA as M increases at the cost of poor energy efficiency. Thus, M should be carefully selected to achieve the desired trade-off between performance and energy efficiency.

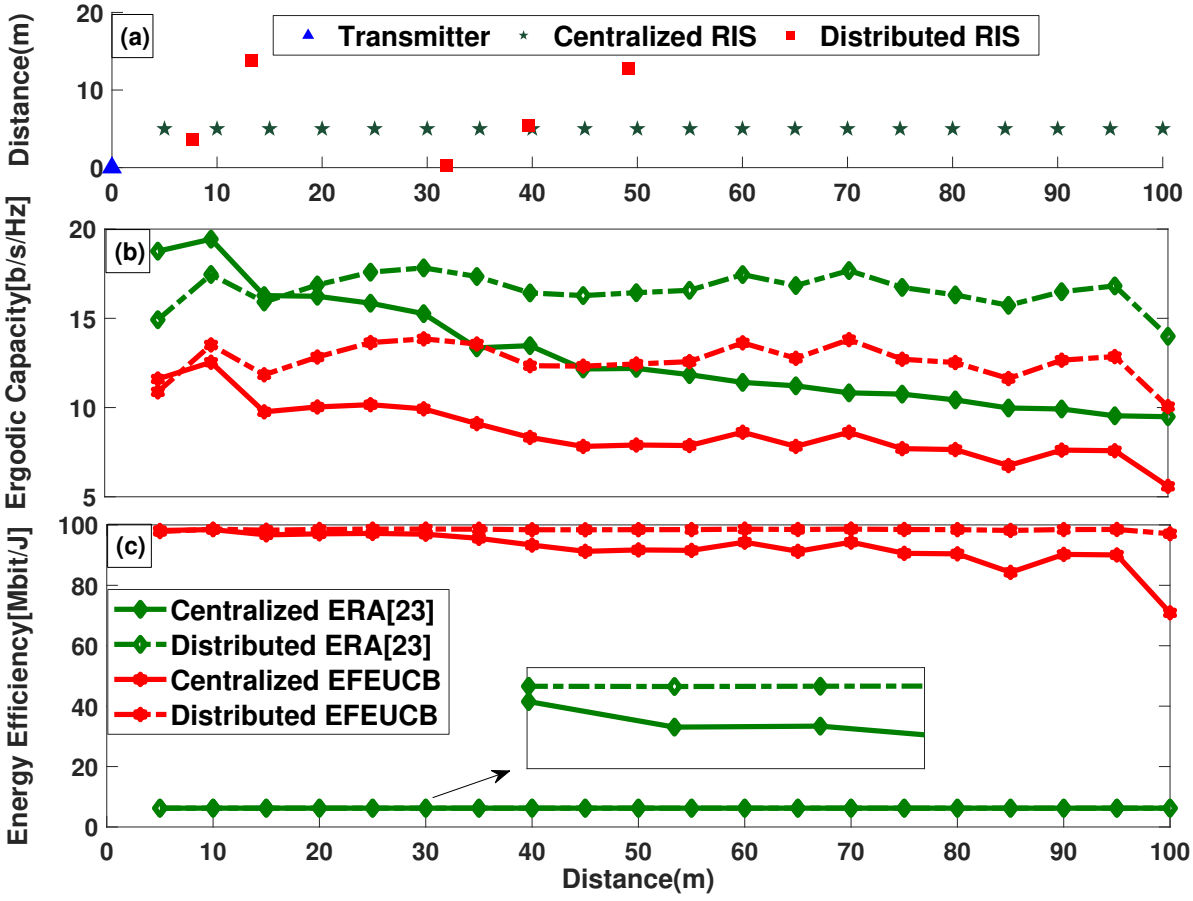

Figure 3.13: Comparison between centralized and distributed RIS approaches: (a) Fixed locations of the five distributed RIS and different locations of centralized RIS. The locations of four receivers is selected randomly in each experiment, (b) Comparison of ergodic capacity for different position of centralized RIS, and (b) Comparison of energy efficiency for different position of centralized RIS. The transmit power and achievable rate for energy efficiency is fixed at 20 dBm and 10 b/s/Hz, respectively.

3.5.6 Centralized vs. Distributed RIS Approach

Next, we compare the performance of the centralized and distributed RIS approaches in Fig. 3.13. In a centralized approach, we use a single RIS with the same number of elements as the total elements of K RIS in a distributed approach. We randomly select the position of the receivers in each experiment and average the results over multiple experiments. In Fig. 3.13, we compare the Ergodic capacity of the ERA and EFEUCB algorithms for different positions of the centralized RIS between transmitter and receiver. As expected, the distributed approach offers higher Ergodic capacity than the centralized approach in both cases, thereby validating the need for a multi-RIS-aided distributed system. In Fig. 3.13, we also compare the energy efficiency of the centralized and distributed RIS approaches for ERA and EFEUCB algorithms. The distributed approach offers higher efficiency than the centralized one due to higher ergodic capacity.

Next, we make an in-depth comparison of the ergodic capacity of centralized and distributed approaches to analyze the effect of a number of receivers and their locations. In Table 3.3, we consider three scenarios with 1, 2, and 4 receivers, respectively, along with the specific locations of the receivers in each case. The location of 5 distributed RIS is fixed in all

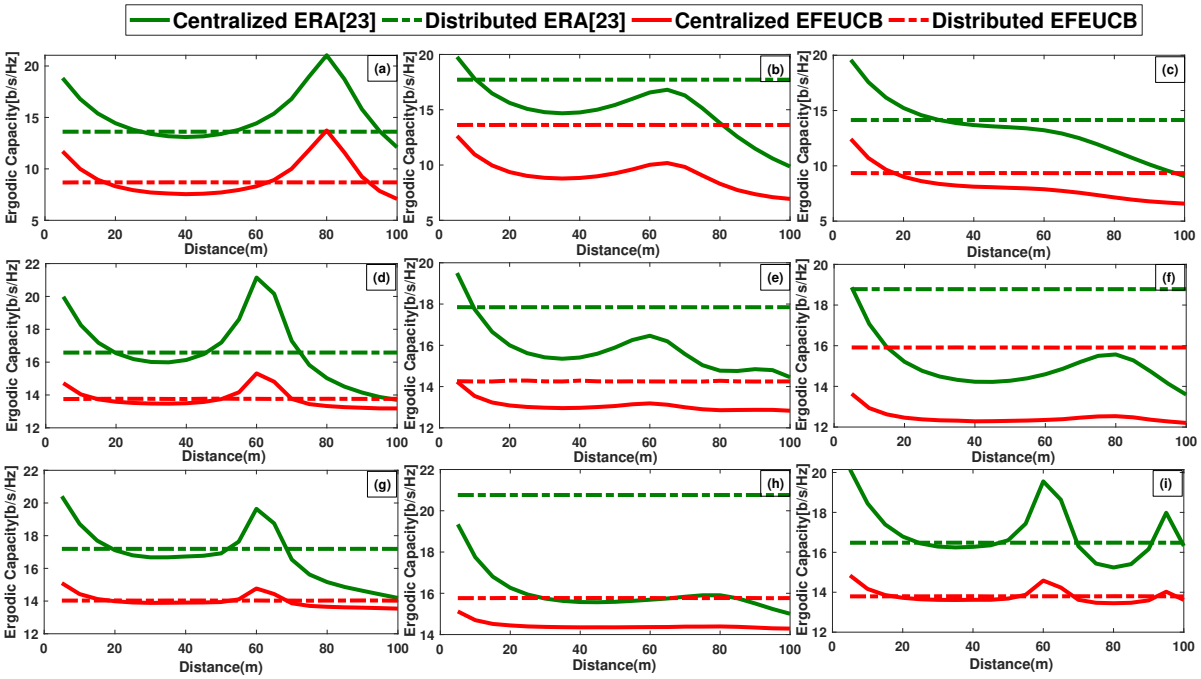

Figure 3.14: Comparative results for Ergodic Capacity at Transmit Power 20dBm, with different scenario (a)-(c) Scenario 1 (d)-(f) Scenario 2 and (g)-(i) Scenario 3.

scenarios. In Fig. 3.14 (a)-(c), we compare the ergodic capacity for scenario 1 with a single receiver. The centralized RIS offers better performance than the distributed RIS in Fig. 3.14 (a) as the receiver is near to the centralized RIS compared to any of the distributed RIS. On the other hand, the distributed approach significantly outperforms the centralized approach in Fig. 3.14 (b)-(c) even though centralized RIS can be moved at any position between transmitter and receiver while the position of distributed RIS is fixed in advance. Similar observations can be made from various results presented in Fig. 3.14 (d)-(i). These results validate the superiority of the distributed approach in Fig. 3.13, where receiver positions are selected randomly. Since the positions of receivers are not fixed in real networks, the distributed approach is more effective and practical.

Table 3.3: Parameters for Experiments in Fig. 3.14

<table>
<tr><th>Scenario</th><th>No. of RX</th><th>Figure</th><th>Location of Receiver(s)</th><th>Location of RIS</th></tr>
<tr><td rowspan="3">1</td><td rowspan="3">1</td><td>Fig.3.14 (a)</td><td>(80.35,0.65)</td><td rowspan="10">Distributed:
RIS1 (16.95,3.58),
RIS2 (94.12,24.38),
RIS3 (81.53,21.45),
RIS4 (57.10,10.55),
RIS5 (63.74,12.33)

Centralized:
Moving along
X axis, and
Y coordinate is
fixed at 5.</td></tr>
<tr><td>Fig. 3.14 (b)</td><td>(66.71,16.95)</td></tr>
<tr><td>Fig. 3.14 (c)</td><td>(65.22,31.41)</td></tr>
<tr><td rowspan="3">2</td><td rowspan="3">2</td><td>Fig. 3.14 (d)</td><td>(74.04,27.20),
(61.99,1.41)</td></tr>
<tr><td>Fig. 3.14 (e)</td><td>(94.93,17.62),
(62.25,16.73)</td></tr>
<tr><td>Fig. 3.14 (f)</td><td>(84.47,20.94),
(79.96,18.58)</td></tr>
<tr><td rowspan="3">3</td><td rowspan="3">4</td><td>Fig. 3.14 (g)</td><td>(45.15,20.15),
(74.04,21.2),
(62,1.5),
(91.33,22.13)</td></tr>
<tr><td>Fig. 3.14 (h)</td><td>(84.47,20.94),
(64.00,26.11),
(79.96,18.58),
(94.22,23.42)</td></tr>
<tr><td>Fig. 3.14 (i)</td><td>(62,1.5),
(96,1.5),
(86.96,20),
(46.37,22.50)</td></tr>
</table>

3.6 Execution Time Comparison on Edge Platforms

In the MAB algorithm deployed in wireless networks, the execution time of the algorithm should be as short as possible. In time-slotted communication, a faster MAB algorithm leads to the early selection of the RIS block, allowing the transmitter more time to transmit the actual data and achieve high throughput. In Table 3.4 and Fig. 3.18, we compare the execution

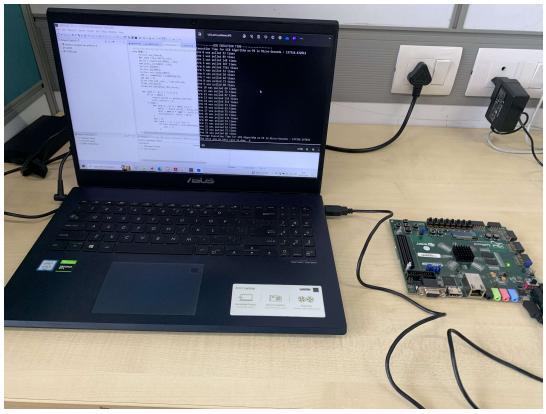

Figure 3.15: Execution Time analysis on Zedboard with ARM Cortex A9 processor working at 666 MHz

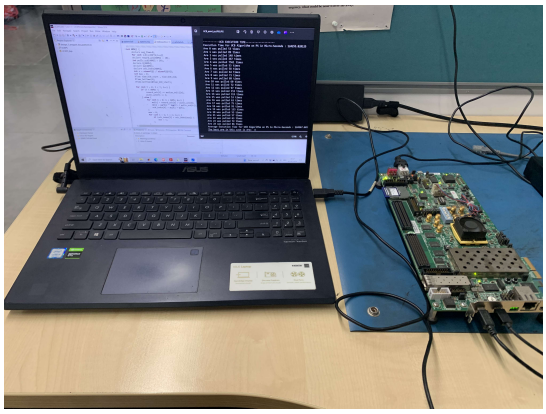

Figure 3.16: Execution Time analysis on ZCU706 board with ARM Cortex A9 processor working at 800 MHz

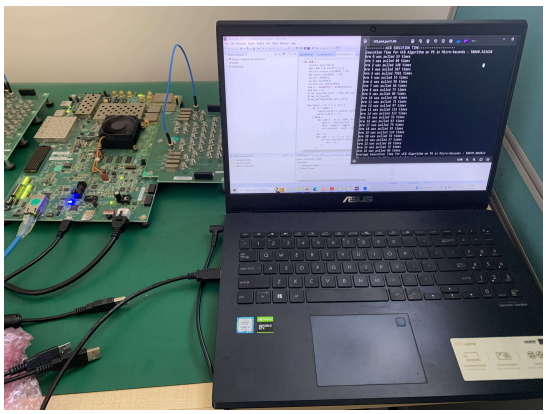

Figure 3.17: Execution Time analysis on ZCU711 board with ARM Cortex A53 processor working at 1.1 GHz

time of three MAB algorithms on three different types of processors used in Edge platforms: 1) ARM Cortex A9 at 666 MHz, 2) ARM Cortex A9 at 800 MHz, and 3) ARM Cortex A53 at 1.1 GHz. We study the effect of single instruction multiple data (SIMD) NEON co-processor and

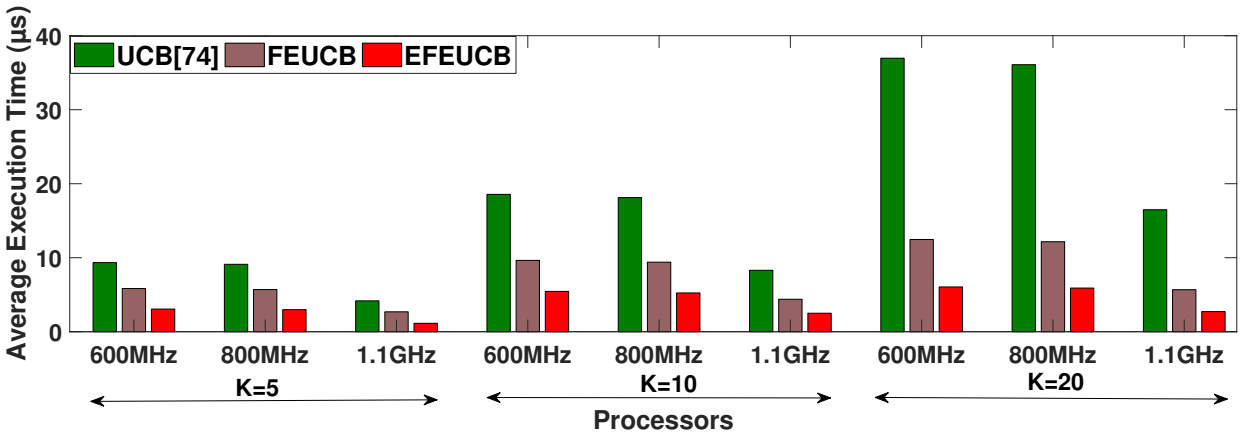

Figure 3.18: Comparison of Average Execution Time (μs) for different algorithms on different processors.

two different word lengths, single precision floating point (SPFL) and double precision floating point (DPFL), on the execution time. For $R = 25$ RIS sub-blocks, we can observe that FEUCB offers around 33-39% lower execution time than UCB while EFEUCB offers around 64-68% lower execution time than UCB. As the number of sub-blocks are increased from 25 to 50 and 100, further improvement in the performance of FEUCB and EFEUCB with respect to UCB is observed. The substantial degradation in the execution time of the UCB can be observed from Fig. 3.18. For 50 sub-blocks, FEUCB and EFEUCB offer around 46-49% and 68-71% lower execution time than UCB, respectively. This improvement is increased to 65-66% and 80-85% for 100 sub-blocks. The huge savings in execution time in proposed algorithms are due to the focused exploration approach, which reduces the number of computational arithmetic operations in each time slot. As expected, execution time with DPFL is slightly higher than SPFL, though we did not observe any degradation in performance with SPFL. Similarly, higher clock frequency and better processors lead to further reduced execution time, as shown in Table 3.4. The use of SIMD-based NEON co-processor results in a significant reduction in execution time for all the algorithms. Further reduction in execution time can be achieved using multiple cores of these processors along with other co-processors such as graphic processing unit (GPU) or field-programmable gate array (FPGA).

Table 3.4: Comparison of Execution Time in Milliseconds of the Various Algorithms on Edge Platforms

S.No.	Sub Blocks	Algorithm	Cortex A9 (666 MHz)		Cortex A9 and NEON (666 MHz)		Cortex A9 (800 MHz)		Cortex A9 and NEON (800 MHz)		Cortex A53 (1.1 GHz)		Cortex A53 and NEON (1.1 GHz)	
			SPFL	**DPFL**	**SPFL**	**DPFL**	**SPFL**	**DPFL**	**SPFL**	**DPFL**	**SPFL**	**DPFL**	**SPFL**	**DPFL**
1	25	UCB(74)	12.73	12.95	9.34	9.21	12.42	12.61	9.11	8.98	5.86	5.91	4.17	4.43
2		FEUCB	8.49 (-33%)	8.60 (-33%)	5.84 (-37%)	5.59 (-39%)	8.29 (-33%)	8.39 (-33%)	5.70 (-37%)	5.45 (-39%)	3.89 (-33%)	3.92 (-33%)	2.69 (-37%)	2.81 (-37%)
3		EFEUCB	4.49 (-65%)	4.57 (-65%)	3.06 (-67%)	2.96 (-68%)	4.35 (-65%)	4.44 (-65%)	2.98 (-67%)	2.93 (-67%)	2.10 (-64%)	2.12 (-64%)	1.41 (-66%)	1.48 (-67%)
4	50	UCB(74)	25.22	25.76	18.56	18.40	24.59	25.14	18.12	17.96	11.64	11.73	8.30	8.81
5		FEUCB	13.55 (-46%)	13.78 (-46%)	9.64 (-48%)	9.37 (-49%)	13.23 (-46%)	13.45 (-46%)	9.40 (-48%)	9.14 (-49%)	6.32 (-46%)	6.37 (-46%)	4.39 (-47%)	4.60 (-48%)
6		EFEUCB	7.84 (-69%)	7.96 (-69%)	5.45 (-70%)	5.40 (-70%)	7.65 (-69%)	7.82 (-69%)	5.24 (-71%)	5.16 (-71%)	3.76 (-68%)	3.79 (-68%)	2.50 (-70%)	2.61 (-70%)
7	100	UCB(74)	49.43	51.25	36.96	36.60	48.24	50.02	36.08	35.72	23.11	23.28	16.48	17.49
8		FEUCB	17.05 (-66%)	17.52 (-66%)	12.46 (-66%)	12.32 (-66%)	16.64 (-66%)	17.10 (-66%)	12.16 (-66%)	12.02 (-66%)	8.08 (-65%)	8.14 (-65%)	5.68 (-65%)	5.98 (-66%)
9		EFEUCB	9.12 (-82%)	9.44 (-82%)	6.05 (-84%)	6.28 (-83%)	8.88 (-82%)	9.24 (-85%)	5.90 (-84%)	5.91 (-83%)	4.67 (-80%)	4.70 (-80%)	2.72 (-83%)	2.96 (-83%)

3.7 Summary

In this chapter, novel UCB-based distributed algorithms, i.e., FEUCB and EFEUCB, have been proposed to select optimal RIS sub-block. The theoretical regret analysis and the in-depth simulation results validate the effectiveness and superiority of our proposed algorithm over the existing state-of-the-art algorithms. The execution time comparison on the edge platforms demonstrates the compute and memory efficient architecture of the proposed algorithms.

Chapter 4

High-Speed Compute-Efficient Bandit Learning for Many Arms

4.1 Overview

MAB are online machine learning algorithms that aim to identify the optimal arm without prior statistical knowledge via the exploration-exploitation trade-off. The performance metric, regret, and computational complexity of the MAB algorithms degrade with the increase in the number of arms, R. In applications such as wireless communication, radar systems, and sensor networks, R, i.e., the number of antennas, beams, bands, etc., is expected to be large. In this work, we consider focused exploration-based MAB, which outperforms conventional MAB for large R, and its mapping on various edge processors and multiprocessor system on a chip (MPSoC) via hardware-software co-design and fixed point analysis.

4.2 Introduction

Multi-armed bandit (MAB) is a subset of reinforcement or online learning algorithms that aim to identify and select the optimal arm as often as possible without prior statistical knowledge (111; 112). This is done through sequential arm selection and reward feedback cycle optimized via exploration-exploitation trade-off over a finite number of slots, i.e., horizon. Analytical traceability of MAB algorithms and the capability to learn in unknown environments without prior training have led to their usefulness and adoption in various practical applications such as circuit design, wireless networks, robotics, and radar systems (113; 114; 115; 101; 116; 117; 118). Some of these applications demand an efficient hardware realization of MAB algorithms on edge platforms such as system-on-chip (SoC) (119; 120; 121; 122) so that decisions can be taken at the edge, thereby minimizing the latency when compared to cloud-based decision making. This demands an efficient hardware realization of MAB algorithms on resource and power-constrained edge platforms.

Every experiment in the MAB consists of a series of time slots indexed by $t \in [T]$, where T is the total time horizon, and out of R arms, one arm is selected in each slot. For each arm, the reward is assumed to be drawn independently across time from distributions that are stationary and independent across arms. However, the reward distribution is unknown. The well-known upper confidence bound (UCB) algorithm selects each arm once in the beginning. After R time slots, UCB calculates the quality factor $UCB^r(t)$ for each arm as (111).

$$UCB^r(t) = \frac{X^r(t)}{S^r(t)} + 2\sqrt{\frac{\log(t)}{S^r(t)}} \tag{4.1}$$

where $X^r(t)$ and $S^r(t)$ are the total received reward and the total number of selections of an arm r till time slot t. Then, the arm I_t with the highest quality factor is selected. The performance metric is the regret, R, which is given as (111).

$$R = TX^{r_*} - \mathbb{E}\left[\sum_{r=1}^{R} X^r(T)S^r(T)\right] \tag{4.2}$$

where r_* is an optimal arm. For fixed T, the distribution-independent lower bound on regret

is of the order $\sqrt{RT}$ (107). It is evident that the UCB suffers from high exploration time and complexity, especially when R is large.

In next-generation wireless and radar applications, R, i.e., the number of antennas, beams, frequency bands, sensors, etc., are expected to be large, ranging from 10-100 (123; 115; 101; 69). Most of these applications are based on a time-slotted approach where each slot is divided into two sub-slots: 1) the First sub-slot is used for MAB-based arm selection, and 2) the Second is for applications such as data communication or radar sensing. Larger K in UCB results in a longer first sub-slot, degrading the performance of communication or radar sensing.

For such large or many arms problems, we proposed two algorithms: 1) Focused Exploration UCB (FEUCB), and 2) Enhanced FEUCB (EFEUCB) in (53). Though they offer improved regret than UCB, their hardware feasibility, regret performance and area, delay, and power analysis are critical for their usefulness in wireless and radar applications. In this chapter, we consider the algorithms to architecture mapping of FEUCB and EFEUCB on Zynq SoC via hardware-software co-design (HSCD) and fixed point (FP) analysis. Though both offer lower regret than UCB, their complexity is higher than UCB. To address this, we introduced a modified EFEUCB (mEFEUCB) that quickly reduces the number of arms from R to $\hat{R} < R$, thereby offering significant savings in complexity over FEUCB and EFEUCB. For wireless systems with intelligent reflecting surface (IRS) based applications, the proposed mEFEUCB outperforms UCB with 67% reduction in average cumulative regret, 84% reduction in execution time on edge processor, 97% reduction in execution time using Field-Programmable Gate Array (FPGA) based accelerator with UCB on processor, and 10% savings in resources over UCB for large $R = 100$. With both UCB and mEFEUCB are on FPGA, mEFEUCB is 51% faster.

4.3 FEUCB and EFEUCB Algorithms

When R is large, it is expected that there will be a limited number of arms with rewards close to that of the optimal arm. Thus, the performance of the MAB algorithm can be improved if we identify the subset of such $\hat{R}(\leq R)$ arms before invoking the MAB algorithm. Assuming this, it is possible to effectively lower the distribution-independent upper bound to $\sqrt{\hat{R}T}$. The flowchart of the proposed FEUCB and EFEUCB algorithms are given in Fig. 4.1. In both cases,

each arm is selected once at the beginning. After the first R slots, instead of computing the UCB quality factor for each arm, we identify the subset of $\hat{R}$ *good* arms $\mathscr{E}(t)$ as follows (53).

$$\mathscr{E}(t) := \left\{ r \in [R] \middle| \frac{X^r(t)}{S^r(t)} \geq 2\sqrt{\frac{\log(t)}{S^r(t)}} \right\} \tag{4.3}$$

If $\hat{R}$ *good* arms are identified, UCB is focused on these $\hat{R}$ arms only. Otherwise, all arms are

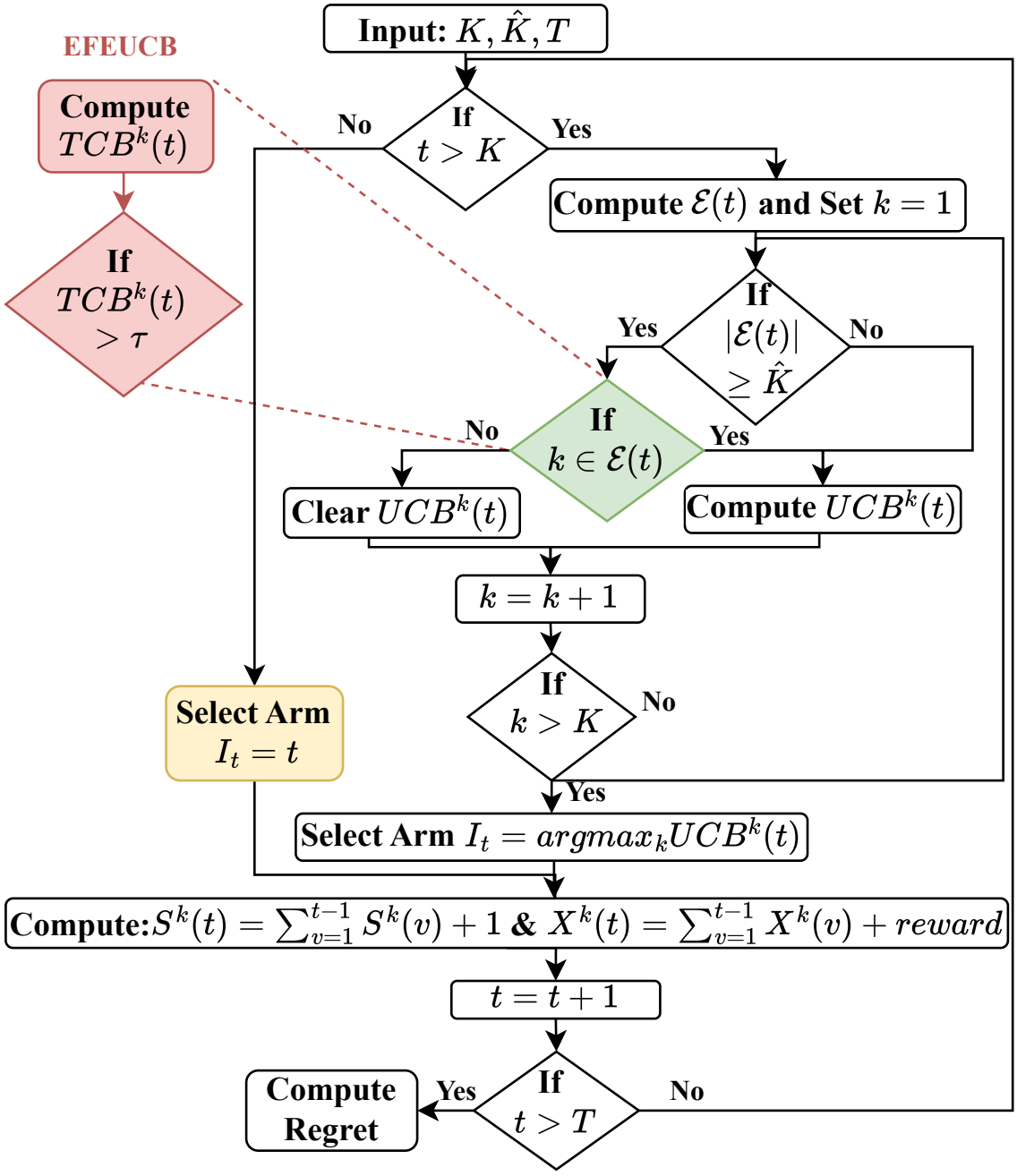

Figure 4.1: Flowchart of the FEUCB and EFEUCB algorithms.

considered. In the extended EFEUCB algorithm, we aim to increase the selection of *good* arms

during the UCB phase of FEUCB algorithm. This is done by limiting the UCB to only those arms whose threshold confidence bound (TCB) given by Eq. 5.2 is greater than threshold, τ (53).

$$TCB^r(t) = \sqrt{S^r(t)}|\hat{\mu}^r(t)| \tag{4.4}$$

where $\hat{\mu}_r(t)$ is the empirical mean reward for arm r up to time t. Such thesholding based approach enables application dependent focused exploration. For instance, in wireless applications, signal-to-noise ratio (SNR) can be used as threshold to identify good beams. In our work (53), we theoretically show that FEUCB offers lower regret than UCB while EFEUCB outperforms FEUCB and UCB in experimental results. However, performance analysis of these algorithms on edge platforms has not been done yet.

4.4 Proposed Architectures

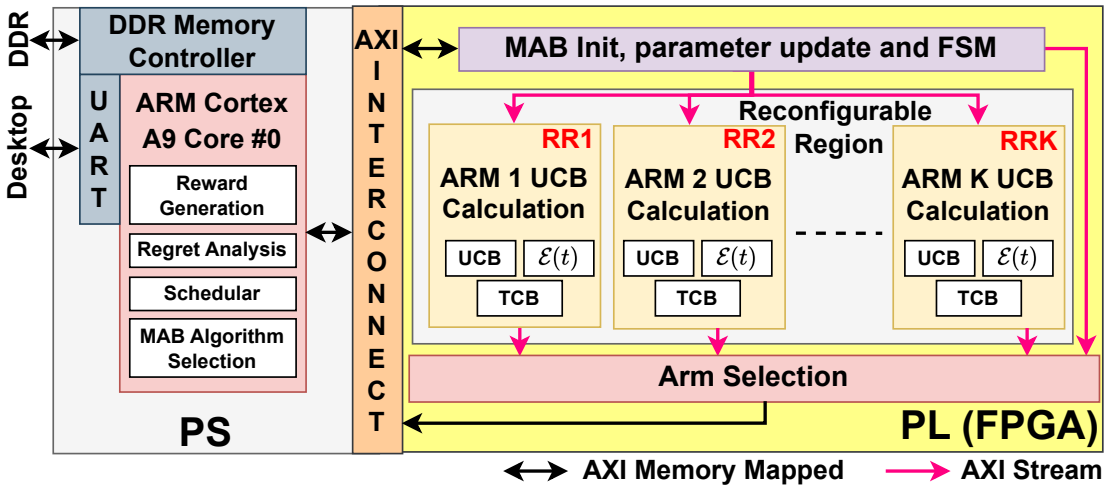

Figure 4.2: Hardware software co-design based reconfigurable architecture.

In this section, we discuss algorithms to architecture mapping of UCB, FEUCB, and EFEUCB algorithms on heterogenous SoC via hardware-software co-design (HSCD) and word length (WL) analysis. The proposed architecture, shown in Fig. 4.2, is realized on Zynq Soc consisting of a processing system (PS) and programmable logic (PL) such as ARM processor and Field-Programmable Gate Array (FPGA), respectively. The sequential and computationally

simple tasks, such as reward generation for selected arm, regret calculation, and control tasks, such as algorithm selection and scheduler, are realized on PS. In each time slot, the scheduler in PS enables the different hardware IPs in PL via the AXI interface, and it receives the selected arm via the arm selector block in PL. The reward of the selected block is calculated in PS and communicated to the parameter update block in PL at the end of the time slot. In Section 4.5, we discuss complete implementation of all algorithms in PS without using PL and corresponding impact of execution time.

4.4.1 Proposed Architecture for FEUCB and EFEUCB

The proposed combined architecture for realizing parameter update, UCB calculation, and Arm Selection tasks in the UCB, FEUCB, and EFEUCB algorithms is shown in Fig. 4.3. In the parameter update task, the reward received from the PS for the selected arm ($X^{I_{t-1}}$) and the corresponding number of selections ($S^{I_{t-1}}$) are accumulated and updated in memory A at the appropriate address. In the UCB calculation task, the quality factor, $UCB^r(t)$, is calculated for each arm, $r \in R$. In this architecture, we use memory B and C to store $\frac{X^r(t)}{S^r(t)}$ and $2\sqrt{\frac{\log(t)}{S^r(t)}}$, respectively. In the case of FEUCB, memory D is used to store $\mathscr{E}(t)$. Similarly, memory E is used to store the TCB of each arm required in EFEUCB.

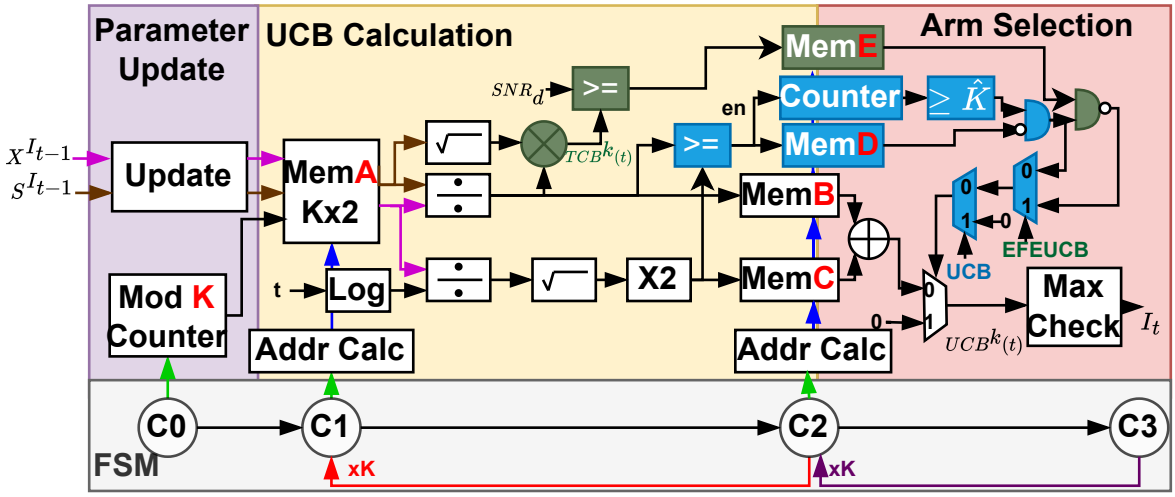

Figure 4.3: Hardware implementation of parameter update, UCB calculation and Arm selection tasks in UCB, FEUCB and EFEUCB algorithms.

In the Arm Selection task, the quality factor is calculated based on the selected algorithm.

In the case of UCB (mux control signal *UCB* set to 0), the element-wise addition between memory B and C is done to obtain the UCB quality factor. The quality factor of all R arms is compared to select the arm I_t with the highest quality factor. In the case of FEUCB (mux control signals *UCB* and *EFEUCB* set to 0) and EFEUCB (mux control signals *UCB* and *EFEUCB* set to 0 and 1, respectively), UCB quality factor calculation for some of the arms is skipped based on the $\mathscr{E}(t)$ and TCB values. Thus, the proposed architecture can dynamically switch between UCB, FEUCB, and EFEUCB by appropriate control signal selection, *UCB*, and *EFEUCB*.

The UCB calculation task is shown to be accomplished sequentially for all R arms. Since the calculation of the UCB factor is independent for each arm, this task can be parallelized depending on the availability of resources on PL. Such serial-parallel configuration helps to reduce the latency at the cost of higher resource utilization and power consumption.

4.4.2 Proposed Architecture for Modified EFEUCB Algorithm

The complexity of the FEUCB and EFEUCB is significantly higher, as observed from the architecture in Fig. 4.3. This is because all computations needed to calculate the UCB quality factor except the final adder are done for all R arms. Since the final adder computation for some of the arms is skipped, FEUCB and EFEUCB offer lower execution times than UCB. Please refer to Section 4.5 for more details. In this section, we present the modified EFEUCB (mEFEUCB) algorithm to reduce its computational complexity.

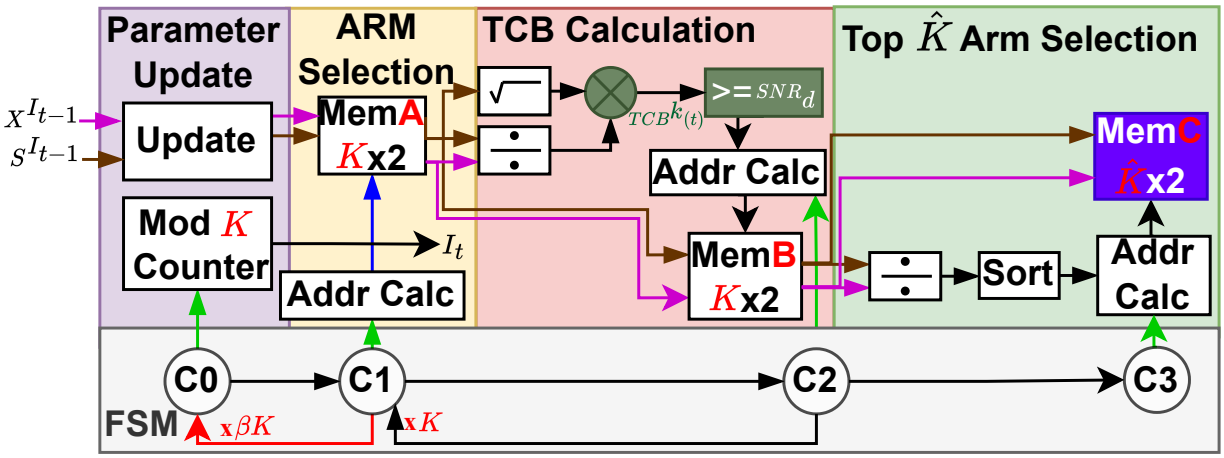

Figure 4.4: Phase 1 of the modified EFEUCB for selection of top $\hat{R}$ arms.

The mEFEUCB algorithm is divided into two phases: 1) Top $\hat{R}$ arms Selection, and 2) EFEUCB. The first phase, shown in Fig. 4.4, consists of a sequential selection of all R arms for β number of times. Thereafter, TCB is calculated for all R arms and the arms with TCB greater than τ are selected as top $\hat{R}$ arms. If there are more than $\hat{R}$ such arms, then the top $\hat{R}$ arms with higher $\frac{X^r(t)}{S^r(t)}$ are selected. The second phase is the implementation of UCB algorithm on the top $\hat{R}$ arms as shown in Fig. 4.5. In this phase, UCB quality factor is calculated for only $\hat{R}$ arms resulting in significant saving in resources and memory.

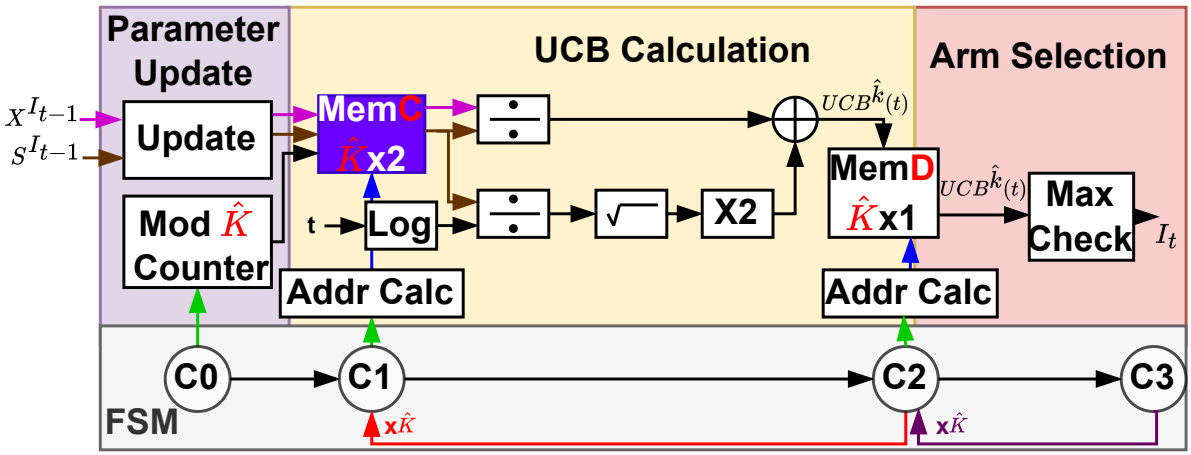

Figure 4.5: Phase 2 of the modified EFEUCB.

For large R, we can safely assume that $\hat{R} \leq R$ where $\hat{R}$ are good arms blocks out of R arms (53). In such a case, it is possible to reduce the distribution-independent upper bound to $\sqrt{\hat{R}T}$ from $\sqrt{RT}$ of the UCB algorithm. The expected regret of the UCB scales as $\mathcal{O}(\sum_{r\in[R]\setminus r_*} \frac{\log T}{\Delta_r})$ where $\Delta_r = X^{r_*} - X^r$ for all $r \neq r_*$ (111). For fixed T, the distribution-independent bounds are of the order $\sqrt{RT}$. It is evident that the UCB suffers from high exploration time, especially when R is large. On the other hand, the expected regret of the FEUCB scales as $\mathcal{O}(\sum_{r\in[\hat{R}]\setminus r_*} \frac{\log T}{\Delta_r})$. For fixed T, the distribution-independent bounds are of the order $\sqrt{\hat{R}T}$. Since $\hat{R} < R$, FEUCB offers lower regret than UCB (53). mEFEUCB is an extension of FEUCB in which we have used round-robin exploration based thresholding and restricted the number of arms to be explored in Stage 2. Thus, the regret of mEFEUCB is expected to be the same or better than FEUCB.

4.5 Performance Analysis

We compare the functionality, complexity, execution time and power consumption of UCB, FEUCB/EFEUCB and mEFEUCB algorithms on Zynq SoC for different WLs and serial-parallel architectures. All the experimental results on Zynq SoC are averaged for 15 independent experiments with $R = \{4, 8, 16, 25, 50, 100\}$, $T = 10000$. A similar setup has been discussed in (53), in which the number of arms is replaced with RIS sub-blocks, and the aim of the player (Transmitter) is to learn and select an RIS sub-block that maximizes the SNR at multiple receivers. UCB architecture for $R = \{4, 8\}$ is discussed in (119; 120) while we design and implement UCB for rest values of R.

4.5.1 Regret Analysis

In Fig. 6.3, we compare the cumulative regret of the three MAB algorithms at different instances of the horizon. Since the regret performance of EFEUCB and mEFEUCB is nearly identical, we skip EFEUCB. We consider single-precision floating point (SPFP) and fixed point (FP) architectures. In the case of FP WL, we use the notation (W, I) where W and I represent the total number of bits and the number of integer bits, respectively. To find appropriate WL, we first identify I for sufficiently large W followed by identification of W for selected I. It can be observed that the regret difference between UCB and proposed mEFEUCB and FEUCB algorithms increases with an increase in R thereby validating the focused exploration approach and its feasibility on the Zynq SoC. For the RIS application considered in this paper, WL of $(17, 5)$ is observed to be optimal as it offers the same performance as SPFP, as shown in Fig. 6.3. Single bit reduction in WL to $(17, 4)$ results in significant deviation in regret performance compared to SPFP, validating the selection of $(17, 5)$. As shown in Table 4.1, mEFEUCB and FEUCB offer an improvement of up to {67%,53%} and {67%,54%} in regret performance over UCB for SPFP and FP WLs, respectively.

Table 4.1: Comparison of Cumulative Regret for Different WLs and R.

WL	MAB	R=4	R=16	R=50	R=100
SPFP	**UCB (120; 119)**	67.40	461.50	1632.03	2977.53
	FEUCB (53)	56.92 (-18%)	392.38 (-15%)	1076.9 (-34%)	1412.5 (-53%)
	mEFEUCB	56 (-19%)	255.84 (-45%)	724.75 (-56%)	983.44 (-67%)
{17,5}	**UCB (120; 119)**	66.43	451.03	1601.15	2930
	FEUCB (53)	55.61 (-16%)	385.86 (-14%)	1064.78 (-34%)	1362.84 (-54%)
	mEFEUCB	55.58 (-16%)	243.75 (-46%)	693.55 (-57%)	955.32 -67%

Table 4.2: Comparison of Resource Utilisation, Power Consumption and Execution Time On Zynq SoC.

K	MAB	SLICE LUTs		Flip-Flops		DSP		Power (W)		Execution Time (μs)				
		SPFP	{17,5}	SPFP	{17,5}	SPFP	{17,5}	SPFP	{17,5}	PS	SPFP (PL)		{17,5} (PL)	
											100 MHz	152 MHz	100 MHz	152 MHz
4	**UCB (119; 120)**	6590	2985	8194	2832	**77**	**0**	1.9	**1.9**	2.1	2.35	2.28	1.35	1.33
	FEUCB	7511 (+13.9%)	2918 (-2.2%)	8727 (+6.5%)	2854 (+0.8%)	80	1	1.9	1.7	1.7	**2.15**	**2.10 (-8%)**	**1.34**	**1.3 (-2%)**
	mEFEUCB	**5646 (-14.3%)**	**2795 (-6.3%)**	**7313 (-10.7%)**	**2484 (-12.3%)**	**77**	1	**1.9**	**1.7**	**1.6**	2.24	2.18 (-4%)	1.57	1.42 (-7%)
16	**UCB**	7547	2970	10175	3231	**77**	**0**	1.9	1.7	8.2	2.5	2.48	1.7	1.65
	FEUCB	9614 (+27%)	3199 (+8%)	11565 (+14%)	3367 (+4%)	80	1	1.9	1.7	5.6	2.3	2.27 (-8%)	**1.5**	1.63 (-1%)
	mEFEUCB	**6556 (-13.3%)**	**2805 (-5%)**	**9126 (-10.3%)**	**2642 (-18.2%)**	**77**	1	**1.8**	**1.6**	**2.9**	**2.2**	**2.08 (-16%)**	1.6	**1.59 (-4%)**
50	**UCB**	10608	3110	15794	3951	**77**	**0**	1.9	**1.7**	25.2	3.4	3.24	2.3	2.27
	FEUCB	16935 (+37%)	3456 (+11%)	18384 (+24%)	4281 (+8%)	80	1	2	1.8	13.6	3.27	3 (-7%)	1.9	1.78 (-22%)
	mEFEUCB	**9269 (-13%)**	**2980 (-4.2%)**	**14172 (-10.2%)**	**3269 (-17%)**	**77**	1	**1.9**	**1.7**	**7.8**	**2.7**	**2.34 (-28%)**	**1.7**	**1.45 (-36%)**
100	**UCB**	15454	3296	23926	4907	**77**	**0**	2	**1.8**	50.5	5	4.87	3.4	3.36
	FEUCB	23683 (+53%)	3734 (+13%)	28606 (+19.5%)	5956 (+21%)	80	1	2.1	1.9	17	4.6	4.45 (-9%)	2	1.88 (-44%)
	mEFEUCB	**13100 (-15%)**	**3106 (-6%)**	**21100 (-11.2%)**	**3910 (-20%)**	**77**	1	**1.9**	**1.8**	**9**	**2.8**	**2.67 (-45%)**	**1.7**	**1.64 (-51%)**

4.5.2 Complexity and Power Comparison

In Table 4.2, we compare the resource utilization, execution time, and power consumption of three MAB algorithms for $R = \{4, 16, 50, 100\}$. The clock frequency of PS is 666 MHz, and

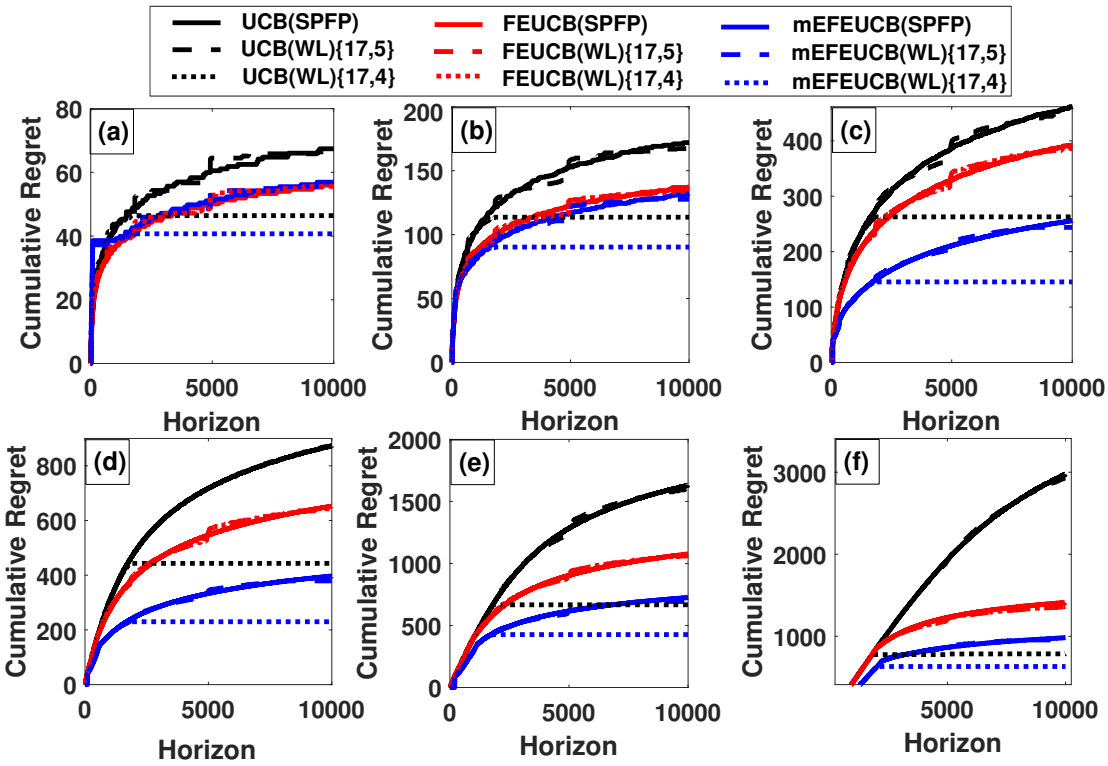

Figure 4.6: Regret Analysis for SPFP and WL $(17,5)$ and $(17,4)$ implementation of our proposed algorithms for (a) $R = 4$ (b) $R = 8$ (c) $R = 16$ (d) $R = 25$ (e) $R = 50$ and (f) $R = 100$.

two PL frequencies of 100 MHz and 152 MHz are used.

For small $R = 4$, the UCB SPFP architecture offers 13.9%, and 6.5% savings in LUTs and FFs, respectively, over FEUCB along with 3 lesser DSP units. This is due to the additional thresholding stage in FEUCB for all R arms. For $R = 100$, the savings in LUTs and FFs are increased to 53% and 19.5%, respectively. The power consumption of UCB is marginally lower than that of FEUCB. Thus, though FEUCB outperforms UCB in regret, its resource utilization and power consumption are higher. When compared to mEFEUCB, UCB needs to compute the UCB quality factor for R arms, while in mEFEUCB, TCB is calculated for R arms, while the UCB quality factor is calculated for $\hat{R}$ arms only. The complexity of TCB is significantly lower than that of the UCB quality factor. Thus, mEFEUCB SPFP architecture offers savings of around 14% and 10% in LUTs and FFs over UCB. Thus, proposed modifications to EFEUCB result in significant savings in resource utilization and lower regret.

The FP WL of {17,5} offers more than 50% savings in LUTs and FFs. In addition, the need for DSP units can be eliminated as the corresponding multiplication and accumulation operations can be realized in LUTs. The FP architecture can be parallelized to reduce the

execution time further at the cost of increased resource utilization.

4.5.3 Execution Time Comparison

In wireless and radar sensing applications, each time slot consists of arm selection using MAB in the beginning. Thus, the execution time of the MAB should be as small as possible to have more time available for communication or sensing tasks. In Table 4.2 (last five columns), we compare the execution time of three algorithms on PS (ARM Cortex A9 at 666 MHz) and PL (FPGA at 100 MHz and 152 MHz). For $R = 4$, PS realization of FEUCB and mEFEUCB offer 19% and 24% reduction in execution time over UCB. For $R = 100$, PS realization of FEUCB and mEFEUCB offer 66% and 82% reduction in execution time over UCB. Thus, as R increases, mEFEUCB outperforms FEUCB by more than 50%.

Next, we accelerate MAB algorithms on PL. For small $R = 4$, acceleration on PL is not possible due to significant overhead in data communication between PS and PL when compared to actual computation. For large $R = 100$, the mEFEUCB in PL offers 70% lower execution time than mEFEUCB in PS, resulting in an overall reduction of 94% and 83% over UCB and EFEUCB in PS, respectively. Further reduction in execution time is observed when the WL is changed from SPFP to FP and increasing the PL clock frequency to the maximum possible 152 MHz by efficiently pipelining the circuit. As shown in the last column of Table 4.2, the mEFEUCB in PL offers 82% lower execution time than mEFEUCB in PS, resulting in an overall reduction of 97% and 90% over UCB and EFEUCB in PS, respectively. Further, mEFEUCB in PL is up to 51% and 13% faster than UCB and FEUCB in PL, respectively.

To analyze the performance on different edge platforms, we compare the execution time on Cortex A9 and Cortex A53 ARM processors operating at different frequencies and augmented with single instruction multiple data (SIMD) NEON co-processor. The detailed comparison for different R is shown in Table 4.3. For $R = 4$, FEUCB and mEFEUCB offer execution time reduction by $18\% - 22\%$ and $22\% - 26\%$, respectively. The savings in execution time increases substantially with an increase in R, and for $R = 100$, FEUCB and mEFEUCB offer a reduction of $65\% - 66\%$ and $80\% - 84\%$, respectively. These results validates the superiority of the mEFEUCB algorithm on different types of edge platforms.

Table 4.3: Comparison of Execution Time on Edge Platforms

Arms (R)	Algorithm	Cortex A9 (666MHz)	Cortex A9 and NEON (666MHz)	Cortex A9 (800MHz)	Cortex A9 and NEON (800MHz)	Cortex A53 (1.1GHz)	Cortex A53 and NEON (1.1GHz)
4	UCB	2.11	1.52	2.09	1.51	0.94	0.68
	FEUCB	1.73 (-18%)	1.17 (-23%)	1.71 (-18%)	1.15 (-24%)	0.78 (-17%)	0.54 (-21%)
	mEFEUCB	1.65 (-22%)	1.12 (-26%)	1.63 (-22%)	1.10 (-27%)	0.74 (-21%)	0.51 (-24%)
16	UCB	8.19	5.97	8.14	5.93	3.78	2.73
	FEUCB	5.65 (-31%)	3.88 (-35%)	5.58 (-31%)	3.82 (-35%)	2.59 (-31%)	1.78 (-36%)
	mEFEUCB	2.87 (-65%)	1.92 (-68%)	2.81 (-65%)	1.85 (-69%)	1.33 (-65%)	0.89 (-67%)
50	UCB	25.18	18.57	24.58	18.11	11.6	8.30
	FEUCB	13.57 (-46%)	9.62 (-48%)	13.23 (-46%)	9.40 (-48%)	6.32 (-46%)	4.39 (-47%)
	mEFEUCB	7.79 (-69%)	5.38 (-70%)	7.65 (-69%)	5.24 (-71%)	3.76 (-68%)	2.50 (-70%)
100	UCB	50.46	36.96	49.24	36.07	23.11	16.48
	FEUCB	16.99 (-66%)	12.53 (-66%)	16.67 (-66%)	12.14 (-66%)	8.08 (-65%)	5.68 (-65%)
	mEFEUCB	9.09 (-82%)	6.05 (-84%)	8.87 (-82%)	5.91 (-84%)	4.67 (-80%)	2.84 (-83%)

4.6 Summary

In this chapter, we consider the design and implementation of multi-armed bandit (MAB) algorithms with many arms on Zynq system on chip (SoC) via hardware software co-design and fixed-point analysis. The proposed algorithm and its architecture offers a 67% reduction in average cumulative regret, a 97% reduction in execution time, and 10% savings in resources over state-of-the-art MABs for 100 arms. This is due to focussed exploration approach in which number of candidate arms are limited in the beginning thereby minimizing the selection of sub-optimal arms and corresponding computations in hardware.

This page was intentionally left blank.

Chapter 5

Optimizing RIS Block Selection for Power Consumption

5.1 Overview

The next generation of wireless networks aims to achieve higher data rates, seamless connectivity, and enhanced energy efficiency. In multi-Reconfigurable Intelligent Surface assisted systems, optimizing RIS selection to maximize throughput and signal-to-noise ratio presents a significant challenge due to the large number of RIS sub-blocks. Conventional approaches often select the RIS sub-block that yields the highest SNR; however, this method can lead to excessive power consumption. Determining the optimal number of active sub-blocks, M is critical for balancing target SNR and power efficiency, as increasing M typically enhances key performance metrics such as outage probability and ergodic capacity. To address this, we propose a novel sensor selection-based multi-armed bandit (MAB) framework that dynamically learns and selects the optimal RIS sub-block, achieving an efficient trade-off between power consumption and SNR.

5.2 Introduction

Sensor selection problems present a powerful abstraction for resource-constrained decision-making tasks and offer useful insights into RIS sub-block selection. In applications such as security surveillance, medical diagnostics, and robotics, a series of sensors—each with different costs and error rates—are queried sequentially to identify events or states with high confidence. The learner must select a sensor that balances the need for reliable information against the overhead of cost and delay.

Formally, sensor selection is often modeled as a cascading structure, where low-indexed sensors are cheaper and offer quick but noisy feedback, while high-indexed sensors offer precise but expensive results. The cascade architecture is naturally suited to real-time systems, where early termination may be necessary to meet latency or energy budgets. In RIS-aided communication systems, the analogy is immediate: small subsets of RIS elements (or sub-blocks) provide partial gains at low power cost, while full configurations maximize signal strength at the expense of energy efficiency.

Many sensor selection problems fall under the umbrella of stochastic partial monitoring, a generalization of MABs where the learner receives indirect or partial feedback. Unlike classical MABs that receive exact rewards, partial monitoring setups return signals that correlate with outcomes, requiring inference over noisy observations. Applications in crowdsourcing (124), resource allocation (125), and medical testing (126) highlight scenarios where exact labels or rewards are infeasible to obtain.

The stochastic nature of feedback and cost-reward trade-offs calls for robust algorithms that can handle budgeted learning under uncertainty. Works such as (127; 128) have laid foundational theories on efficient sensor querying under partial observability. These frameworks define reward structures and budget constraints and propose upper confidence-based selection rules, regret bounds, and cascading optimality conditions.

The relevance to RIS is compelling: intelligent sub-block selection can be seen as querying parts of a reflective surface with increasing spatial resolution. Like cascading sensors, each additional RIS sub-block refines the received signal estimate but adds to power consumption. Therefore, extending sensor selection frameworks to RIS configurations provides a structured

way to achieve energy-efficient link optimization under real-time and uncertainty constraints.

5.3 Network Model

In this chapter we consider the network model same as (53), as shown in Fig. 5.1. The network layout consists of a system with a single transmitter, K RIS, and N receivers. Further, each RIS is divided into L sub-blocks, and each sub-block consists of Z sub-λ sized passive RIS elements. A group composed of any M sub-blocks is denoted as a block. In the selected block, the amplitude A ranges from $[0,1]$, while the phase θ ranges from $[0,2\pi]$. These parameters are optimized to establish communication with N receivers. I_t is the selected block in time slot t. It needs to be noted that each RIS can have different number of elements and different number of sub-blocks. However, in our simulations we have considered all RIS to be identical. It is also considered that, with the increase of number of RIS elements in a sub-block, there will be an increase in the total consumed power. We analyze the channel model outlined in (23), where channels linked to elements within the same RIS are assumed to be independently and identically distributed (IID). Conversely, channels associated with different RISs are regarded as independent but not identically distributed (INID), with the system experiencing Nakagami-m fading.

5.4 Proposed Work

In (53), selecting the best RIS block in systems with multiple RISs and receivers has been compared to solving a multi-armed bandit (MAB) problem,where the transmitter (player) selects RIS blocks (arms). In each time slot, the player selects an arm and receives a reward (equivalent to the SNR at each receiver). The existing MAB algorithms are not always effective in handling the complexities of such scenarios. This work builds on earlier research by focusing on improving the process of selecting a set of M RIS sub-blocks block with multiple receivers, and it highlights the weaknesses of existing MAB algorithms in these situations.

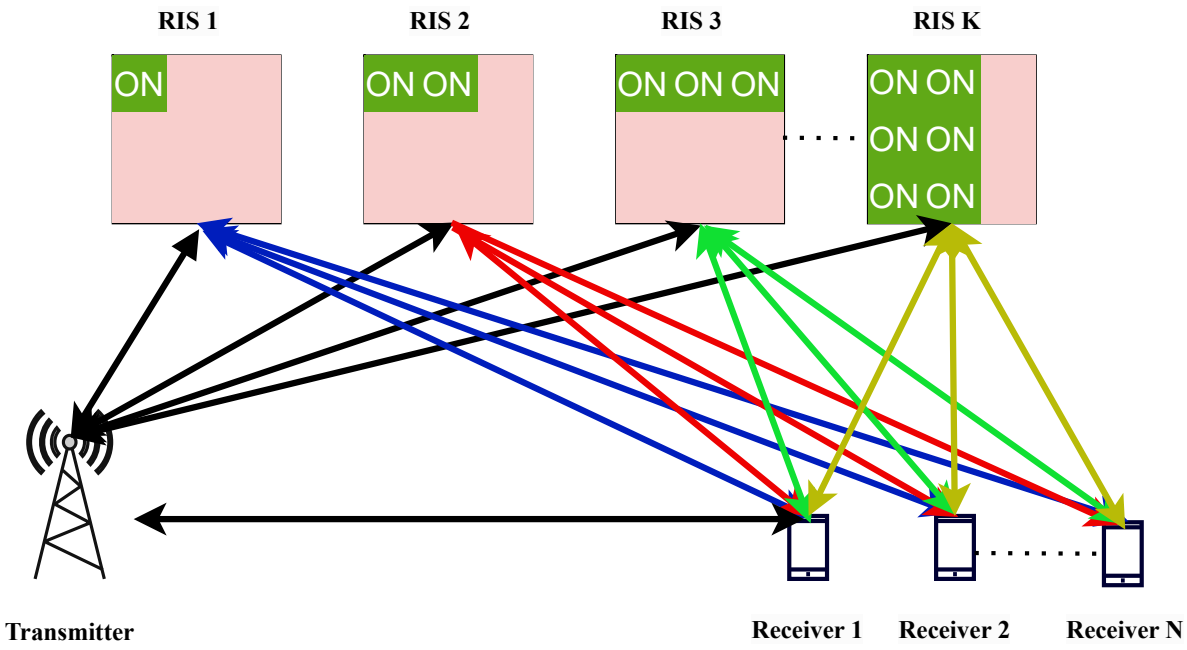

Figure 5.1: Illustrations of network model for multi-RIS-aided wireless system with active M blocks from different RIS

5.4.1 Limitations of Existing MAB Framework in RIS Aided Wireless Communication

In (53), the utilization of the MAB Framework in RIS-aided wireless communication to choose the RIS sub-block that maximizes the SNR at the receivers has been discussed. It considers the RIS selection for fixed M, i.e., all candidate RIS sub-blocks are identical in size and, hence, consume the same power as the consumed power is directly linked to the number Z sub-λ size passive RIS elements. In this work, we do not assume fixed M.

5.4.2 Modifying the Existing MAB Framework for selection of sub-blocks

The proposed work modifies the MAB algorithms in (53) with an aim to learn and select the optimal block size, M, to achieve a target SNR. This is achieved by modifying the reward function which takes the consumed power and the error rate into account. Since the error rate

is not known but strongly correlated to the actual SNR, the modified reward function is written as:

$$X^{I_t} = \sum_{\substack{v=i_1 \\ v\in[I_t]}}^{i_M} \left(\prod_{n=1}^{N} 1_{\mathrm{SNR}_n^r > \Delta} \sum_{n=1}^{N} \mathrm{SNR}_n^r \right) - (Cp_{I_t} + \phi_{I_t}) \tag{5.1}$$

The average reward for the n^{th} receiver during block r is represented as SNR_n^r. Consequently, the total reward in time slot t is computed by incorporating the sum of the consumed power Cp_{I_t} and error rate ϕ_{I_t} of the selected arm I_t is considered as a penalty along with the actual reward. When selecting with a block with smaller number of sub-blocks, the reward is mainly influenced by the error rate. Conversely, in a block with a more number of sub-blocks, the reward is primarily driven by the consumed power. The Enhanced Focused Exploration Upper Confidence Bound

Algorithm 1 EFEUCBwCost

1: **Input:** $R, \hat{R}, T$
2: Initialize $\hat{X}^r(t) \leftarrow 0$, $S^r \leftarrow 0$ for all r
3: **for** $t = 1$ to T **do**
4: Select block $I_t = t$ if $t \leq R$, else compute $\mathscr{E}(t) := \left\{ r \in [R] \middle| \hat{X}^r(t) \geq 2\sqrt{\frac{\log(t)}{S^r(t)}} \right\}$
5: If $|\mathscr{E}(t)| \geq \hat{R}$, set $UCB^r(t) = \frac{\hat{X}^r(t)}{S^r(t)} + 2\sqrt{\frac{\log(t)}{S^r(t)}}$ for $r \in \mathscr{E}(t)$
6: Else, calculate $TCB^r(t) = \sqrt{S^r(t)} \left| \frac{\hat{X}^r(t)}{S^r(t)} \right|$ and update $UCB^r(t)$ for $r \in [R]$ where $TCB^r(t) > SNR_d$
7: Select $I_t = \arg\max_{r\in[R]} (UCB^r(t))$, configure and transmit with I_t, then update $\hat{X}^r(t)$, S^{I_t} based on SNR
8: **end for**

(EFEUCBwCost) algorithm is based on thresholding approach to better identify optimal RIS blocks in scenarios where the number of candidate blocks is high. EFEUCBwCost improves the exploration-exploitation process by incorporating an empirical thresholding approach. In EFEUCBwCost, each RIS block is first assessed for sufficient performance by computing a Threshold Confidence Bound (TCB), as given by:

$$TCB^r(t) = \sqrt{S^r(t)} \left| \frac{\hat{X}^r(t)}{S^r(t)} \right| \tag{5.2}$$

This TCB term helps to select RIS blocks with SNR above a certain threshold SNR, focusing exploration on only the most promising blocks that exceed this threshold. This selective sampling enables EFEUCBwCost to converge to the optimal RIS blocks faster, especially in sparse environments where only a few RIS blocks are highly effective.In Algorithm 1, initially, blocks are sequentially selected (Line 4). The blocks meeting reward thresholds are considered in (Line 5). If the number of such blocks are sufficient, UCB values are computed for them (Line 6), otherwise, TCB is calculated for all blocks, and only those exceeding a defined SNR threshold are updated (Line 7). The block with the maximum UCB is selected.

5.4.3 Conditions for Optimal RIS sub-block

The optimal RIS sub-block m^* is considered to satisfy the following conditions,

$$\forall p < m^* : Cp_m^* - Cp_p \leq \phi_p - \phi_m^* \tag{5.3}$$

$$\forall p > m^* : Cp_p - Cp_m^* > \phi_m^* - \phi_p \tag{5.4}$$

The aforementioned equations are not able to build a trustworthy sub-block selection criterion since the loss of a sub-block is not directly observable. As a result, establishing a link between observable and unobservable components becomes essential. We can calculate the likelihood of disagreement between sub-block in our design by comparing their feedback.

5.4.4 Consumed Power Aware Upper Confidence Bound (CPAUCB) based Selection of Optimal RIS sub-block

The proposed algorithm is based on the sensor selection based approach, in which different groups of RIS-sub blocks are assumed as sensors. With $M = 1$ as the sensor with least consumed power and high error rate, and the sensor with maximum consumed power is assumed when all the sub-blocks of a RIS work together, i.e. $M^k = \sum_{i=1}^{L} i$. The illustration can be well understood

from Fig. 5.1. The algorithm helps to select a combination of M sub-blocks which is ideal in the case of consumed power - SNR_{th} trade-off for a particular achievable rate.

For higher SNR_{th} (keeping same constant achievable rate as earlier), we combine the optimal blocks from different RIS to form a new RIS, now it is interesting to note that it is important to arrange the RIS blocks as a requirement for algorithm, for this, we have arranged the blocks as per the SNR at the receiver. Starting from the least to the maximum SNR. As shown in Fig. 5.2. Since the RIS are randomly placed, therefore it is not necessary that a block with higher value of M will guarantee more SNR as compared to the RIS which is placed near to the receivers and is able to achieve the target SNR with lower values of M. As mentioned in the

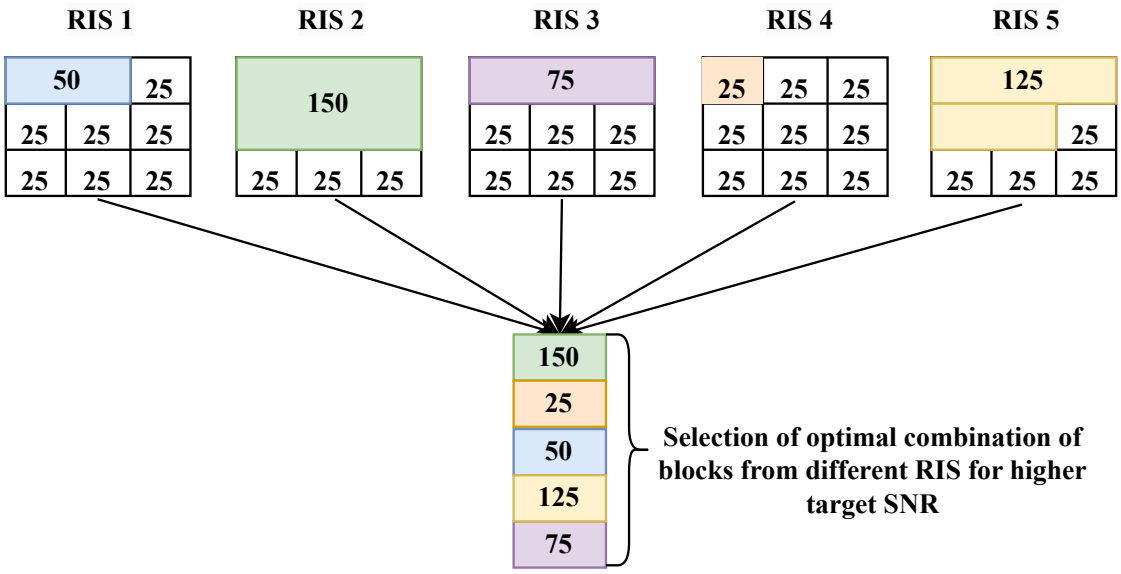

Figure 5.2: Schematic representation for selecting a block for higher target SNR from individual best blocks from each RIS

previous sub-section that Eq. 5.3 and Eq. 5.4 are not able to build a trust-worthy block selection criterion since the loss of a block is not directly observable. Therefore we replace the error rates with the probability of disagreements, i.e. $\mathbb{P}(P^*_{1/0_m} \neq P_{1/0_p})$. Using this term, the Eq. 5.3 and Eq. 5.4, can be modified as

$$\alpha = \left\{ \forall p < m^* : Cp^*_m - Cp_p \leq \mathbb{P}(P^*_{1/0_m} \neq P_{1/0_p}) \right\} \cup \{M_i = \sum_{i=1}^{L=1} i\} \tag{5.5}$$

$$\beta = \left\{ \forall p > m^* : Cp_p - Cp^*_m > \mathbb{P}(P^*_{1/0_m} \neq P_{1/0_p}) \right\} \cup \{M = \sum_{i=1}^{L} i\} \tag{5.6}$$

$\mathbb{P}(P^*_{1/0_m} \neq P_{1/0_p})$ is observable as P_1 and P_0 are easily available at the transmitter, based on SNR_{th}. The following equations are followed individually by each K RIS. Since there are total of LK sub-blocks.

Updating Eq. 5.5 and Eq. 5.6 for all $k \in K$ by the optimistic estimates we get :

$$\alpha = \left\{ \forall p < m^* : Cp^*_m - Cp_p \leq \mu^k_{mp}(t) + \phi^k_{mp}(t) \right\} \cup \{M_i = \sum_{i=1}^{L=1} i\} \tag{5.7}$$

$$\beta = \left\{ \forall p > m^* : Cp_p - Cp^*_m > \mu^k_{mp}(t) + \phi^k_{mp}(t) \right\} \cup \{M = \sum_{i=1}^{L} i\} \tag{5.8}$$

where $\mu^k_{mp}(t)$ is the mean of disagreements $(X^k_{pm}(t))$ and number of times the pair of blocks $m, p \in M^k = \sum_{i=1}^{L} i$ of RIS k has been explored and $\phi^k_{mp}(t)$ is the related optimistic upper confidence bound, derived from (74).

Lemma 4 *The optimal block m^* should satisfy the Eq.5.7 and 5.8. s.t.*

$$m^* = (\alpha \cap \beta) \tag{5.9}$$

Proof: Inspired from the literature (128), and considering our scenario. There are three conditions where a sub-block/block can be considered as optimal :

1. $M_i = \sum_{i=1}^{L=1} i < m^* < M = \sum_{i=1}^{L} i$
2. $m^* = M_i = \sum_{i=1}^{L=1}$
3. $m^* = M = \sum_{i=1}^{L} i$

Case 1). When $M_i = \sum_{i=1}^{L=1} i < m^* < M = \sum_{i=1}^{L} i$

Since, m^* is an optimal sub-block/block, which implies that $\forall p > m^* : Cp_p - Cp^*_m > \mathbb{P}(P^*_{1/0_m} \neq P_{1/0_p}) \Rightarrow Cp_p - Cp^*_m \not\leq \mathbb{P}(P^*_{1/0_m} \neq P_{1/0_p}) \Rightarrow \forall p > m^* \notin \alpha_{-1}(\text{s.t.}\alpha_{-1} = \alpha \backslash M_i = \sum_{i=1}^{L=1} i)$. For any

block $a \in \alpha_{-1}$ then $a \leq m^*$

$$\alpha_{-1} = \{a_1, a_2 \ldots m^*\} \text{where} (1 < a_1 < a_2 \ldots m^*) \tag{5.10}$$

$$\alpha = \alpha_{-1} \cup \left\{ M_i = \sum_{i=1}^{L=1} i \right\} = \left\{ M_i = \sum_{i=1}^{L=1} i, a_1, a_2 \ldots m^* \right\} \tag{5.11}$$

similarly , $\forall p < m^* : Cp^*_m - Cp_p \leq \mathbb{P}(P^*_{1/0_m} \neq P_{1/0_p}) \Rightarrow Cp^*_m - Cp_p \not> (P^*_{1/0_m} \neq P_{1/0_p}) \Rightarrow \forall p < m^* \notin \beta_{-M=\sum_{i=1}^{L} i} (\text{s.t.} \beta_{-M=\sum_{i=1}^{L} i} = \beta \backslash M = \sum_{i=1}^{L} i)$. For any block $b \in \beta_{-M=\sum_{i=1}^{L} i}$ then $b \geq m^*$

$$\beta_{-M=\sum_{i=1}^{L} i} = \{m^*, b_1 \ldots M = \sum_{i=1}^{L} i\} \text{i.e.} (m^* < b_1 < \ldots M = \sum_{i=1}^{L} i) \tag{5.12}$$

$$\beta = \beta_{-M=\sum_{i=1}^{L} i} \cup \left\{ M = \sum_{i=1}^{L} i \right\} = \left\{ m^*, b_1 \ldots M = \sum_{i=1}^{L} i \right\} \tag{5.13}$$

combining the Eq. 5.11 and Eq. 5.13, we get

$$m^* = (\alpha \cap \beta) \tag{5.14}$$

Case 2). When $m^* = M_i = \sum_{i=1}^{L=1}$

From Eq. 5.10, it is clear that $\alpha_{-1} = \emptyset$, therefore $\alpha = M_i = \sum_{i=1}^{L=1} i$, and from Eq. 5.13, we have $\beta = \left\{ M_i = \sum_{i=1}^{L=1} i, b_1, b_2, \ldots M = \sum_{i=1}^{L} \right\} i$ that also implies that :

$$m^* = \left\{ M_i = \sum_{i=1}^{L=1} i \right\} \tag{5.15}$$

Case 3). When $m^* = M = \sum_{i=1}^{L} i$

From Eq. 5.12, it is clear that $\beta_{-M=\sum_{i=1}^{L} i} = \emptyset$, therefore $\beta = M = \sum_{i=1}^{L} i$, and from Eq. 5.11, we

have $\beta = \left\{ M_i = \sum_{i=1}^{L=1} i, a_1, a_2, \ldots M = \sum_{i=1}^{L} \right\} i$ that also implies that :

$$m^* = \left\{ M = \sum_{i=1}^{L} i \right\} \tag{5.16}$$

The lemma presented above states all the possible conditions of an optimal block m^*.

Algorithm 1 CPAUCB

Input: K, L, T, SNR_{th}
Initialize: $X_{pm}^k(t) \leftarrow 0$ and $S_{pm}^k(t) \leftarrow 0$ for all m
Select block $I_t^k(1) = M^k$
for $k = 1, 2 \ldots K$, **do**
    Observe $P_{1/0_1}(1) \ldots P_{1/0_{M^k}}(1)$ based on SNR_{th}
    Update $X_{pm}^k(1) \leftarrow X_{pm}^k(1) + 1_{(P_{1/0_m} \neq P_{1/0_p})}\ \forall m < p \leq M^k$
    Update $S_{pm}^k(1) \leftarrow S_{pm}^k(1) + 1\ \forall m < p \leq M^k$
    for $t = 2, 3 \ldots T$ **do**
        $\mu_{mp}^k(t) = \frac{X_{pm}^k(t-1)}{S_{pm}^k(t-1)}\ \forall m < p \leq M^k$
        $\phi_{mp}^k(t) = \sqrt{\frac{2\log(t)}{S_{pm}^k(t-1)}}\ \forall m < p \leq M^k$
        $I_t^k(t) = \min\{(\alpha \cap \beta) \cup M^k\}$
        Observe $P_{1/0_1}(t) \ldots P_{1/0_{I_t^k}}(t)$ based on SNR_{th}
        Update $X_{pm}^k(t) \leftarrow X_{pm}^k(t-1) + 1_{(P_{1/0_m} \neq P_{1/0_p})}\ \forall m < p \leq I_t^k(t)$
        Update $S_{pm}^k(t) \leftarrow S_{pm}^k(t-1) + 1\ \forall m < p \leq I_t^k(t)$
    end for
end for
Arrange the learnt optimal blocks from each K RIS based on the SNR achieved at the Receivers.
Update the SNR_{th}, since this RIS made up of selected blocks from other K RIS will target higher values of SNR_{th}.

5.5 Performance Analysis

We investigate the performance of a time-slotted multi-RIS-assisted communication system comprising a single transmitter and multiple receivers. In this framework, the transmitter dynamically selects one Reconfigurable Intelligent Surface (RIS) block per time slot to transmit data packets. The primary objective is to maximize the selection frequency of the optimal RIS block, thereby enhancing resource utilization efficiency.

To assess the system's effectiveness, we evaluate the proposed Consumed Power-Aware Upper Confidence Bound (CPAUCB) algorithm against established benchmark algorithms designed for multi-RIS-aided communication scenarios. The comparative study includes learning-based methods such as FEUCB and EFEUCB (53; 129), in addition to a modified version of the classical UCB algorithm (74) that integrates power consumption and signal-to-noise ratio (SNR) constraints to guide optimal arm (RIS block) selection.

In our system model, we consider a multi-RIS-assisted communication framework with number of receivers $N = 4$, number of RIS $K = [5, 10]$. Each RIS is partitioned into five sub-blocks, with the number of elements per sub-block represented by the array $Z = [25, 25, 25, 25, 25]$.The aggregated elements within each block are denoted as $M_i = \sum_{i=1}^{L} i = \{25, 50, 75, 125, 225\}$. The horizon size, denoted as T, ranges within the interval of 10000. Each result is averaged over 15 independent experiments over the horizon size. For each trial, the receiver and RIS locations are randomly initialized, while the transmitter position remains fixed. The equivalent noise power at the receiver is given as:$\sigma_n^2 = N_0 + 10log(BW) + NoiseFigure[dBm]$ where N_0 is the thermal noise power density.

The outage probability is the probability that the SNR falls below a predefined threshold. Outage Probability is affected by multiple factors, including channel fading, receiver noise, and RIS block selection, which collectively impact the SNR. A lower outage probability is correlated with an increased number of selected RIS blocks M, as larger RIS configurations enhance signal reflection and improve link reliability.

The ergodic capacity is the maximum achievable data rate at which information can be reliably transmitted over a selected channel, averaged over all possible channel realizations. It is the average capacity in scenarios where the channel undergoes stochastic variations over time,

Table 5.1: Parameters

Parameters	**Values**
Location of Transmitter	(0,0)
Location of Receivers	Random
Location of RISs	Random
Transmit Power [dBm]	[-40,40]
Amplitude reflection coef., A	1(53)
Number of RIS	5 and 10
Number of Sub-blocks per RIS	5
Number of Elements in sub-blocks	[25,25,25,25,25]
Threshold SNR (SNR_{th})	15dBm and 31dBm
Bandwidth	10 MHz (53)
Noise Figure	10 (53)
Thermal noise power density,(N_0)	-174 (53)
Antenna gain[dB]	5 (53)
Carrier Frequency [GHz]	3 (53)
Circuit Dissipated Power in RIS [mW]	7.8 (53)
Circuit Dissipated Power in TX and RX[mW]	10 (53)

assuming both the transmitter and receiver have perfect CSI. Analogous to outage probability, higher ergodic capacity is typically observed for larger values of M, as increase in number of RIS elements enhances signal diversity and improves spectral efficiency.

As outlined in the previous sections, the primary objective of this work is to determine the optimal value of M that achieves a balanced trade-off between power consumption and the target SNR threshold. While alternative learning algorithms may require lower transmit power to attain reduced outage probability and may yield higher ergodic capacity, this is attributed to the reduced frequency of selecting the optimal arm compared to the CPAUCB algorithm. Furthermore, existing learning algorithms tend to select sub-optimal arms with higher M values more frequently, leading to inefficient resource utilization, whereas the CPAUCB algorithm prioritizes energy-efficient arm selection to optimize system performance.

5.5.1 Regret Comparison

We begin by evaluating multiple MAB algorithms for RIS selection, considering $K = [5, 10]$, $L = 5$ and $N = 4$. Figure 6.3 illustrates the cumulative regret over a horizon of 10,000 time steps. Among all learning-based algorithms, our proposed CPAUCB algorithm demonstrates superior performance compared to EFEUCB, FEUCB, and UCB (53; 129; 74). This advantage arises because CPAUCB eliminates suboptimal RIS blocks from the outset, specifically those that do not satisfy Equations 5.7 and 5.8, thereby reducing the exploration time. In contrast, FEUCB, EFEUCB, and UCB initially explore all arms, only filtering out suboptimal blocks over time. Due to the thresholding mechanism employed in FEUCB and EFEUCB, they outperform UCB by converging faster to optimal selections. In Figure 6.3(a), where the number of arms is relatively small, the performance difference between FEUCB and EFEUCB is negligible. However, as seen in Figure 6.3(b), where the number of arms increases, EFEUCB exhibits a slight performance gain over FEUCB, indicating its improved adaptability in larger action spaces. Additionally, the Opportunistic RIS-Aided (ORA) scheme (23), which selects the RIS block solely based on maximum SNR without considering power consumption, exhibits consistently high regret in both cases. This is because ORA lacks a learning mechanism and always selects the block with the maximum number of RIS elements, leading to suboptimal power efficiency and excessive regret accumulation.

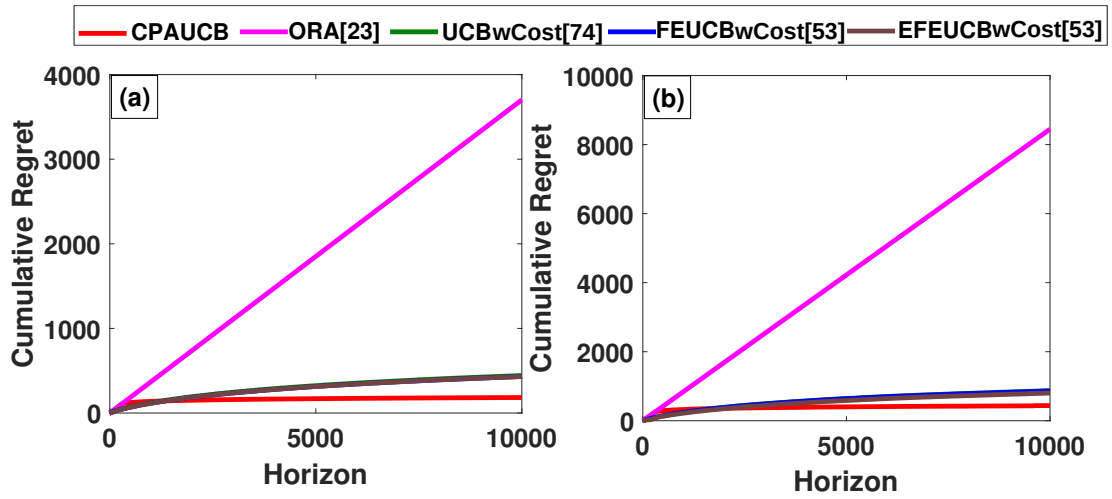

Figure 5.3: Comparison of Average Cumulative Regret for different algorithms for (a) $K = 5, L = 5$ (b) $K = 10, L = 5$

5.5.2 Achievable Rate

In this sub-section, we analyze the impact of consumed power and transmit power on the achievable rate, which represents the maximum data rate that can be effectively transmitted over a communication channel under given constraints. It is a measure that quantifies the effective throughput, considering channel impairments, noise, interference, and modulation schemes. Figure 5.4 illustrates the transmit power required to achieve a specific target data rate. It is observed that the ORA scheme requires the least transmit power, as it selects the RIS block with the maximum number of elements, thereby maximizing SNR. In contrast, learning-based schemes optimize RIS block selection for a specific target SNR, leading to a higher transmit power requirement. However, as seen in Figure 5.5, the total consumed power of the ORA scheme is significantly higher than that of the learning-based approaches. This is due to the ORA scheme activating the maximum number of RIS sub-blocks, leading to excessive power consumption. Among the learning-based algorithms, our proposed CPAUCB algorithm outperforms the other schemes, as it is tightly bounded and rapidly converges to the optimal RIS block, thereby ensuring an efficient trade-off between achievable rate and power consumption, as discussed in the previous sub-section.

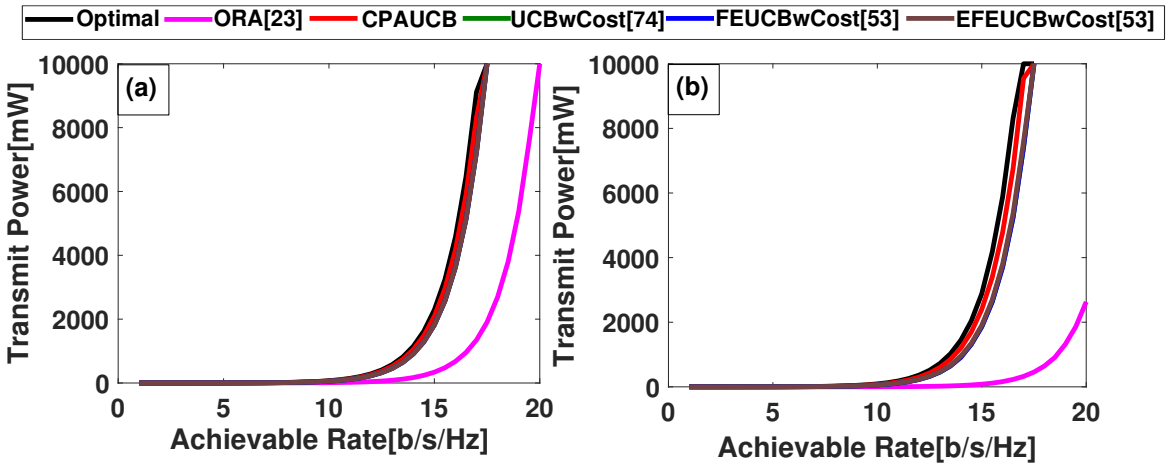

Figure 5.4: Comparison of Transmit Power vs. Achievable Rate for different algorithms for (a) $K = 5, L = 5$ (b) $K = 10, L = 5$

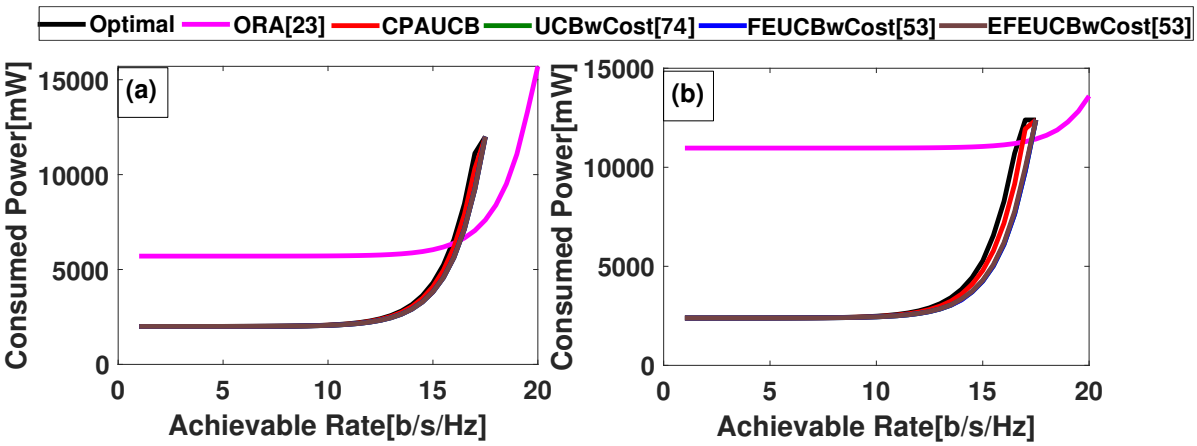

Figure 5.5: Comparison of Consumed Power vs. Achievable Rate for different algorithms for (a) $K = 5, L = 5$ (b) $K = 10, L = 5$

5.5.3 Ergodic Capacity and Outage Probability

In Figure 5.6, we evaluate the performance of Outage Probability (OP) and Ergodic Capacity (EC) with Transmit Power. Figures 5.6 (a) and (b) illustrate that the ORA scheme achieves the highest ergodic capacity, as it utilises the maximum number of RIS sub-blocks, thereby maximizing SNR. Among the learning-based algorithms, the CPAUCB algorithm closely approximates the optimal case, efficiently balancing target SNR and power consumption. However, EFEUCB, FEUCB, and UCB exhibit slightly higher ergodic capacities due to their exploration of sub-optimal RIS blocks containing a greater number of sub-blocks, thereby introducing fluctuations in the average reliable data transmission rate over a selected channel. In Figures 5.6

(c) and (d), we observe that the ORA scheme requires the lowest transmit power to achieve lower outage probabilities, owing to its selection of RIS blocks with a higher number of sub-blocks (M). Among the learning-based algorithms, EFEUCB, FEUCB, and UCB demonstrate lower outage probabilities than CPAUCB, as highlighted in the figure. This is attributed to their selection of sub-optimal RIS blocks with larger M,which inherently results in lower outage probabilities but at the cost of increased power consumption. Meanwhile, CPAUCB achieves performance closest to the optimal case, which is defined as the configuration satisfying Equations 5.3 and 5.4, assumed to be known in hindsight. Furthermore, the penalty of excessive exploration is evident in the lower outage probabilities, where deviation from the optimal RIS block selection becomes noticeable due to the exploration of a greater number of arms.

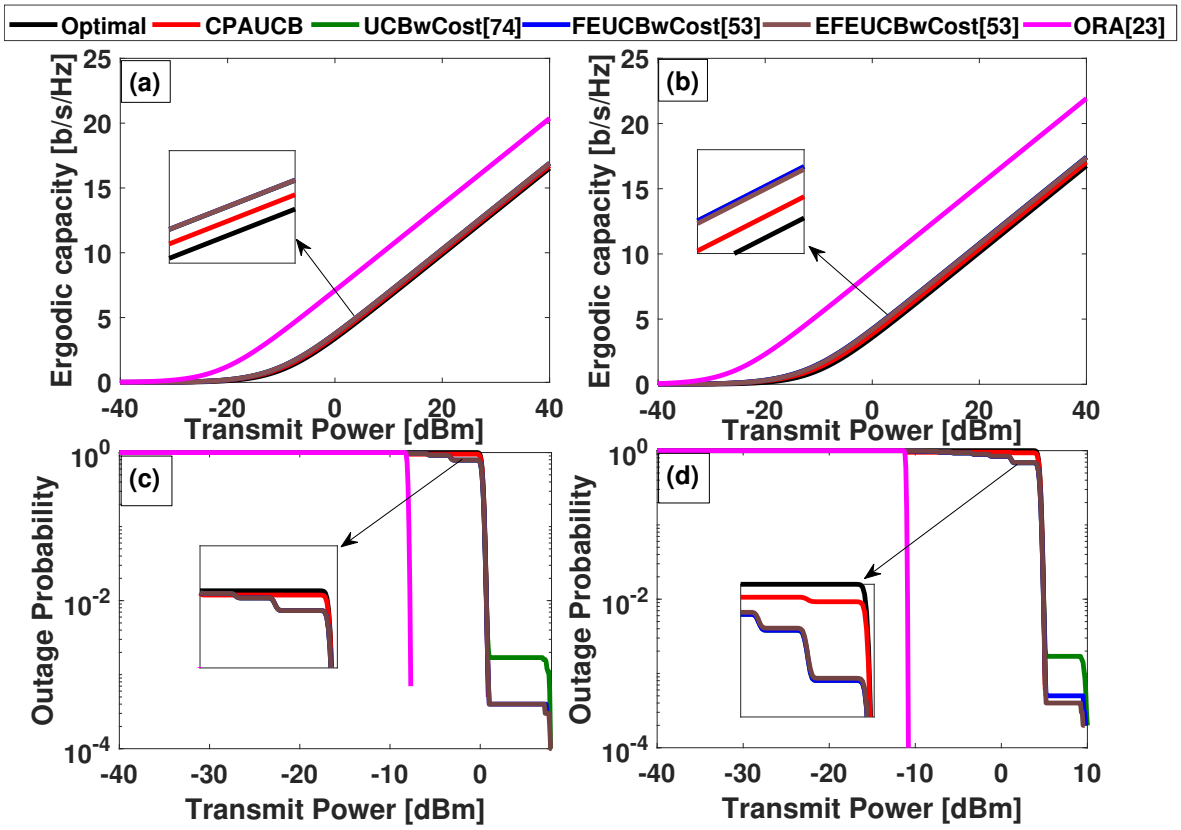

Figure 5.6: Comparison of Ergodic Capacity and Outage Probability with respect to Transmit Power for different algorithms for (a),(c) $K = 5, L = 5$ (b),(d)$K = 10, L = 5$

Similarly, we can see in Fig. 5.7 (c) and (d) where we have compared the outage probability with the consumed power. It is evident that the ORA scheme requires significantly higher power consumption compared to the learning-based approaches, owing to its selection of RIS blocks with the maximum number of sub-blocks, without optimizing for energy efficiency. Additionally, the impact of selecting sub-optimal blocks during the exploration phase

of the learning algorithms can be observed, as they gradually converge toward the optimal RIS configuration. However, the proposed CPAUCB algorithm demonstrates superior convergence characteristics, achieving zero outage probability at a considerably faster rate than the other learning-based schemes, highlighting its efficient trade-off between outage reduction and power consumption optimization.

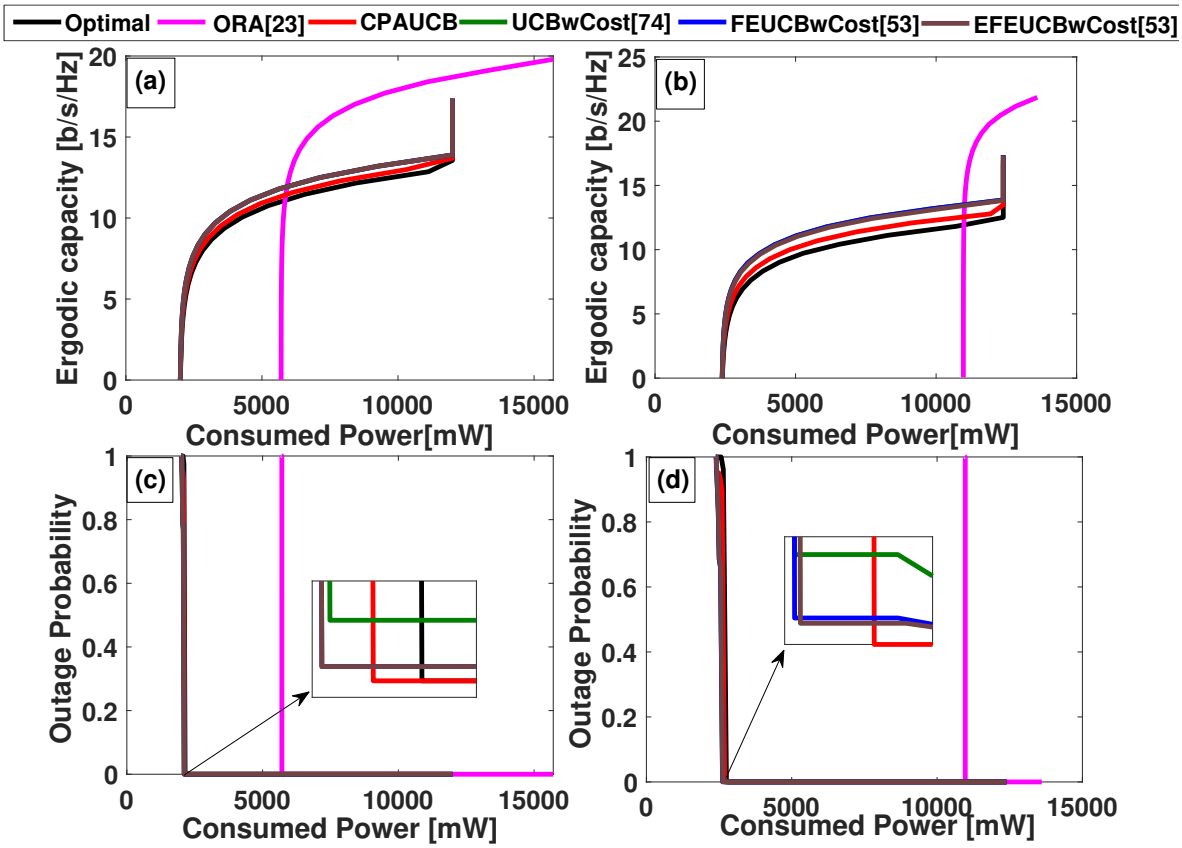

Figure 5.7: Comparison of Ergodic Capacity and Outage Probability with respect to Consumed Power for different algorithms for (a),(c) $K = 5, L = 5$ (b),(d)$K = 10, L = 5$

5.5.4 Execution Time

The execution time of MAB algorithms plays a vital role in applications like wireless networks. An ideal MAB algorithm should converge to selecting the optimal arm, especially when choosing RIS blocks while balancing power consumption and target SNR. In this work as shown in Tab.**5.2**, we evaluated the execution time of our proposed CPAUCB algorithm with the existing UCB, FEUCB and EFEUCB algorithms. Performance analysis were carried out on edge computing processors: (1) ARM Cortex-A9 (666 MHz) and (2) ARM Cortex-A9 (800

Table 5.2: Comparison of Execution Time on Edge Platforms

Sub-Blocks (M)	**Algorithm**	**Cortex A9 (666 MHz)**		**Cortex A9 and NEON (666 MHz)**		**Cortex A9 (800 MHz)**		**Cortex A9 and NEON (800 MHz)**	
		SPFP	**DPFP**	**SPFP**	**DPFP**	**SPFP**	**DPFP**	**SPFP**	**DPFP**
5	UCB	2.57	2.61	1.85	1.81	2.12	2.17	1.54	1.51
	FEUCB	2.26	2.29	1.52	1.48	1.88	1.91	1.26	1.23
	EFEUCB	2.22	2.28	1.5	1.47	1.87	1.90	1.25	1.22
	CPAUCB	5.51	5.54	4.18	4.27	4.58	4.59	3.48	3.55
10	UCB	5.04	5.13	3.67	3.57	4.18	4.27	3	2.97
	FEUCB	4.93	5	3.36	3.29	4.10	4.16	2.79	2.73
	EFEUCB	4.35	4.41	2.95	2.9	3.62	3.67	2.46	2.41
	CPAUCB	16.19	14.97	12.2	11.9	12.39	13.31	10	10.2

MHz). Additionally, the impact of the Single Instruction Multiple Data (SIMD) NEON co-processor was analyzed, along with the i different numerical precision Single Precision Floating Point (SPFP) and Double Precision Floating Point (DPFP)—on execution time, providing insights into computational efficiency across varying hardware configurations. Results show that EFEUCB achieves a 12˜14% reduction in execution time compared to UCB. This improvement is primarily due to the algorithm's focused exploration strategy, which reduces the number of arithmetic computations required per time slot. However, the proposed CPAUCB algorithm takes more time as compared to the existing algorithms, this is because of exploring each block which are lined up before the selected block, therefore the complexity increases.

Chapter 6

Online Learning and Change Detection based Multi-RIS-Aided Wireless Systems for Dynamic Environment

6.1 Overview

Reconfigurable intelligent systems offer on-the-fly control over the radio propagation environment, and various works have demonstrated the need for multiple RIS to support a wide range of mobile users (MU). In such a multi-RIS-aided wireless system, a multi-armed bandit based online learning has been explored to select the subset of RIS blocks for a given MUs. However, due to a large number of subblocks and dynamic environments where optimal RIS subblock changes over time, existing change detection-based MAB perform poorly. In this work, we propose a dynamic discounting and restarting (DDR) based enhanced focussed exploration based upper confidence bound (DDR-EFEUCB) algorithm. The DDR-EFEUCB addresses the challenge of a large number of subblocks via focussed exploration and the challenge of the dynamic

environment via the DDR approach without any prior knowledge of change intervals.

6.2 Introduction

In highly dynamic wireless environments—such as vehicular networks, UAV communication, or mobile cellular systems—the statistical properties of the communication channel can vary over time. This dynamicity renders standard stationary learning algorithms insufficient, motivating the need for online learning models that can adapt to changing reward distributions. Change detection in MABs is a prominent approach used to handle non-stationarity in reward structures.

The two dominant approaches for change adaptation in MABs are: active change detection and passive adaptation.

Active change detection relies on statistical hypothesis testing over sliding windows or cumulative sum (CUSUM) statistics. When a significant deviation in the observed rewards is detected compared to historical trends, the learner triggers a "restart" by resetting empirical estimates and reinitiating exploration. Notable algorithms like Discounted UCB, Sliding Window UCB, and EXP3.S use these principles (130; 131). These algorithms are effective in identifying abrupt shifts in environments, such as rapid user mobility or sudden link failures.

Passive change adaptation uses discount factors to assign exponentially decreasing weights to older observations. This allows the learner to "forget" outdated information gradually and prioritize recent data. Passive methods like Discounted Thompson Sampling and AdaUCB avoid abrupt resets but may adapt slowly to sharp changes. Their strength lies in handling smoothly evolving environments where transitions occur gradually.

However, applying these approaches to RIS-aided systems faces significant challenges:

The dimensionality of the action space is large due to the number of sub-blocks and combinations.

The frequency of changes may vary, requiring flexible thresholds and adaptive strategies.

The heterogeneity in reward impact means that not all changes are equally relevant (e.g.,

a minor SNR drop vs. a complete link outage).

Moreover, simultaneous change detection across multiple arms can lead to computational overhead and false positives. Recent works have begun exploring hybrid methods that combine active restart with passive decay, allowing nuanced adaptation to dynamic conditions.

In RIS-aided systems, where the channel between the transmitter, RIS, and receiver can fluctuate due to environmental mobility, such frameworks enable robust, low-latency adaptation of RIS configurations. This is particularly important in multi-user scenarios, where optimal configurations vary across spatial locations and times. The integration of change detection into scalable bandit algorithms presents a promising direction for intelligent wireless system design under realistic deployment conditions.

6.3 Network Model

We consider a multi-RIS-aided system consisting of a single transmitter and MU. Each of K RIS consists of L sub-blocks and each sub-block consists of Z sub-lambda-sized passive elements, as shown in Fig. 6.1. The RIS block may consist of single or multiple sub-blocks spanned across multiple RIS as shown in Fig. 6.1 where MU in the car is supported by multiple RIS and selected RIS block changes over time.

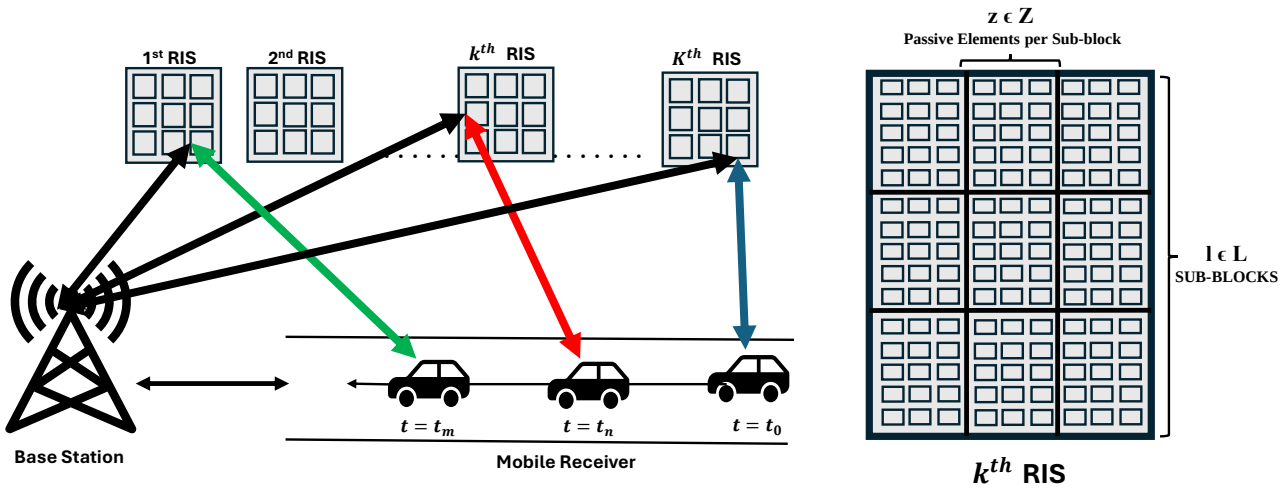

Figure 6.1: Illustrations of network model for multi-RIS-aided wireless system with mobile receiver for $n \in \{0, m\}$ indicates a change point.

We consider the channel model in (53), which assumes that the channels associated with sub-blocks of a given RIS are independent and identically distributed (i.i.d.), while for channels associated with the sub-blocks of different RISs, independence is assumed but not identical distributions. We consider Nakagami-m fading. The complex channel coefficient between the transmitter and the z-th element of the RIS sub-block, l, of the k^{th} RIS is defined in polar form as $G_{klz} = G_{klz}e^{j\psi_{klz}}$. Similarly, the channel between the z-th element of the RIS sub-block, l, and the receiver is also expressed in polar form as $H_{klz} = H_{klz}e^{j\phi_{klz}}$. For the direct path between the transmitter and the receiver, the channel coefficient is given as $G = Ge^{j\psi}$.

When transmitting a symbol x_s with power P_s (in dBm), the signal received through the RIS sub-block l, is:

$$y_{kl} = \sqrt{P_s}\left(G + \sum_{z=1}^{Z} G_{klz}\alpha_{klz}e^{j(\theta_{klz}+\phi_{klz})}H_{klz}\right)x_s + w \tag{6.1}$$

Here, α_{klz} represents the amplitude reflection coefficient, and θ_{klz} refers to the phase shift of the z-th element of the RIS. The term w corresponds to the Additive White Gaussian Noise (AWGN), which has zero mean and a variance of σ^2. With optimal RIS configuration (assuming zero phase error), the SNR, can be calculated as:

$$\text{SNR} = \frac{P_s}{\sigma^2}\left|G + \sum_{z=1}^{Z} G_{klz}\alpha_{klz}H_{klz}\right|^2. \tag{6.2}$$

The goal of the MAB algorithm is to explore and exploit the optimal RIS block with the highest SNR. In a dynamic environment, the horizon is divided into multiple sub-horizons of different unknown durations, and the task of the MAB algorithm is to detect the transitions between sub-horizons and identify the new optimal block as quickly as possible.

6.4 Proposed Work

6.4.1 Limitations of State-of-the-art

In (53), two algorithms, namely Focused Exploration UCB (FEUCB) and Enhanced Focused Exploration UCB (EFEUCB), were proposed. These algorithms are based on the concepts of sparsity (132) and thresholding (133). They aim to quickly learn and select the optimal RIS block from a large number of sub-blocks, outperforming the standard upper confidence bound (UCB) algorithm by reducing the regret bound from $\sqrt{RT}$ to $\sqrt{\hat{R}T}$, where R is the total number of arms i.e. $(R = KL)$ and T is the total time horizon and $\hat{R}$ is shortlisted set of top arms after focussed exploration stage. For the dynamic case with a large number of arms, the performance of UCB is expected to be poor due to large exploration time between change points, while it is not clear how FEUCB and EFEUCB can adapt due to focused exploration that restrict the exploration to subset of $\hat{R}$ arms. Thus, there is a need for a change detection mechanism for MAB algorithms with a large number of arms.

6.4.2 Proposed Algorithm

Identifying the need for unlearning the previously optimal arms post a change point we propose a new class of algorithms CD-EFEUCB and CD-FEUCB, which are equipped with active change-detectors to restart the learning once change in arm statistics has been observed by the change-detector. A necessary assumption in such cases is 1. that changes occur after the duration required for sufficient sampling of arms described by the value γ times the time-horizon as given in (53) that is we can only expect changes once initial thresholding of the top $\hat{R}$ arms has been done and 2. Changes occur in all arms simultaneously. We examine the change-detectors used in CD-EFEUCB and CD-FEUCB thoroughly in the subsequent sections.

We equip the thresholding based EFEUCB with an active change-detection algorithm, Monitored UCB or MUCB (134), which restarts the learning upon encountering a change-point. We begin by uniformly sampling all R arms once and entering the TCB stage of EFEUCB. Once

sufficient sampling, using the TCB algorithm (53) has been done, we carry out thresholding to perform a focused learning over "top" arms, $\hat{R}$. Here, we sample and locate the optimal arm using MUCB. This algorithm utilizes a windowing approach over the past rewards encountered to calculate the value of MCD 6.9. MUCB utilizes the UCB algorithm to sample the most optimal arm and then estimates the value of MCD by taking the difference of the summation of the second half window and first half window of expected rewards encountered uptil the current time instant. It then makes the decision if a change-point has been encountered. If the value of MCD exceeds the threshold value b, the algorithm flags a change point and restarts learning by refreshing its memory and re-populating the set of "top" arms.

In more realistic situations, a condition like a "weak change-point" (less significant change in the arm-statistics of a specific sub-block) might occur where even though the arm characteristics change, the previously optimal arms continue to remain optimal and there is less possibility of a previously sub-optimal arm to enter into the set $\hat{R}$ therefore $\hat{R}$ is expected to remain the same. Similarly, a "strong change-point" (significant change in the arm-statistics of a specific sub-block) indicates that one or more arms from the previous $\hat{R}$ set of arms have fallen to a below optimal behavior and one or more sub-optimal arms are now near-optimal arms. In case of weak change points, learning can continue, but it has to be restarted in the case of strong change points so that the latest optimal arms can be thresholded. Thus, we propose Dynamic Discounting and Restart EFEUCB (DDR-EFEUCB) which deploys a mixture of both Active and Passive Change Detection algorithms, which restarts learning only when it is needed and when a "strong" change-point is observed thus saving on regret by preventing unnecessary restarts.

We start the algorithm by initially exploring each of the $r \in R$ arm once **(lines 1-4)**. After sufficient exploration has been made using the TCB algorithm **(lines 6-8)** we enter the phase of thresholding **(line 9)** akin to EFEUCB. (53).

$$\mathrm{TCB}^r(t) = \sqrt{S^r(t)}\,|\mu_r(t)|\,. \tag{6.3}$$

here,

$$\mu_r(t) = \frac{\hat{X}^r(t)}{S^r(t)} \tag{6.4}$$

The equation that describes the thresholding is given by (**lines 8-9**) of the algorithm.

Once the thresholded set of optimal arms have been achieved in $\hat{R}$ we run our passive-change detector, based on discounting technique (131), to handle weak-changepoints **(lines 11)** and select the optimal arm I_t. This can be explained by the following equation as given in (130),(131).

$$DUCB_r(t) = \arg\max_r \left(\mu_r(t) + \sqrt{\frac{\max(\mu_r(t)(1-\mu_r(t)), \varepsilon) \log t}{S^r(t)_d}} \right) \tag{6.5}$$

where,

$$\hat{X}^r(t)_d = \sum_{\tau=0}^{t} \mathbb{I}(l_\tau = I_\tau) \gamma'^{t-\tau} \hat{X}^r(t), \tag{6.6}$$

$$S^r(t)_d = \sum_{\tau=0}^{t} \mathbb{I}(l_\tau = I_\tau) \gamma'^{t-\tau}, \tag{6.7}$$

$$\mu_r(t) = \frac{\hat{X}^r(t)_d}{S^r(t)_d}. \tag{6.8}$$

It is to be noted that γ' describes the discounting-factor of the DUCB algorithm.The active-change detector **MUCB** is then used to handle strong ones **(lines 12-13)** and restart learning by refreshing the algorithms memory of arm-statistics and number of pulls that is X^r and S^r respectively. The decision-making of MUCB is described as follows in (134).

$$\text{MCD}_r(t) = \left| \sum_{i=W/2+1}^{t} \mu_r(i) - \sum_{i=t-W}^{W/2} \mu_r(i) \right|, \text{where } t >= W \tag{6.9}$$

A change is detected whenever the value of MCD(t) exceeds a threshold b, for a window length W. In case no "strong-change point" is detected we continue our learning by performing the incremental step **(lines 14)**

Algorithm 5 DDR-EFEUCB

Input: R,T,γ,W,b,γ'

Initialize: $\hat{X}^r(t) \leftarrow 0$, $S^r \leftarrow 0$ for all r, $t' \leftarrow 0$, ts,$t' \leftarrow 1$

for $t = 1$ ***to*** T **do**

  if $t' \leq R$ **then**

    $I_t \leftarrow t'$

  else

    if $ts = 1$ **then**

      while $t \leq T$ & $t' \leq \gamma * T$ **do**

        $I_t \leftarrow \max(TCB_{r\in R}(t'))$ as in 6.3

      $\hat{R} := \left\{ r \in [R] \middle| \hat{X}^r(t) \geq 2\sqrt{\frac{\log(t)}{S^r(t)}} \right\}$

      $t_s = 0$

    else

      $I_t \leftarrow \max(DUCB_{r\in\hat{R}}(t'))$ as in 6.5

      if $MCD_{I_t}(t') > b$ **then**

        Restart learning:

        $t' \leftarrow 0$

        $t \leftarrow t+1$

        $\hat{X}^r(t) \leftarrow 0$ for all r

        $S^r \leftarrow 0$ for all r

        $ts \leftarrow 1$

  Increment: $t' \leftarrow t'+1$, $t \leftarrow t+1$ $S^r \leftarrow S^r + 1$, $\hat{X}^r(t) \leftarrow \hat{X}^r(t) + X_t^r$ Transmitter configures RIS block I_t Each receiver observes instantaneous normalized SNR and communicates to the transmitter

6.4.3 Mathematical Analysis

For the regret analysis of DDR-EFEUCB algorithm, an adaptive learning method based on the M-UCB framework, we refer to the analysis presented in (135).

Let the time horizon be denoted by T, and consider an environment with R arms and M piecewise-stationary segments, where the change-points are represented as $t_0, t_1, \ldots, t_M$, with $t_0 = 0$ and $t_M = T$. Each segment $[t_{m-1}, t_m)$ corresponds to a stationary interval during which the reward distribution remains fixed. Within each segment, the expected reward of arm $r \in \mathscr{R}$ is assumed to be constant. $\mu_r(t_k < t < t_{k+1})$ or μ_r^k lies between 0-1. We define a basic assumption that allows us to tune the parameters used in calculating 6.9 based on the analysis presented by (135).

Assumption 1. The learning agent can choose w, the window length for M-UCB and γ, the factor of time-horizon which is required for sufficient sampling, such that:

(a) $M < \lfloor T/L \rfloor$ and $t_{i+1} - t_i > L, \forall 0 \leq i \leq M-1$, and

(b) $\forall 1 \leq i \leq M-1, \exists r \in \mathscr{R}$ such that:

$$\delta_r^{(i)} \geq 2\sqrt{\log(2RT^2)/w} + 2\sqrt{\log(2T)/w}.$$

where L is given by w$\lceil R/\gamma \rceil and the amplitude of the change of arm r at the ith change - point as \delta_r^{(i)} = |\mu_r^{i+1} - \mu_r^i|, \quad \forall 1 \leq i \leq M-1, r \in \mathscr{R}$.(6.10)

It should be noted that this assumption is only necessary for the regret analysis and the proposed algorithm can be implemented regardless. We now come to our main result, the overall regret bound of DDR-EFEUCB.

Theorem 1. Regret/Cost of running exploration in DDR-EFEUCB can be evaluated due to two prominent cases **(53)**:

1. $A_r(t)$: Cost/Regret an event of RIS sub-block, r, being selected during the D-UCB phase when $\hat{R}$ sub-blocks are sampled sufficiently .

2. $B_r(t)$: Cost/Regret an event of RIS sub-block, r, being selected during the UCB phase

when $\hat{R}$ sub-blocks are not sampled sufficiently.

Therefore

$$\sum_{i=1}^{M} \tilde{C}_i \leq A_r(t) + B_r(t) \tag{6.11}$$

Alternatively we know that the expected regret can be written as**(53)**,

$$\mathbb{E}[R(T)] \leq \log(T) \sum_{r \in [\hat{\mathscr{R}}] \setminus r_*} \frac{1}{\Delta_r}. \tag{6.12}$$

Where,

$$\begin{aligned} \mathbb{E}[R(T)] = 1 + \sum_{r \in \hat{\mathscr{R}} \setminus r_*} \Delta_r \mathbb{E}\left[\sum_{t=1}^{T} \mathbf{1}_{A_r(t)}\right] \\ + \sum_{r \in \hat{\mathscr{R}} \setminus r_*} \Delta_r \mathbb{E}\left[\sum_{t=1}^{T} \mathbf{1}_{B_r(t)}\right] \end{aligned} \tag{6.13}$$

Thus, we can rewrite (6.11) as;

$$\begin{aligned} \sum_{i=1}^{M} \tilde{C}_i \leq \sum_{r \in \hat{\mathscr{R}} \setminus r_*} \Delta_r \mathbb{E}\left[\sum_{t=1}^{T} \mathbf{1}_{A_r(t)}\right] \\ + \sum_{r \in \hat{\mathscr{R}} \setminus r_*} \Delta_r \mathbb{E}\left[\sum_{t=1}^{T} \mathbf{1}_{B_r(t)}\right] \end{aligned} \tag{6.14}$$

Now we may find the regret-bounds of both the events individually as in **Lemma 1** and **Lemma 2** and rewrite (6.14) to achieve cummulative-regret without considering restarts.

Lemma 1: The regret due to event $A_r(t)$ can be written as described in (130) where γ' gives the discounting-factor

$$\begin{aligned} \sum_{r \in \hat{\mathscr{R}} \setminus r_*} \Delta_r \mathbb{E}\left[\sum_{t=1}^{T} \mathbf{1}_{A_r(t)}\right] \leq \sum_{r \in \hat{\mathscr{R}} \setminus r_*} \Delta_r \Big(B(\gamma') T (1-\gamma') \log(\frac{1}{1-\gamma'}) \\ + C(\gamma') \frac{\Upsilon_T}{1-\gamma'} \log\left(\frac{1}{1-\gamma'}\right) \Big) \end{aligned} \tag{6.15}$$

Where:

$$B(\gamma') = \frac{16B^2\xi}{\gamma'^{1/(1-\gamma')}(\Delta_r)^2} \cdot \frac{\lceil T(1-\gamma')\rceil}{T(1-\gamma')} + \frac{2}{C_1 \cdot \log(1-\gamma')} \cdot C_2 \tag{6.16}$$

$$C_1 = \left\lceil \frac{-\log(1-\gamma')}{\log\left(1+4\sqrt{1-\frac{1}{2\xi}}\right)} \right\rceil, \qquad C_2 = \left(1-\gamma'^{1/(1-\gamma')}\right) \tag{6.17}$$

$$C(\gamma') = \frac{\gamma'-1}{\log(1-\gamma')\log\gamma'} \cdot \log\left((1-\gamma')\xi \log n_K(\gamma')\right) \tag{6.18}$$

For the event $B_r(t)$, the regret bounds would be similar to the FEUCB algorithm as presented in (53), the EFEUCB algorithm's exploration-exploitation balance is enhanced by this thresholding strategy, which reduces the number of number of arms, thereby accelerating convergence to the optimal configuration compared to traditional methods such as FEUCB. However, the nonlinear nature of the thresholding operation introduces challenges in deriving a closed-form expression for the algorithm's regret. **Theorem 2.** DDR-EFEUCB, with w and γ satisfying **Assumption 1** and $b = [w\log(2KT^2)/2]^{1/2}$, has the following upper-bound when restarts are taken into consideration.

$$\begin{aligned} R(T) \leq & \underbrace{\sum_{i=1}^{M} \tilde{C}_i}_{\text{(a)}} + \underbrace{\gamma T}_{\text{(b)}} \\ & + \underbrace{\sum_{i=1}^{M-1} 2K \cdot \min\left(\frac{w}{2}, \frac{\lceil b/\delta^{(i)}\rceil}{\gamma} + 3\sqrt{w}\right)}_{\text{(c)}} + \underbrace{3M}_{\text{(d)}}. \end{aligned} \tag{6.19}$$

where $\delta^{(i)} = \max_{k\in\mathscr{K}} \delta_k^{(i)}$ and $\tilde{C}_i$ is given by **Theorem 1**. **Theorem 2** reveals that the regret incurred by DDR-EFEUCB can be decomposed into four terms:

- bounds the cost of the EFEUCB-based exploration given by **Theorem 1**,
- bounds the cost of the uniform sampling required after each restart caused due to change

detection,

- bounds the cost associated with the detection delay of the CD algorithm as given in (134), and described in **Lemma 3**,and

- is incurred by the unsuccessful and incorrect detection of the $M+1$ change-points, where M indicates the number of segments caused due to change-pointsas described in **Lemma 4**.

Lemma 3: Consider the M-UCB aided DDR-EFEUCB algorithm operating under Assumption 1, with parameters w, $b = \sqrt{\frac{w}{2}\log(2KT^2)}$, and γ. Let τ_i denote the detection time of the i^{th} change-point ν_i. Then, the expected cumulative regret incurred by detection delays over all $M-1$ change-points is bounded by:

$$\mathbb{E}\left[\sum_{i=1}^{M-1}(\tau_i - \nu_i)\right] \leq \sum_{i=1}^{M-1} \frac{2K \cdot \min\left(\frac{w}{2}, \left\lceil \frac{b}{\delta^{(i)}} \right\rceil + 3\sqrt{w}\right)}{\gamma}, \tag{6.20}$$

where $\delta^{(i)} = \max_{k \in \mathscr{K}} |\mu_k^{(i+1)} - \mu_k^{(i)}|$ denotes the maximum magnitude of change at the i^{th} change-point. **Proof:** Lemma 4 in (134) explicitly provides the bound for the expected detection delay conditioned on successfully detecting the change. Specifically, Lemma 4 states:

$$\mathbb{E}[\tau_i - \nu_i \mid \nu_i < \tau_i \leq \nu_i + L/2] \leq \frac{\min\left(\frac{L}{2}, \left\lceil \frac{b}{\delta^{(i)}} \right\rceil + 3\sqrt{w} \cdot \left\lceil \frac{K}{\gamma} \right\rceil\right)}{1 - 2\exp(-wc^2/4)}, \tag{6.21}$$

where $L = w\left\lceil \frac{K}{\gamma} \right\rceil$ and $c = 2\sqrt{\frac{\log(2T)}{w}}$ is chosen to ensure high-probability successful detection.

Given **Assumption 1** and the choice of parameters b and c, we ensure that the denominator $1 - 2\exp(-wc^2/4)$ is very close to 1, thus simplifying the expected detection delay to:

$$\mathbb{E}[\tau_i - \nu_i \mid \nu_i < \tau_i \leq \nu_i + L/2] \leq \min\left(\frac{L}{2}, \left\lceil \frac{b}{\delta^{(i)}} \right\rceil + 3\sqrt{w} \cdot \left\lceil \frac{K}{\gamma} \right\rceil\right). \tag{6.22}$$

Since DDR-EFEUCB uses uniform sampling to ensure detection within a window of length $L/2$ with probability at least $1 - \frac{1}{T}$, the total expected regret from detection delays for all

$M-1$ change-points can be summed, yielding:

$$\mathbb{E}\left[\sum_{i=1}^{M-1}(\tau_i - \nu_i)\right] \leq \sum_{i=1}^{M-1} \frac{2K \cdot \min\left(\frac{w}{2}, \left\lceil \frac{b}{\delta^{(i)}} \right\rceil + 3\sqrt{w}\right)}{\gamma}. \tag{6.23}$$

This completes the proof, giving us the cumulative regret term due to detection delays.

Lemma 4: Consider the M-UCB aided DDR-EFEUCB algorithm operating under Assumption 1, with parameters w, $b = \sqrt{\frac{w}{2}\log(2KT^2)}$, and γ. Let τ_i denote the detection time of the i^{th} change-point ν_i. Then, the expected cumulative regret incurred by incorrect detections (false alarms) over all M stationary segments is bounded by:

$$\mathbb{E}\left[\text{Regret due to incorrect detections}\right] \leq 3M. \tag{6.24}$$

Proof: Consider a stationary scenario (no true change-points), and define τ_1 as the first detection time of a false alarm. Lemma 2 in (134) provides an explicit upper bound on the probability of incorrectly raising a false alarm:

$$P(\tau_1 \leq T) \leq wK\left(1 - \left(1 - 2\exp\left(-\frac{2b^2}{w}\right)\right)^{\lfloor T/w \rfloor}\right). \tag{6.25}$$

Using the inequality $(1-x)^a \geq 1 - ax$ for $0 < x < 1$ and $a > 1$, we simplify the probability bound:

$$P(\tau_1 \leq T) \leq wK \cdot \frac{2T}{w}\exp\left(-\frac{2b^2}{w}\right) \tag{6.26}$$

$$= 2KT\exp\left(-\frac{2b^2}{w}\right). \tag{6.27}$$

Substituting the choice of $b = \sqrt{\frac{w}{2}\log(2KT^2)}$ yields:

$$P(\tau_1 \leq T) \leq 2KT\exp\left(-\frac{2}{w}\cdot\frac{w}{2}\log(2KT^2)\right) \tag{6.28}$$

$$= 2KT \cdot \frac{1}{2KT^2} = \frac{1}{T}. \tag{6.29}$$

Therefore, the expected number of false alarms during each stationary segment (with length at most T) is at most 1. Since the total number of segments is M, and each false alarm contributes at most a constant regret of 3 (due to immediate resets and short-term suboptimal actions), the cumulative regret due to incorrect detections is thus bounded by:

$$\mathbb{E}[\text{Regret due to incorrect detections}] \leq 3M. \tag{6.30}$$

6.5 Performance Analysis

In this section, we present the synthetic results and its analysis to validate our proposed algorithm and compare its performance with the existing state-of-the-arts i.e.,UCB, FEUCB and EFEUCB algorithms equiped with Change Detectors (74; 53). We compare their performance in terms of outage probability, ergodic capacity, energy efficiency and regret.

We consider a MU, and $K = 5$ RIS. Each RIS is divided into $L = 5$ sub-blocks. The horizon size, T, is between 12000. Each result is averaged over 15 independent experiments over the selected horizon size. In each experiment, the positions of the transmitter, RIS are fixed, and the receiver changes its position in a piecewise stationary manner.

For an accurate consideration of possible scenarios of user-mobility we formulate three different cases while performing the analysis.

- Case 1: Receiver changes its position such that each RIS sub-block witnesses a trivial increase or decrease in performance in a piecewise-stationary manner indicating weak changepoints.
- Case 2:Receiver changes its position such that few RIS-sub blocks observe a non-trivial performance degradation, indicating strong changepoints, while some RIS-sub blocks see a non-trivial improvement, thus a shift in optimal arms is certain and focused exploration must be repeated to locate the optimal arm which might lie outside the existing set of thresholded arms.
- Case 3: Receiver changes its position randomly, thus indicating a random combination of weak and strong changepoints

The Fig. 6.2 illustrates how the mean (μ) of arms changes across different scenarios, simulating the effect of receiver movement on the RIS-assisted communication system(only for illustration purposes). In Case 1, the receiver shifts position in a manner that causes each RIS sub-block to experience only a slight, piecewise-stationary variation in performance. These minor fluctuations correspond to weak change points, where the optimal arm likely remains unchanged or shifts subtly. In Case 2, the receiver's movement induces significant performance degradation in some RIS sub-blocks and notable improvements in others. These strong change points result in a definitive shift in the optimal arm, necessitating renewed exploration, especially outside the previously shortlisted set of arms. Finally, in Case 3, the receiver moves unpredictably, causing a mixture of weak and strong change points across the RIS sub-blocks. This scenario represents a more chaotic environment, requiring a bandit strategy robust to both subtle and drastic shifts in arm performance.

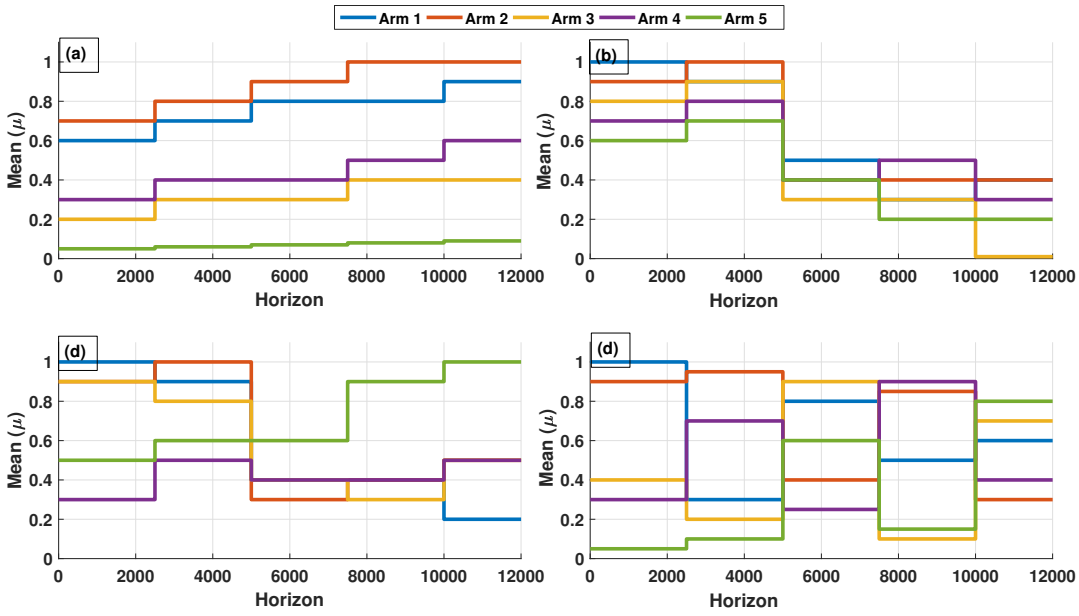

Figure 6.2: Illustration of the change in the mean (μ) of arms after each change point.

6.5.1 Regret Analysis

We begin our analysis by benchmarking the proposed algorithm against other mentioned alternatives by performing a comparison of their cumulative regret in a slotted time interval over a horizon of 12000 time samples. On plotting the regret curves of existing techniques like

UCB, FEUCB and EFEUCB, we observe significantly large values of regret due to absence of mechanisms for change-point detection. Since the necessity of change-point handling mechanisms is evident, we compare the variants of FEUCB and EFEUCB with change-detectors with the proposed algorithm all of which outperform UCB and other variants. It is observed that in Case 1, DDR-EFEUCB significantly outperforms existing thresholding techniques with change-detectors like CD-FEUCB and CD-EFEUCB as the earlier refrains itself from restarting on change-points that do not cause the thresholded set of arms to be outdated. DDR-EFEUCB only restarts when it suspects the possibility of a previously sub-optimal arm being optimal where as the other change-detection algorithms restart at each change-point forcing the learning algorithm to undergo thresholding again and again unnecessarily. Similarly in Case 2 Fig. 6.3 (c), DDR-EFEUCB is able to detect change-points which are followed by shift in behavior of a sub-optimal arm to being optimal and vice versa. DDR-EFEUCB only restarts learning and thresholding in such "strong" change-points thus minimizing regret as compared to other Change-Detector-aided-thresholding-based algorithms which restart learning at both "strong" and "weak" change-points .

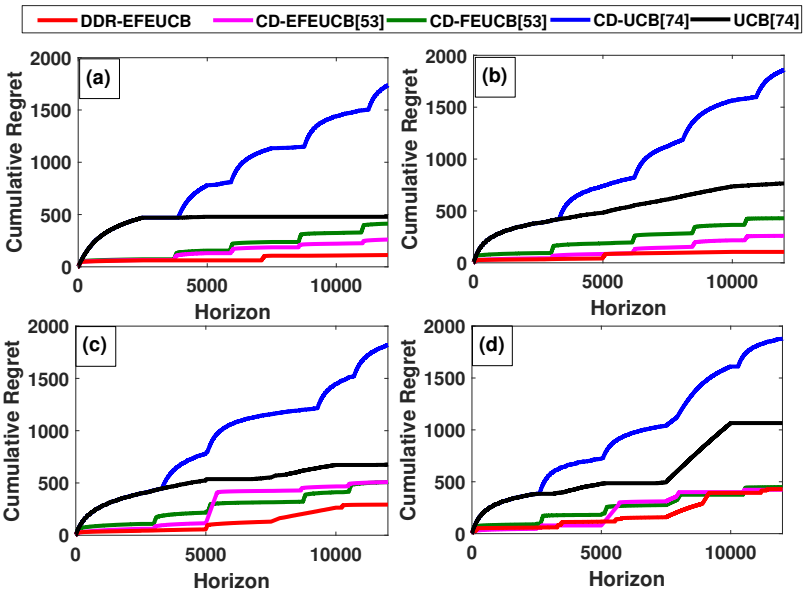

Figure 6.3: Regret Analysis of Proposed Algorithms in a Changing Environment

6.5.2 Comparison of Achievable Rate, Outage Probability, Ergodic Capacity and Energy Efficiency

In this sub-section, we analyze the effect of the transmit power on the achievable rate. As shown in Fig. 6.4 (a), different algorithms need different transmit power to achieve a given rate. All learning-based approaches need a similar transmit power, validating the successful learning and frequent selection of optimal RIS. However, our proposed DDR-EFEUCB algorithm achieves the maximum achievable rate at a given transmit power compared to CD-EFEUCB,CD-FEUCB and CD-UCB because of its superior adaptability to changes in the receiver's location.

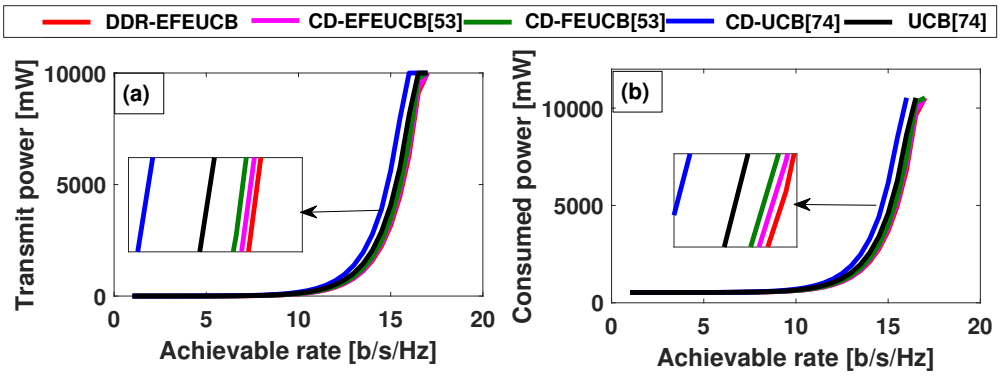

Figure 6.4: Comparison of Achievable Rate with (a) Transmit Power and (b) Consumed Power

Another notable remark is that, contrary to the intuition, CD-UCB requires a higher transmit power due to the absence of sufficient time-samples to perform learning-restarts, making it difficult for the algorithm to locate the optimal-arm in time. Thus, for larger number of arms, the presence of a Change-Detector in a non-thresholding traditional UCB algorithm is counter-productive. Similarly in Fig. 6.4 (b), DDR-EFEUCB offers the maximum achievable rate at a given consumed power compared to other algorithms.

Next we compare the outage probability, which is significant in the optimization and design of wireless communication systems since it helps to balance system capacity and quality of service trade-offs. In Fig. 6.5 (a), outage probability has been measured against the transmit power, and as expected, the outage probability decreases with the increase in the transmit power. Since the most optimal sub-block is being selected by the DDR-EFEUCB algorithm more often, the transmit power required by DDR-EFEUCB is less than other learning algorithms, followed by CD based EFEUCB and FEUCB algorithms. And finally, CD based UCB

consume maximum transmit power to reach lower values of outage probability. This is because UCB is able to adapt its learning post change-points and shift focus to the newly optimal arm before UCB-CD undergoes a restart and sufficient exploration, which increases the selection of sub-optimal arms. As the DDR-EFEUCB utilizes least transmit power to reach lower values of outage probability, the total consumed power required by DDR-EFEUCB is also the least as shown in Fig. 6.5 (b), followed by other learning algorithms, and as expected the CD based UCB algorithms consumes the maximum total consumed power.

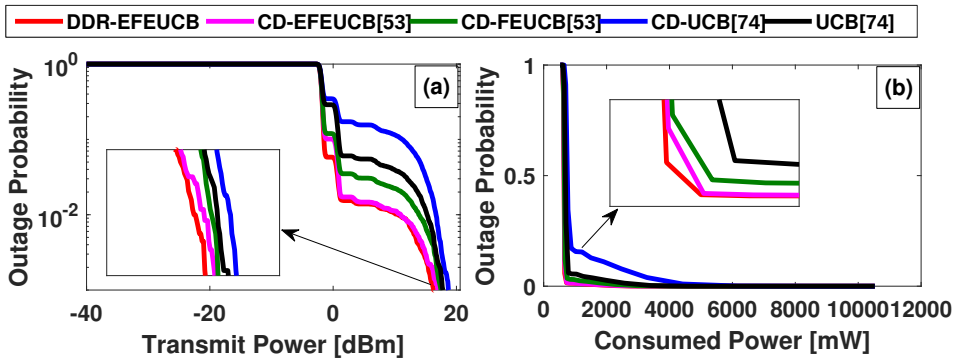

Figure 6.5: Comparison of Outage Probability with (a) Transmit Power and (b) Consumed Power

Similarly, in Fig. 6.6, we compare the ergodic capacity with respect to both transmit power and consumed power. As expected, the DDR-EFEUCB algorithm achieves the highest ergodic capacity due to its ability to adapt quickly to dynamic environments and learn the optimal RIS sub-blocks in each piecewise stationary environment. It efficiently identifies and selects the sub-blocks that maximize the ergodic capacity, enabling better performance compared to other methods. Furthermore, for a given consumed power, the DDR-EFEUCB algorithm demonstrates remarkable efficiency by achieving the maximum ergodic capacity while maintaining optimal resource utilization. This adaptability and power efficiency make it particularly suitable for scenarios where power constraints are critical, ensuring consistent outperformance under varying transmit and consumed power levels. This can also be shown in Fig. 6.7, the DDR-EFEUCB algorithm not only maximizes ergodic capacity but also demonstrates superior energy efficiency. By adapting to varying conditions and selecting the optimal RIS sub-blocks, it ensures that the achieved capacity is maximized for a given consumed power. This efficient utilization of power resources minimizes energy wastage, making DDR-EFEUCB particularly advantageous in power-constrained scenarios. Its ability to balance high performance and en-

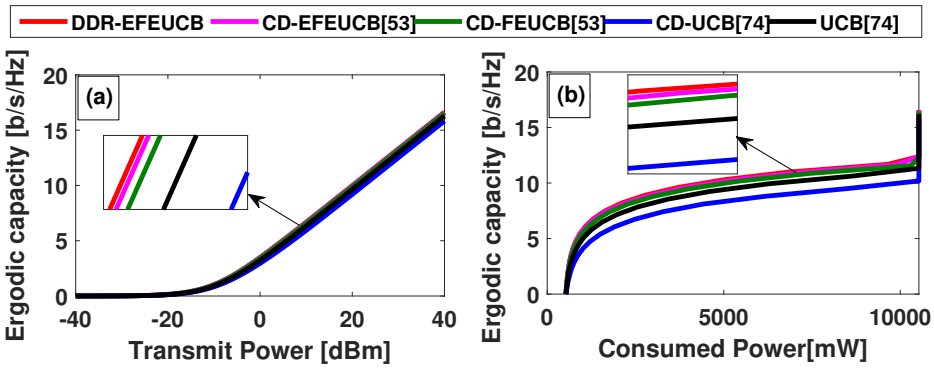

Figure 6.6: Comparison of Ergodic Capacity with (a) Transmit Power and (b) Consumed Power

ergy consumption enables sustainable operation in dynamic environments. Compared to other methods, DDR-EFEUCB achieves a higher capacity-to-power ratio, highlighting its potential for energy-efficient applications in modern wireless communication and sensing systems where power efficiency is paramount.

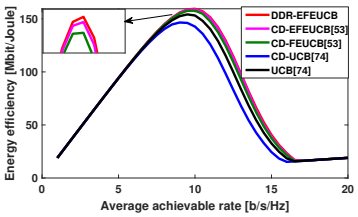

Figure 6.7: Comparison of energy efficiency and average achievable rate for different algorithms

6.5.3 Timing Analysis on Edge Platforms

In practical edge deployments of MAB algorithms for RIS configuration, minimizing execution time is crucial for real-time decision making in time-slotted wireless communication. A faster algorithm enables quicker RIS sub-block selection, allowing more time for actual data transmission and thus improving system throughput. Table 6.1 presents a detailed comparison of execution time (in milliseconds) for different MAB algorithms implemented using single-precision floating-point (SPFL) arithmetic on three types of widely-used ARM Cortex processors: 1) Cortex-A9 at 666 MHz, 2) Cortex-A9 at 800 MHz, and 3) Cortex-A53 at 1.1 GHz.

We have compared CD-based UCB, FEUCB, and EFEUCB, along with our proposed

novel DDR-EFEUCB algorithm, against the baseline UCB, FEUCB, and EFEUCB algorithms to assess the impact of change-detection based learning on execution time. As shown in Table 6.1, the CD-based variants (CD-UCB, CD-FEUCB, and CD-EFEUCB) exhibit a slight increase in execution time compared to their non-CD counterparts, likely due to the additional overhead introduced by the change detector. Notably, the proposed DDR-EFEUCB incurs the highest execution time among all algorithms. This increase is attributed to the integration of discounting and both active and passive change detection mechanisms, which add computational complexity to support improved adaptability in dynamic environments.

Table 6.1: Performance Comparison of Different Algorithms on Cortex Architectures (Sub Blocks = 25)

S.No.	Algorithm	Arm Cortex Architectures		
		A9 (666 MHz)	A9 (800 MHz)	A53 (1.1 GHz)
1	UCB(74; 53)	12.73	12.42	5.86
2	FEUCB(53)	8.49 **(-33.31%)**	8.29 **(-33.25%)**	3.89 **(-33.62%)**
3	EFEUCB(53)	4.49 **(-64.73%)**	4.35 **(-64.98%)**	2.10 **(-64.16%)**
4	CD-UCB	15.00 **(+17.83%)**	14.23 **(+14.57%)**	7.43 **(+26.79%)**
5	CD-FEUCB	11.88 **(-6.68%)**	11.61 **(-6.52%)**	4.60 **(-21.50%)**
6	CD-EFEUCB	4.82 **(-62.14%)**	4.56 **(-63.29%)**	2.21 **(-62.29%)**
7	DDR-EFEUCB	16.96 **(+33.23%)**	16.53 **(+33.09%)**	7.80 **(+33.11%)**

6.6 Summary

In this work, the DDR-EFEUCB algorithm has been proposed along with a new class of change-detectors equipped EFEUCB and FEUCB to efficiently select the optimal RIS sub-blocks in a piecewise stationary environment. Through intelligent adaptation and learning, DDR-EFEUCB ensures improved performance in dynamic conditions. Extensive and in-depth simulation results validate the effectiveness and demonstrate the superiority of the proposed algorithm compared to existing state-of-the-art methods in terms of ergodic capacity and energy efficiency. These results highlight its potential for real-world applications. Future work will focus on the hardware implementation of these algorithms to evaluate their performance in real-time, dynamically varying environments.

Chapter 7

Conclusions and Future Works

This thesis has presented a comprehensive investigation into online learning-based optimization strategies for Reconfigurable Intelligent Surfaces (RIS) in dynamic wireless communication environments. Across multiple contributions, the thesis addressed the central challenge of scalable and adaptive RIS configuration in the face of high-dimensional configuration spaces, limited feedback, and changing channel conditions. Through novel algorithm design, theoretical analysis, extensive simulation studies, and hardware-oriented implementations, the work has laid the foundation for efficient real-time decision-making in RIS-assisted wireless systems.

7.1 Conclusion

In **Chapter 3**, we proposed two novel distributed learning algorithms, *Focused Exploration UCB (FEUCB)* and its enhanced variant *Enhanced FEUCB (EFEUCB)*, for selecting optimal RIS sub-blocks. These algorithms were designed to operate in large action spaces under limited feedback, leveraging the exploration-exploitation tradeoff of the Multi-Armed Bandit (MAB) framework. Theoretical regret analysis and extensive simulation results established the superior

performance of these algorithms over traditional UCB and -greedy approaches. Furthermore, implementation on embedded edge platforms demonstrated their computational efficiency, with significant reductions in execution time. This work validated the viability of distributed online learning for RIS configuration and motivated future extensions toward hardware-software co-design and deployment in quasi-stationary vehicular environments. Future directions also include the design of directional analog front-ends and antenna architectures to support multi-RIS communication.

In **Chapter 4**, we investigated the problem of bandit learning at scale, focusing on MAB algorithms designed for large arm sets. A hardware-aware implementation of these algorithms was developed on a Xilinx Zynq SoC, where performance bottlenecks related to computation and memory were addressed through algorithm-architecture co-design. The proposed architecture achieved a **67%** reduction in cumulative regret, **97%** improvement in execution time, and **10%** resource savings compared to state-of-the-art MAB implementations for 100 arms. These gains were made possible by leveraging focused exploration techniques that limit candidate arms early in the learning process. The results emphasize the importance of hardware-software synergy in scaling learning algorithms for practical deployment. Future work in this direction will involve integrating change detection capabilities to support dynamic environments and adaptively adjust exploration strategies based on reward distribution shifts.

In **Chapter 5**, we introduced a power-aware RIS sub-block selection algorithm called *Consumed Power Aware UCB (CPAUCB)*. This algorithm is based on a sensor selection analogy, where RIS sub-block groups are treated as sensors of varying cost (power consumption) and accuracy (SNR). The CPAUCB algorithm aims to strike a balance between power efficiency and communication performance. Rigorous regret analysis and simulations demonstrated its effectiveness in reducing power consumption while achieving competitive SNR performance. The distributed nature of the algorithm makes it especially suitable for edge deployments with multiple RIS panels. Future research will extend CPAUCB for dynamic vehicular environments, incorporating detection of environment changes to enable adaptive sub-block selection.

In **Chapter 6**, we addressed the challenge of dynamically varying channel environments through the *Dynamic Discounted Restart FEUCB (DDR-EFEUCB)* algorithm. This contribution presented a class of FEUCB-based algorithms equipped with change detection mechanisms—both active and passive—to handle piecewise-stationary reward processes. DDR-

EFEUCB allows the system to autonomously detect significant shifts in RIS performance due to environmental changes and adapt its configuration policy accordingly. Simulation studies across multiple dynamic scenarios showed DDR-EFEUCB to outperform existing baselines in ergodic capacity, regret, and power efficiency. These results underscore the robustness of the proposed learning framework under real-world dynamics. The next step involves implementing DDR-EFEUCB on embedded platforms to evaluate its latency and energy performance in practical deployments.

7.2 Future Work

While the current work advances the state of the art in learning-based RIS optimization, several promising research directions remain open:

- **Joint Optimization with Beamforming:** Integrating RIS configuration with active beamforming strategies at the transmitter and receiver could unlock synergistic performance gains. Such joint design is particularly valuable in massive MIMO and coordinated multipoint (CoMP) networks, where spatial diversity and directionality can be exploited for both energy and spectral efficiency.

- **Contextual and Structured Bandits:** Future research should incorporate side information (context) such as user location, channel statistics, or quality-of-service constraints into the decision-making process. Structured bandits can further exploit correlations among arms to improve sample efficiency, making the system responsive to heterogeneous environments.

- **Hardware-in-the-Loop Prototyping:** A critical step for real-world adoption involves evaluating the proposed algorithms on hardware-in-the-loop testbeds with real channel emulation. FPGA, SoC, and mixed-signal platforms should be explored to benchmark latency, power, and scalability. Such platforms could validate RIS optimization under realistic feedback, noise, and propagation delays.

- **Multi-Agent Learning in Cooperative RIS Networks:** As RIS deployment becomes dense, future systems will likely feature multiple RIS panels managed by decentralized

controllers. Cooperative and federated learning approaches could be employed to coordinate RIS behaviors under communication constraints and partial observability.

- **RIS in ISAC (Integrated Sensing and Communication):** Future networks aim to unify communication and sensing functionalities. Extending MAB-based RIS optimization to also enhance sensing metrics (e.g., radar cross-section, angular resolution) introduces a new class of multi-objective bandit problems with potential for military, vehicular, and smart infrastructure use-cases.

This thesis thus contributes a comprehensive suite of algorithms, analysis, and implementations for learning-based RIS optimization. It sets the stage for a new generation of intelligent and adaptive wireless systems capable of meeting the demands of 6G and beyond.

References

[1] Y. Liu, X. Liu, X. Mu, T. Hou, J. Xu, M. Di Renzo, and N. Al-Dhahir, "Reconfigurable intelligent surfaces: Principles and opportunities," *IEEE Communications Surveys Tutorials*, vol. 23, no. 3, pp. 1546–1577, 2021.

[2] 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, 2023.

[3] 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, 2020.

[4] J. Huang, C.-X. Wang, Y. Sun, R. Feng, J. Huang, B. Guo, Z. Zhong, and T. J. Cui, "Reconfigurable intelligent surfaces: Channel characterization and modeling," *Proceedings of the IEEE*, vol. 110, no. 9, pp. 1290–1311, 2022.

[5] M. Ahmed, A. Wahid, W. U. Khan, F. Khan, A. Ihsan, Z. Ali, K. M. Rabie, T. Shongwe, and Z. Han, "A survey on ris advances in terahertz communications: Emerging paradigms and research frontiers," *IEEE Access*, vol. 12, pp. 173867–173901, 2024.

[6] S. Basharat, S. A. Hassan, H. Pervaiz, A. Mahmood, Z. Ding, and M. Gidlund, "Reconfigurable intelligent surfaces: Potentials, applications, and challenges for 6g wireless networks," *IEEE Wireless Communications*, vol. 28, no. 6, pp. 184–191, 2021.

[7] K. Keykhosravi, M. F. Keskin, S. Dwivedi, G. Seco-Granados, and H. Wymeersch, "Semi-passive 3d positioning of multiple ris-enabled users," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 10, pp. 11073–11077, 2021.

[8] M. Luan, B. Wang, Y. Zhao, Z. Feng, and F. Hu, "Phase design and near-field target localization for ris-assisted regional localization system," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 2, pp. 1766–1777, 2022.

[9] H. D. Tuan, A. A. Nasir, E. Dutkiewicz, H. V. Poor, and L. Hanzo, "Ris-aided multiple-input multiple-output broadcast channel capacity," *IEEE Transactions on Communications*, vol. 72, no. 1, pp. 117–132, 2024.

[10] H. Luo, R. Liu, M. Li, Y. Liu, and Q. Liu, "Joint beamforming design for ris-assisted integrated sensing and communication systems," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 12, pp. 13393–13397, 2022.

[11] E. Shtaiwi, H. Zhang, S. Vishwanath, M. Youssef, A. Abdelhadi, and Z. Han, "Channel estimation approach for ris assisted mimo systems," *IEEE Transactions on Cognitive Communications and Networking*, vol. 7, no. 2, pp. 452–465, 2021.

[12] H. Guo and V. K. N. Lau, "Uplink cascaded channel estimation for intelligent reflecting surface assisted multiuser miso systems," *IEEE Transactions on Signal Processing*, vol. 70, pp. 3964–3977, July 2022.

[13] Z. Chu, J. Zhong, P. Xiao, D. Mi, W. Hao, R. Tafazolli, and A. P. Feresidis, "Ris assisted wireless powered iot networks with phase shift error and transceiver hardware impairment," *IEEE Transactions on Communications*, vol. 70, no. 7, pp. 4910–4924, 2022.

[14] 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, 2023.

[15] M. Wang, W. Duan, G. Zhang, M. Wen, J. Choi, and P.-H. Ho, "On the achievable capacity of cooperative noma networks: Ris or relay?," *IEEE Wireless Communications Letters*, vol. 11, no. 8, pp. 1624–1628, 2022.

[16] E. Basar and I. Yildirim, "Reconfigurable intelligent surfaces for future wireless networks: A channel modeling perspective," *IEEE Wireless Communications*, vol. 28, no. 3, pp. 108–114, 2021.

[17] S. Shen, B. Clerckx, and R. Murch, "Modeling and architecture design of reconfigurable intelligent surfaces using scattering parameter network analysis," *IEEE Transactions on Wireless Communications*, vol. 21, no. 2, pp. 1229–1243, 2022.

[18] B. Yang, X. Cao, J. Xu, C. Huang, G. C. Alexandropoulos, L. Dai, M. Debbah, H. V. Poor, and C. Yuen, "Reconfigurable intelligent computational surfaces: When wave propagation control meets computing," *IEEE Wireless Communications*, vol. 30, no. 3, pp. 120–128, 2023.

[19] A. M. Salhab and M. H. Samuh, "Accurate performance analysis of reconfigurable intelligent surfaces over rician fading channels," *IEEE Wireless Communications Letters*, vol. 10, no. 5, pp. 1051–1055, 2021.

[20] I. Yildirim, A. Uyrus, and E. Basar, "Modeling and analysis of reconfigurable intelligent surfaces for indoor and outdoor applications in future wireless networks," *IEEE Transactions on Communications*, vol. 69, no. 2, pp. 1290–1301, 2021.

[21] E. Björnson and L. Sanguinetti, "Rayleigh fading modeling and channel hardening for reconfigurable intelligent surfaces," *IEEE Wireless Communications Letters*, vol. 10, no. 4, pp. 830–834, 2021.

[22] 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. 116753–116773, 2019.

[23] T. N. Do, G. Kaddoum, T. L. Nguyen, D. B. Da Costa, and Z. J. Haas, "Multi-ris-aided wireless systems: Statistical characterization and performance analysis," *IEEE Transactions on Communications*, vol. 69, no. 12, pp. 8641–8658, 2021.

[24] Z. Chen, L. X. Cai, and X. Hao, "Near-field and far-field beamforming design for ris-enabled millimeter wave systems," in *2024 IEEE 99th Vehicular Technology Conference (VTC2024-Spring)*, pp. 1–6, 2024.

[25] D.-R. Emenonye, H. S. Dhillon, and R. M. Buehrer, "Fundamentals of ris-aided localization in the far-field," *IEEE Transactions on Wireless Communications*, vol. 23, no. 4, pp. 3408–3424, 2024.

[26] M. Delbari, G. C. Alexandropoulos, R. Schober, and V. Jamali, "Far-versus near-field ris modeling and beam design," *arXiv preprint arXiv:2401.08237*, 2024.

[27] S. Abeywickrama, R. Zhang, Q. Wu, and C. Yuen, "Intelligent reflecting surface: Practical phase shift model and beamforming optimization," *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5849–5863, 2020.

[28] K. Ardah, S. Gherekhloo, A. L. F. de Almeida, and M. Haardt, "Double-ris versus single-ris aided systems: Tensor-based mimo channel estimation and design perspectives," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5183–5187, 2022.

[29] L. Yang, Y. Yang, D. B. d. Costa, and I. Trigui, "Outage probability and capacity scaling law of multiple ris-aided networks," *IEEE Wireless Communications Letters*, vol. 10, no. 2, pp. 256–260, 2021.

[30] M. He, W. Xu, H. Shen, G. Xie, C. Zhao, and M. Di Renzo, "Cooperative multi-ris communications for wideband mmwave miso-ofdm systems," *IEEE Wireless Communications Letters*, vol. 10, no. 11, pp. 2360–2364, 2021.

[31] Y. Zhao, W. Xu, X. You, N. Wang, and H. Sun, "Cooperative reflection and synchronization design for distributed multiple-ris communications," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 5, pp. 980–994, 2022.

[32] M. Lan, Y. Hei, M. Huo, H. Li, and W. Li, "A new framework of ris-aided user-centric cell-free massive mimo system for iot networks," *IEEE Internet of Things Journal*, vol. 11, no. 1, pp. 1110–1121, 2024.

[33] H. Lu, D. Zhao, Y. Wang, C. Kong, and W. Chen, "Joint power control and passive beamforming in reconfigurable intelligent surface assisted user-centric networks," *IEEE Transactions on Communications*, vol. 70, no. 7, pp. 4852–4866, 2022.

[34] S. Yang, J. Zhang, W. Xia, Y. Ren, H. Yin, and H. Zhu, "A unified framework for distributed ris-aided downlink systems between mimo-noma and mimo-sdma," *IEEE Transactions on Communications*, vol. 70, no. 9, pp. 6310–6324, 2022.

[35] R. Zhong, X. Mu, Y. Liu, Y. Chen, J. Zhang, and P. Zhang, "Star-riss assisted noma networks: A distributed learning approach," *IEEE Journal of Selected Topics in Signal Processing*, vol. 17, no. 1, pp. 264–278, 2023.

[36] H. Du, J. Zhang, J. Cheng, and B. Ai, "Millimeter wave communications with reconfigurable intelligent surfaces: Performance analysis and optimization," *IEEE Transactions on Communications*, vol. 69, no. 4, pp. 2752–2768, 2021.

[37] F. Zhu, X. Wang, C. Huang, Z. Yang, X. Chen, A. Al Hammadi, Z. Zhang, C. Yuen, and M. Debbah, "Robust beamforming for ris-aided communications: Gradient-based manifold meta learning," *IEEE Transactions on Wireless Communications*, vol. 23, no. 11, pp. 15945–15956, 2024.

[38] H. Zhao, W. Sun, Y. Ni, W. Xia, G. Gui, and C. Zhu, "Deep deterministic policy gradient-based rate maximization for ris-uav-assisted vehicular communication networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 11, pp. 15732–15744, 2024.

[39] C. A. Pitz, M. V. Matsuo, and R. Seara, "A gradient-based algorithm for joint beamforming and reflection design in ris-assisted mobile communications," *Digital Signal Processing*, p. 105298, 2025.

[40] Q. Zhang, Y.-C. Liang, and H. V. Poor, "Reconfigurable intelligent surface assisted mimo symbiotic radio networks," *IEEE Transactions on Communications*, vol. 69, no. 7, pp. 4832–4846, 2021.

[41] X. He, L. Huang, and J. Wang, "Novel relax-and-retract algorithm for intelligent reflecting surface design," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 2, pp. 1995–2000, 2021.

[42] S. Mao, L. Liu, N. Zhang, M. Dong, J. Zhao, J. Wu, and V. C. M. Leung, "Reconfigurable intelligent surface-assisted secure mobile edge computing networks," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 6, pp. 6647–6660, 2022.

[43] Z. Yang, J. Shi, Z. Li, M. Chen, W. Xu, and M. Shikh-Bahaei, "Energy efficient rate splitting multiple access (rsma) with reconfigurable intelligent surface," in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–6, 2020.

[44] N. Panuganti, P. Ranjan, and A. Shukla, "Impact of metaheuristic optimization algorithms on wireless network coverage enhancement with reconfigurable intelligent surfaces," *International Journal of Communication Systems*, vol. 38, no. 5, p. e70026, 2025.

[45] S. Kumar, J. K. Rai, P. Ranjan, and R. Chowdhury, "Secured and energy efficient wireless system using reconfigurable intelligent surface and pso algorithm," in *2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, vol. 2, pp. 1–5, 2024.

[46] X. Xu, P. Xu, Y. Wang, Z. Wang, K. Yu, H. Shi, D. Ge, X. Ma, G. Leng, M. Wang, and C. Wang, "Intelligent design of reconfigurable microstrip antenna based on adaptive immune annealing algorithm," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–14, 2022.

[47] J. An, C. Xu, Q. Wu, D. W. K. Ng, M. Di Renzo, C. Yuen, and L. Hanzo, "Codebook-based solutions for reconfigurable intelligent surfaces and their open challenges," *IEEE Wireless Communications*, vol. 31, no. 2, pp. 134–141, 2024.

[48] Z. Zhang and W. Yu, "Learning beamforming codebooks for active sensing with reconfigurable intelligent surface," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2025.

[49] X. Wei, D. Shen, and L. Dai, "Channel estimation for ris assisted wireless communications—part i: Fundamentals, solutions, and future opportunities," *IEEE communications letters*, vol. 25, no. 5, pp. 1398–1402, 2021.

[50] D. An, J. Hu, K. Zhong, and Y. Cong, "Ris assisted multiple user interference mitigation via an accelerated coordinate descent method," *IEEE Transactions on Cognitive Communications and Networking*, vol. 9, no. 1, pp. 159–169, 2023.

[51] H. Yang, Z. Xiong, J. Zhao, D. Niyato, L. Xiao, and Q. Wu, "Deep reinforcement learning-based intelligent reflecting surface for secure wireless communications," *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 375–388, 2020.

[52] H. Zhou, M. Erol-Kantarci, Y. Liu, and H. V. Poor, "A survey on model-based, heuristic, and machine learning optimization approaches in ris-aided wireless networks," *IEEE Communications Surveys Tutorials*, vol. 26, no. 2, pp. 781–823, 2024.

[53] I. Sharma, R. Kumar, and S. J. Darak, "Online-learning-based multi-ris-aided wireless systems," *IEEE Systems Journal*, vol. 18, no. 2, pp. 1174–1185, 2024.

[54] M. Munochiveyi, A. C. Pogaku, D.-T. Do, A.-T. Le, M. Voznak, and N. D. Nguyen, "Reconfigurable intelligent surface aided multi-user communications: State-of-the-art techniques and open issues," *IEEE Access*, vol. 9, pp. 118584–118605, 2021.

[55] S. Yang, W. Lyu, D. Wang, and Z. Zhang, "Separate channel estimation with hybrid ris-aided multi-user communications," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 1, pp. 1318–1324, 2023.

[56] Z. Zhou, H. Yin, L. Tan, R. Zhang, K. Wang, and Y. Liu, "Multi-user passive beamforming in ris-aided communications and experimental validations," *IEEE Transactions on Communications*, vol. 72, no. 10, pp. 6569–6582, 2024.

[57] B. J. Qeryaqos and S. A. Ayoob, "Proposed multiple reconfigurable intelligent surfaces to mitigate the inter-user-interference problem in nlos," *Journal of Communications Software and Systems*, vol. 20, no. 3, pp. 245–252, 2024.

[58] P. Zheng, S. Tarboush, H. Sarieddeen, and T. Y. Al-Naffouri, "Mutual coupling-aware channel estimation and beamforming for ris-assisted communications," *arXiv preprint arXiv:2410.04110*, 2024.

[59] Z. Peng, G. Zhou, C. Pan, H. Ren, A. L. Swindlehurst, P. Popovski, and G. Wu, "Channel estimation for ris-aided multi-user mmwave systems with uniform planar arrays," *IEEE Transactions on Communications*, vol. 70, no. 12, pp. 8105–8122, 2022.

[60] Z. Xie, W. Yi, X. Wu, Y. Liu, and A. Nallanathan, "Downlink multi-ris aided transmission in backhaul limited networks," *IEEE Wireless Communications Letters*, vol. 11, no. 7, pp. 1458–1462, 2022.

[61] K. Xu, S. Gong, M. Cui, G. Zhang, and S. Ma, "Statistically robust transceiver design for multi-ris assisted multi-user mimo systems," *IEEE Communications Letters*, vol. 26, no. 6, pp. 1428–1432, 2022.

[62] Y. Yu, X. Liu, and V. C. M. Leung, "Fair downlink communications for ris-uav enabled mobile vehicles," *IEEE Wireless Communications Letters*, vol. 11, no. 5, pp. 1042–1046, 2022.

[63] A. M. Elbir, A. Papazafeiropoulos, P. Kourtessis, and S. Chatzinotas, "Deep channel learning for large intelligent surfaces aided mm-wave massive mimo systems," *IEEE Wireless Communications Letters*, vol. 9, no. 9, pp. 1447–1451, 2020.

[64] W. Ni, Y. Liu, Z. Yang, H. Tian, and X. Shen, "Federated learning in multi-ris-aided systems," *IEEE Internet of Things Journal*, vol. 9, no. 12, pp. 9608–9624, 2022.

[65] M. Ahmed, A. Wahid, W. U. Khan, F. Khan, A. Ihsan, Z. Ali, K. M. Rabie, T. Shongwe, and Z. Han, "A survey on ris advances in terahertz communications: Emerging paradigms and research frontiers," *IEEE Access*, vol. 12, pp. 173867–173901, 2024.

[66] H. Du, J. Zhang, K. Guan, D. Niyato, H. Jiao, Z. Wang, and T. Kürner, "Performance and optimization of reconfigurable intelligent surface aided thz communications," *IEEE Transactions on Communications*, vol. 70, no. 5, pp. 3575–3593, 2022.

[67] Y. Yin, L. Yang, X. Li, H. Liu, K. Guo, and Y. Li, "On the performance of active ris-assisted mixed rf-thz relaying systems," *IEEE Internet of Things Journal*, vol. 12, no. 10, pp. 14938–14951, 2025.

[68] H. Xia, Q. Xue, Y. Liu, B. Zhou, M. Hua, and Q. Chen, "Intelligent angle map-based beam alignment for ris-aided mmwave communication networks," *arXiv preprint arXiv:2410.23919*, 2024.

[69] Q. Xue, C. Ji, S. Ma, J. Guo, Y. Xu, Q. Chen, and W. Zhang, "A survey of beam management for mmwave and thz communications towards 6g," *IEEE Communications Surveys & Tutorials*, vol. 26, no. 3, pp. 1520–1559, 2024.

[70] E. Shi, J. Zhang, H. Du, B. Ai, C. Yuen, D. Niyato, K. B. Letaief, and X. Shen, "Ris-aided cell-free massive mimo systems for 6g: Fundamentals, system design, and applications," *Proceedings of the IEEE*, vol. 112, no. 4, pp. 331–364, 2024.

[71] M. Diamanti, P. Charatsaris, E. E. Tsiropoulou, and S. Papavassiliou, "The prospect of reconfigurable intelligent surfaces in integrated access and backhaul networks," *IEEE Transactions on Green Communications and Networking*, vol. 6, no. 2, pp. 859–872, 2022.

[72] H. Du, J. Zhang, K. Guan, D. Niyato, H. Jiao, Z. Wang, and T. Kürner, "Performance and optimization of reconfigurable intelligent surface aided thz communications," *IEEE Transactions on Communications*, vol. 70, no. 5, pp. 3575–3593, 2022.

[73] T. L. Nguyen, T. N. Do, G. Kaddoum, D. B. d. Costa, and Z. J. Haas, "Channel characterization for ris-aided terahertz communications: A stochastic approach," *IEEE Wireless Communications Letters*, vol. 11, no. 9, pp. 1890–1894, 2022.

[74] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine learning*, vol. 47, pp. 235–256, 2002.

[75] O. Chapelle and L. Li, "An empirical evaluation of thompson sampling," in *Advances in neural information processing systems*, pp. 2249–2257, 2011.

[76] S. Agrawal and N. Goyal, "Analysis of thompson sampling for the multi-armed bandit problem," in *Conference on Learning Theory*, pp. 39–1, 2012.

[77] Q.-Y. Yu, H.-C. Lin, and H.-H. Chen, "Intelligent radio for next generation wireless communications: An overview," *IEEE Wireless Communications*, vol. 26, no. 4, pp. 94–101, 2019.

[78] M. Bkassiny, Y. Li, and S. K. Jayaweera, "A survey on machine-learning techniques in cognitive radios," *IEEE Communications Surveys Tutorials*, vol. 15, no. 3, pp. 1136–1159, 2013.

[79] 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, 2022.

[80] H. Du, J. Zhang, J. Cheng, and B. Ai, "Millimeter wave communications with reconfigurable intelligent surfaces: Performance analysis and optimization," *IEEE Transactions on Communications*, vol. 69, no. 4, pp. 2752–2768, 2021.

[81] H. Zhang, B. Di, L. Song, and Z. Han, "Reconfigurable intelligent surfaces assisted communications with limited phase shifts: How many phase shifts are enough?," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 4, pp. 4498–4502, 2020.

[82] E. Basar, "Reconfigurable intelligent surface-based index modulation: A new beyond mimo paradigm for 6g," *IEEE Transactions on Communications*, vol. 68, no. 5, pp. 3187–3196, 2020.

[83] S. Arzykulov, G. Nauryzbayev, A. Celik, and A. M. Eltawil, "Ris-assisted full-duplex relay systems," *IEEE Systems Journal*, vol. 16, no. 4, pp. 5729–5740, 2022.

[84] P. Zhang, J. Zhang, H. Xiao, H. Du, D. Niyato, and B. Ai, "Ris-aided 6g communication system with accurate traceable user mobility," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 2, pp. 2718–2722, 2023.

[85] T. Jiang, H. V. Cheng, and W. Yu, "Learning to reflect and to beamform for intelligent reflecting surface with implicit channel estimation," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 7, pp. 1931–1945, 2021.

[86] D. Shen and L. Dai, "Dimension reduced channel feedback for reconfigurable intelligent surface aided wireless communications," *IEEE Transactions on Communications*, vol. 69, no. 11, pp. 7748–7760, 2021.

[87] X. Ma, Y. Fang, H. Zhang, S. Guo, and D. Yuan, "Cooperative beamforming design for multiple ris-assisted communication systems," *IEEE Transactions on Wireless Communications*, vol. 21, no. 12, pp. 10949–10963, 2022.

[88] K. Keykhosravi and H. Wymeersch, "Multi-ris discrete-phase encoding for interpath-interference-free channel estimation," *arXiv preprint arXiv:2106.07065*, 2021.

[89] G. Zhou, C. Pan, H. Ren, K. Wang, and M. D. Renzo, "Fairness-oriented multiple ris-aided mmwave transmission: Stochastic optimization methods," *IEEE Transactions on Signal Processing*, vol. 70, pp. 1402–1417, 2022.

[90] Y. Wang, W. Zhang, Y. Chen, C.-X. Wang, and J. Sun, "Novel multiple ris-assisted communications for 6g networks," *IEEE Communications Letters*, vol. 26, no. 6, pp. 1413–1417, 2022.

[91] Y. Huo, X. Dong, and N. Ferdinand, "Distributed reconfigurable intelligent surfaces for energy-efficient indoor terahertz wireless communications," *IEEE Internet of Things Journal*, vol. 10, no. 3, pp. 2728–2742, 2023.

[92] V. K. Chapala and S. M. Zafaruddin, "Multiple ris-assisted mixed fso-rf transmission over generalized fading channels," *IEEE Systems Journal*, vol. 17, no. 3, pp. 3515–3526, 2023.

[93] J. Tong, H. Zhang, L. Fu, A. Leshem, and Z. Han, "Two-stage resource allocation in reconfigurable intelligent surface assisted hybrid networks via multi-player bandits," *IEEE Transactions on Communications*, vol. 70, no. 5, pp. 3526–3541, 2022.

[94] E. M. Mohamed, S. Hashima, K. Hatano, and M. M. Fouda, "Cost-effective mab approaches for reconfigurable intelligent surface aided millimeter wave relaying," *IEEE Access*, vol. 10, pp. 81642–81653, 2022.

[95] S. V. S. Santosh and S. J. Darak, "Multiarmed bandit algorithms on zynq system-on-chip: Go frequentist or bayesian?," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–14, 2022.

[96] M. K. Hanawal and S. J. Darak, "Multiplayer bandits: A trekking approach," *IEEE Transactions on Automatic Control*, vol. 67, no. 5, pp. 2237–2252, 2022.

[97] H. Tibrewal, S. Patchala, M. K. Hanawal, and S. J. Darak, "Distributed learning and optimal assignment in multiplayer heterogeneous networks," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, pp. 1693–1701, 2019.

[98] D. Ghosh, M. K. Hanawal, and N. Zlatanov, "Learning to optimize energy efficiency in energy harvesting wireless sensor networks," *IEEE Wireless Communications Letters*, vol. 10, no. 6, pp. 1153–1157, 2021.

[99] Y. Zhang and R. W. Heath, "Multi-armed bandit for link configuration in millimeter-wave networks: An approach for solving sequential decision-making problems," *IEEE Vehicular Technology Magazine*, pp. 2–13, 2023.

[100] Z. Kuai and S. Wang, "Thompson sampling-based antenna selection with partial csi for tdd massive mimo systems," *IEEE Transactions on Communications*, vol. 68, no. 12, pp. 7533–7546, 2020.

[101] Y. Song, C. Liu, W. Zhang, Y. Liu, H. Zhou, and X. Shen, "Two stage beamforming in massive mimo: A combinatorial multi-armed bandit based approach," *IEEE Transactions on Vehicular Technology*, pp. 1–6, 2023.

[102] M.-J. Youssef, V. V. Veeravalli, J. Farah, C. A. Nour, and C. Douillard, "Resource allocation in noma-based self-organizing networks using stochastic multi-armed bandits," *IEEE Transactions on Communications*, 2021.

[103] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine learning*, pp. 235–256, 2002.

[104] A. Garivier and O. Cappé, "The kl-ucb algorithm for bounded stochastic bandits and beyond," in *Proceedings of the 24th annual Conference On Learning Theory*, pp. 359–376, 2011.

[105] S. Agrawal and N. Goyal, "Further optimal regret bounds for thompson sampling," in *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics* (C. M. Carvalho and P. Ravikumar, eds.), vol. 31 of *Proceedings of Machine Learning Research*, (Scottsdale, Arizona, USA), pp. 99–107, PMLR, 29 Apr–01 May 2013.

[106] E. M. Mohamed, S. Hashima, and K. Hatano, "Energy aware multiarmed bandit for millimeter wave-based uav mounted ris networks," *IEEE Wireless Communications Letters*, vol. 11, no. 6, pp. 1293–1297, 2022.

[107] S. Bubeck, N. Cesa-Bianchi, *et al.*, "Regret analysis of stochastic and nonstochastic multi-armed bandit problems," *Foundations and Trends® in Machine Learning*, vol. 5, no. 1, pp. 1–122, 2012.

[108] J. Kwon, V. Perchet, and C. Vernade, "Sparse stochastic bandits," in *Proceedings of the 2017 Conference on Learning Theory* (S. Kale and O. Shamir, eds.), vol. 65 of *Proceedings of Machine Learning Research*, pp. 1269–1270, PMLR, 07–10 Jul 2017.

[109] E. Björnson, Ö. Özdogan, and E. G. Larsson, "Intelligent reflecting surface versus decode-and-forward: How large surfaces are needed to beat relaying?," *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 244–248, 2019.

[110] 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, 2019.

[111] A. Slivkins *et al.*, "Introduction to multi-armed bandits," *Foundations and Trends® in Machine Learning*, vol. 12, no. 1-2, pp. 1–286, 2019.

[112] N. Singh and S. J. Darak, "Enhancing wireless phy with adaptive ofdm and multi-armed bandit learning on zynq system on chip," *IEEE Transactions on Very Large Scale Integration Systems (TVLSI)*, pp. 1–13, 2024.

[113] C. Hou and Q. Zhao, "Stopping-time management of smart sensing nodes based on trade-offs between accuracy and power consumption," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 25, no. 9, pp. 2472–2485, 2017.

[114] R. Chen, S. Lu, M. A. Elgammal, P. Chun, V. Betz, and D. Niu, "Vpr-gym: A platform for exploring ai techniques in fpga placement optimization," in *2023 33rd International Conference on Field-Programmable Logic and Applications (FPL)*, pp. 72–78, 2023.

[115] D. Ghosh, M. K. Hanawal, and N. Zlatanov, "Holobeam: Learning optimal beamforming in far-field holographic metasurface transceivers," in *IEEE INFOCOM 2024 - IEEE Conference on Computer Communications*, pp. 301–310, 2024.

[116] F. Li, D. Yu, H. Yang, J. Yu, H. Karl, and X. Cheng, "Multi-armed-bandit-based spectrum scheduling algorithms in wireless networks: A survey," *IEEE Wireless Communications*, vol. 27, no. 1, pp. 24–30, 2020.

[117] D. Bouneffouf, I. Rish, and C. Aggarwal, "Survey on applications of multi-armed and contextual bandits," in *2020 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1–8, 2020.

[118] N. Singh, S. V. S. Santosh, and S. J. Darak, "Toward intelligent reconfigurable wireless physical layer (phy)," *IEEE Open Journal of Circuits and Systems*, vol. 2, pp. 226–240, 2021.

[119] S. V. S. Santosh and S. J. Darak, "Multiarmed bandit algorithms on zynq system-on-chip: Go frequentist or bayesian?," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–14, 2022.

[120] S. V. S. Santosh and S. J. Darak, "Intelligent and reconfigurable architecture for kl divergence-based multi-armed bandit algorithms," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 68, no. 3, pp. 1008–1012, 2021.

[121] A. Sneh, S. S. Ram, S. J. Darak, and A. Tewari, "Beam alignment in multipath environments for integrated sensing and communication using bandit learning," *IEEE Journal of Selected Topics in Signal Processing*, 2024.

[122] S. V. S. Santosh and S. J. Darak, "Reconfigurable and computationally efficient architecture for multi-armed bandit algorithms," in *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–5, 2020.

[123] Y. Wei, Z. Zhong, V. Y. F. Tan, and C. Wang, "Fast beam alignment via pure exploration in multi-armed bandits," in *2022 IEEE International Symposium on Information Theory (ISIT)*, pp. 1886–1891, 2022.

[124] T. Bonald and R. Combes, "A minimax optimal algorithm for crowdsourcing," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[125] A. Verma, M. Hanawal, A. Rajkumar, and R. Sankaran, "Censored semi-bandits: A framework for resource allocation with censored feedback," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[126] A. Verma, M. K. Hanawal, and N. Hemachandra, "Unsupervised online feature selection for cost-sensitive medical diagnosis," in *2020 International Conference on COMmunication Systems NETworkS (COMSNETS)*, pp. 1–6, 2020.

[127] M. Hanawal, C. Szepesvari, and V. Saligrama, "Unsupervised Sequential Sensor Acquisition," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics* (A. Singh and J. Zhu, eds.), vol. 54 of *Proceedings of Machine Learning Research*, pp. 803–811, PMLR, 20–22 Apr 2017.

[128] A. Verma, M. K. Hanawal, C. Szepesvári, and V. Saligrama, "Online algorithm for unsupervised sensor selection," *CoRR*, vol. abs/1901.04676, 2019.

[129] I. Sharma, R. Kumar, and S. J. Darak, "Optimizing ris block selection for power consumption," in *2025 17th International Conference on COMmunication Systems and NETworks (COMSNETS)*, pp. 989–993, 2025.

[130] A. Garivier and E. Moulines, "On upper-confidence bound policies for switching bandit problems," in *Algorithmic Learning Theory* (J. Kivinen, C. Szepesvári, E. Ukkonen,

and T. Zeugmann, eds.), (Berlin, Heidelberg), pp. 174–188, Springer Berlin Heidelberg, 2011.

[131] L. Kocsis and C. Szepesvári, "Discounted ucb," in *2nd PASCAL Challenges Workshop*, vol. 2, pp. 51–134, 2006.

[132] J. Kwon, V. Perchet, and C. Vernade, "Sparse stochastic bandits," *CoRR*, vol. abs/1706.01383, 2017.

[133] A. Locatelli, M. Gutzeit, and A. Carpentier, "An optimal algorithm for the thresholding bandit problem," in *International Conference on Machine Learning*, pp. 1690–1698, PMLR, 2016.

[134] Y. Cao, Z. Wen, B. Kveton, and Y. Xie, "Nearly optimal adaptive procedure with change detection for piecewise-stationary bandit," in *International Conference on Artificial Intelligence and Statistics*, 2018.

This page was intentionally left blank.

List of Publications

Journals

1. **I. Sharma**, R. Kumar, and S. J. Darak, "Online-Learning-Based Multi-RIS-Aided Wireless Systems," *IEEE Systems Journal*, vol. 18, no. 2, pp. 1174–1185, June 2024, doi: 10.1109/JSYST.2024.3391856.

2. **I. Sharma**, S. J. Darak, and R. Kumar, "High-Speed Compute-Efficient Bandit Learning for Many Arms," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 2025, accepted for publication, doi: 10.1109/TVLSI.2025.3573924.

Conferences

1. **I. Sharma**, R. Kumar, and S. J. Darak, "Efficient Hardware Implementation of Multi-Armed Bandit Algorithms for RIS-Aided Wireless Networks," in *Proceedings of the IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, Gwalior, India, pp. 1–4, 2025, doi: 10.1109/IATMSI64286.2025.10985441.

2. **I. Sharma**, R. Kumar, and S. J. Darak, "Optimizing RIS Block Selection for Power Consumption," in *Proceedings of the 17th International Conference on Communication Systems & Networks (COMSNETS 2025: Poster)*, India, Jan. 2025, doi: 10.1109/COMSNETS63942.2025.10885684.

Communicated Journals

1. **Ishaan Sharma**, Rohit Kumar, and Sumit J. Darak, "Optimizing RIS Selection: Balancing Power Consumption and SNR", ***IEEE Transactions on Mobile Computing***

2. **Ishaan Sharma**, Rohit Kumar, and Sumit J. Darak, "Optimal RIS Selection with Varying Performance - A Dynamic Multi-Armed Bandit Framework", ***IEEE Transactions on Vehicular Technology***

DELHI TECHNOLOGICAL UNIVERSITY
formerly Delhi College of Engineering
Shahbad Daulatpur, Main Bawana Road,
Delhi–110042

Plagiarism Verification

Thesis Title: Distributed Learning for Reconfigurable Intelligent Surfaces

Name of the Scholar: Ishaan Sharma

Supervisor: Dr. Rohit Kumar

Designation: Assistant Professor

Department: Electronics and Communication Engineering

This is to report that the above thesis was scanned for similarity detection. Processand outcomes are given below:

Software Used: Turnitin

Submission ID: 27535:106590929

Similarity Index:

Self-Publication(s) Similarity Index:

Final Total Similarity Index: 4%

Total Word Count: 38,292

Date: 31/07/2025

Signature of Supervisor

Signature of Candidate

Ishaan Sharma
New Delhi, India
: +91 9911431188
: ishaanaero@gmail.com, ishaan5@mail.ru

Educational Background

Doctor of Philosophy (Ph. D.)(CGPA = 8.78) 2020 - 2025
(Thesis submitted)

Department of Electronics and Communication Engg.
Delhi Technological University (DTU) Delhi, India.
(in association with Indraprastha Institute of Information Technology, Delhi)

- Thesis Title: "Distributed Learning for Reconfigurable Intelligent Surfaces"
- Ph. D. Supervisors:
  - Dr. Rohit Kumar-Asst. Professor, DTU Delhi
  - Dr. Sumit J. Darak- Professor, IIIT Delhi

Master of Technology (M. Tech.) (CGPA = 7.6) 2018 -2020

Department of Communication Engg.
Vellore Institute of Technology
VIT University, Vellore, Tamil Nadu, India.

- Thesis Title: "Design of multiband multi-port antenna for sub-6GHz 5G Applications"
- Thesis Supervisor: Dr. Rajesh A.,Associate Professor, VIT Vellore
  (Now in Sastra University)

Bachelor of Technology (B. Tech.) (CGPA = 7.11) 2014 - 2018

Department of Electronics and Communication Engg.
SRM Institute of Science and Technology, NCR Campus, Ghaziabad, UP, India
SRM University, Chennai, Tamil Nadu, India.

Technical Skills

- **Languages:** C, MATLAB, basics of Python
- **Hardware Tools:** Xilinx Vivado, HSCD, IP Design, Zedboard, ZCU706

Projects

- **RIS for Smart Villages (DST-SERB Project)**
  - Developed RIS-aided wireless system to improve energy-efficient connectivity in rural areas.
  - Applied reinforcement learning and multi-armed bandit techniques for adaptive RIS configuration.
  - Deployed hardware-software co-design using Zynq SoC for real-time inference.
- **Custom Hardware IP Design and Integration on Zynq SoC**
  - Designed custom IP in Vivado HLS and integrated with ARM Cortex-A9 via AXI interfaces.
  - Implemented pipelining, loop unrolling, and memory optimizations.
  - Used DMA for efficient data transfer between processing system (PS) and logic (PL).
  - Evaluated trade-offs in word-length reduction and accuracy.
- **Multiport Multiband Antenna Design for Sub-6 GHz 5G (M.Tech Thesis)**
  - Designed and simulated dual-port multiband antenna using CST.

  - Optimized for sub-6 GHz bands for LTE/5G with low VSWR and high gain.
  - Validated design via fabricated prototype.
- **ISAC-RIS-Aided Bandit Learning Simulation Framework**
  - Currently collaborating with Dr. Galymzhan Nauryzbayev and Dr. Sultangali Arzykulov, Nazarbayev University, Astana, Kazakhstan.
  - Developed a simulation framework for RIS-assisted ISAC systems using contextual bandit learning.
  - Evaluated performance using Rician fading, log-scale SNR normalization, and CVX-based power allocation.

PUBLICATIONS:
IEEE Journals(Accepted)

2. **Ishaan Sharma**, Sumit J. Darak, and Rohit Kumar, "High-Speed Compute-Efficient Bandit Learning for Many Arms", ***IEEE Transactions on Very Large Scale Integration Systems***,vol. 33, no. 7, pp. 2099-2103 July 2025. doi:10.1109/TVLSI.2025.3573924

1. **Ishaan Sharma**, Rohit Kumar, and Sumit J. Darak, "Online-Learning-Based Multi-RIS-Aided Wireless Systems", ***IEEE Systems Journal***, vol. 18, no. 2, pp. 1174–1185, June 2024. doi:10.1109/JSYST.2024.3391856

PUBLICATIONS:
IEEE Journals(Under Revision)

2. **Ishaan Sharma**, Rohit Kumar, and Sumit J. Darak, "Optimizing RIS Selection: Balancing Power Consumption and SNR", ***IEEE Transactions on Mobile Computing***

1. **Ishaan Sharma**, Rohit Kumar, and Sumit J. Darak, "Optimal RIS Selection with Varying Performance - A Dynamic Multi-Armed Bandit Framework", ***IEEE Transactions on Vehicular Technology***

PUBLICATIONS:
Conferences

4. **Ishaan Sharma**, Rohit Kumar, and Sumit J. Darak, "Optimizing RIS Block Selection for Power Consumption", in *Proc. 17th International Conference on Communication Systems and Networks (COMSNETS)*, Bengaluru, India, 2025, pp. 989–993. doi:10.1109/COMSNETS63942.2025.10885684

3. **Ishaan Sharma**, Rohit Kumar, and Sumit J. Darak, "Efficient Hardware Implementation of Multi-Armed Bandit Algorithms for RIS-Aided Wireless Networks", in *Proc. IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, Gwalior, India, 2025, pp. 1–4. doi:10.1109/IATMSI64286.2025.10985441

2. Khagendra Joshi, Sana Ali Naqvi, Vivek A. Bohra, and **Ishaan Sharma**, "Performance Characterization of an IRS-Assisted OFDM System with HPA Memory Effects and IRS Phase Noise", in *Proc. IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*, Guwahati, India, 2024, pp. 1–6. doi:10.1109/ANTS63515.2024.10898233

1. **Ishaan Sharma**, T. Shankar, and A. Rajesh, "Design of Dual-Port Multiband Antenna for Sub-6 GHz 5G Applications", in *Advances in Automation, Signal Processing, Instrumentation, and Control (i-CASIC 2020)*, vol. 700, Springer, Singapore, 2021.

Honours and Awards

- Represented Delhi Technological University and IIIT Delhi in the VLSID Design Contest at **VLSID Conference 2023**.
- Travel Grant to attend **COMSNETS 2025** at Bangalore, India, Jan 2019.
- Selected for Graduate Forum at **COMSNETS 2025**, Bengaluru, India.
- Nominated for Research and Innovation Excellence Award 2025 at DTU.

Teaching Experience

- Teaching Assistant, DTU/IIIT-D: Digital Design, Comm. Systems, RL, FPGA Summer School
- Guided B.Tech and M.Tech students in hands-on hardware–software co-design projects and thesis work.