Bulletin of Electrical Engineering and Informatics
Vol. 15, No. 3, June 2026, pp. 2162~2177
ISSN: 2302-9285, DOI: 10.11591/eei.v15i3.11909

Risk-aware dynamic spectrum allocation using learning-augmented RIS control for cognitive radio networks

Mallikarjun Dheshmuk[1,2], Santosh B. Kumbalavati[2]
[1]Department of Electronics and Communication Engineering, BLDEA's V. P. Dr. P. G. Halakatti College of Engineering and Technology (Affiliated to Visvesvaraya Technological University), Vijayapura, India
[2]Department of Electronics and Communication Engineering, Basaveshwar Engineering College (Affiliated to Visvesvaraya Technological University), Bagalkote, India

Article Info

Article history:

Received Nov 19, 2025
Revised Mar 20, 2026
Accepted Apr 19, 2026

Keywords:

Bilevel optimization
Geometric programming
Primary user
Secondary user
Semidefinite relaxation

ABSTRACT

The proposed RADIANT-CRN framework introduces a risk-aware dynamic spectrum allocation approach for reconfigurable intelligent surface (RIS)-enabled cognitive radio networks while maintaining reliable protection for primary users (PUs). It incorporates our previously developed bidirectional long short-term memory and adaptive manta-ray foraging optimization (BiLSTM-AMRFO) spectrum prediction and deep channel estimation models into a bilevel optimization framework, where semidefinite relaxation (SDR) adjusts RIS phase shifts, geometric programming (GP) allocates transmit power, and entropy-regularized assignment performs channel selection. A primal-dual actor–critic framework coordinates these modules. On a 120 MHz testbed with 256-element RIS, RADIANT-CRN gets a sum rate of 846±18 Mbps with a 0.7±0.2% PU-violation and 99.3% chance-constraint coverage. This is about 25% higher than a greedy non-risk baseline and about 43% higher than a no-RIS optimizer. It also lowers interference CVaR_0.95 from 9.1 mW to 2.4 mW. These results demonstrate that RADIANT-CRN is the first framework to enforce both chance and CVaR guarantees in RIS-assisted CRNs, achieving high spectral efficiency with statistically certifiable PU protection that aligns with Federal Communications Commission interference requirements. The framework is validated on a prototype SDR testbed (Rician fading, K=6 dB, bursty PU activity); implementation uses Python 3.11/PyTorch 2.x with convex optimization python (CVXPy/MOSEK), and synthetic testbed traces are available upon request.

Corresponding Author:

Mallikarjun Dheshmuk
Department of Electronics and Communication Engineering
BLDEAs V. P. Dr. P. G. Halakatti College of Engineering and Technology
(Affiliated to Visvesvaraya Technological University)
Ashram Road, Vijayapura-586103, India
Email: mallikarjun.deshmukh1@mail.com

1. INTRODUCTION

When primary users (PUs) are not actively using licensed bands, dynamic spectrum access (DSA) can opportunistically reuse them, and promising significant spectral advantages [1]. Two obstacles still stand in the way of practical deployments, even though sensing and learning have come a long way in the past decade: i) the precariousness of decisions made in the face of uncertainty (PU activity, channels, and mobility) [2] and ii) the lack of environmental degrees of freedom to safely and efficiently route energy [3]-[6]. Robust, calibrated decision-making tackles [7], reconfigurable intelligent surfaces (RIS) tackle by

manipulating propagation with near-passive phase meta-elements [8], [9]. Online, at control-loop latencies consistent with actual radios, the primary concern is how to simultaneously distribute channels, powers, and RIS phases so that secondary users (SUs) achieve high throughput [10] without breaking PU interference budgets [11], [12] — a requirement directly mandated by Federal Communications Commission (FCC) Part 15 secondary-user rules and International Telecommunication Union (ITU-R) SM.1048 interference protection criteria, which are increasingly relevant to practical 5G/6G deployments in urban dense networks and internet of things (IoT) environments.

Our earlier works [13] laid the groundwork for two main areas. First one, smart spectrum management in cognitive radio networks (CRNs) with RIS assist: a method using adaptive manta-ray foraging optimization (AMRFO) and bidirectional long short-term memory (BiLSTM) by merging time-related features with attention-guided metaheuristic tuning, fine-tuning produced reliable idle-beliefs and channel surrogates. Secondly, next-generation CRNs with an optimized deep learning framework for spectrum sensing and channel estimation have better estimators that are stable regardless of the site or signal-to-noise ratio (SNR) [14]. Within a risk-aware optimization and control stack, to promote these predictors from "offline analytics" to "online priors" in this paper. As a learning-augmented bilevel framework with four interdependent notions, to present RADIANT-CRN: i) reduce needless exposure, the system rewards opportunities where PUs are expected to be absent, according to belief-weighted utility, which is achieved by weighing spectral utility with idle probabilities, ii) forego heuristic margins in favor of chance limitations and conditional value-at-risk (CVaR) on interference, which regulates average behavior as well as regulatorily significant tail occurrences, iii) three optimization methods are used: semidefinite relaxation (SDR) with randomization for RIS phases, geometric programming (GP) for powers, and an entropy-regularized relaxation followed by rounding for discrete channel picks. The computational tractability and feasibility are both maintained by this decomposition, and iv) continuous learning: the blocks are coordinated by a primal-dual actor-critic, with the outputs of BiLSTM and AMRFO serving as warm-starts. The actor suggests (x, p, and Θ), the inner SDR/GP improves them, the dual updates the price of PU-risk, and the critic makes decisions on long-term rewards, which encompass switching costs and fairness. For these reasons, this mixture being used-while pure optimization is respectful of restrictions, it may be sluggish or myopic in non-stationary settings; pure learning, on the other hand, adapts rapidly but has trouble with harsh constraints and worst-case guarantees. RADIANT-CRN integrates both: learning offers potential courses of action and adjusts for uncertainty; optimization ensures practicability and risk management; dual variables convey when the system is secure; and entropy and codebook projection manage the hardware's actual capabilities. The aim is to tackle the following key challenges: i) partial observability, which can be reduced through belief calibration (expected calibration error (ECE)); ii) non-stationarity, which can be overcome through actor-critic adaptation and warm-start mapping; iii) mixed discrete-continuous decisions and unit-modulus RIS constraints, which can be resolved through relaxation and projection; and iv) solver latency, which can be measured and broken down to fulfill real-time budgets.

Our contributions are as follows: i) a risk-aware objective that couple’s belief-weighted throughput with explicit CVaR and chance constraints, turning regulatory limits into tractable control laws, ii) a bilevel decomposition (SDR for RIS, GP for powers, and entropic assignment for channels) proven effective by ablations: removing SDR yields the largest throughput drop (−178 Mbps), removing GP raises tail risk (+1.6 mW CVaR) and violations (+0.7 pp), iii) a learning-augmented primal–dual controller that achieves 846±18 Mbps with 0.7% violations and 99.3% coverage—≈25% over greedy and ≈43% over no-RIS—while improving fairness (Jain’s 0.94) and cutting user-level outage to 1.3%, and iv) a runtime analysis showing 25.9 ms per slot at N=256(≈39 Hz loop), with RIS SDR as the dominant block; at N=128 the loop reaches 57 Hz, enabling tighter control without sacrificing safety. To the best of our knowledge, RADIANT-CRN is the first RIS-assisted CRN framework to jointly enforce chance and CVaR interference constraints within a learning-augmented bilevel optimization architecture. The Pareto analysis confirms that baselines require disproportionately higher CVaR to push throughput from 800→850 Mbps, whereas RADIANT-CRN achieves the same region with significantly reduced tail risk through RIS-guided PU-benign beamforming and power redistribution. Reliability graphs confirm low ECE, meaning predicted risks match observed outcomes—a prerequisite for regulatory acceptance of automated DSA systems. In summary, RADIANT-CRN recasts RIS-assisted DSA as an optimized, learning-accelerated, and real-time-feasible control problem with certifiable safety guarantees.

The increasing demand for spectrum efficiency and the rapid growth of wireless devices have intensified the need for intelligent spectrum management in CRNs. While recent studies have explored machine learning (ML)-based spectrum prediction and optimization-driven resource allocation, several challenges remain unresolved. In particular, most existing approaches either rely solely on learning-based decision mechanisms, which often lack strict interference guarantees for PUs, or depend entirely on optimization frameworks that may struggle to adapt to highly dynamic wireless environments. Furthermore, the integration of RIS introduces additional decision variables such as phase configuration and beam steering,

which significantly increase the complexity of dynamic spectrum allocation. These limitations highlight the need for a unified framework capable of combining learning-based adaptability with optimization-based reliability while ensuring risk-aware PU protection.

To address these challenges, this paper proposes RADIANT-CRN, a learning-augmented optimization framework for risk-aware dynamic spectrum allocation in RIS-assisted CRNs. The proposed framework integrates deep learning-based spectrum intelligence with bilevel optimization techniques to jointly control channel assignment, power allocation, and RIS phase configuration while enforcing probabilistic interference guarantees for PUs. The main contributions of this work are summarized as follows:

a. A risk-aware spectrum allocation formulation that integrates belief-weighted throughput optimization with chance constraints and CVaR to ensure reliable protection for PUs.
b. A bilevel optimization framework that decomposes the joint allocation problem into RIS phase optimization using SDR, power allocation via GP, and entropy-regularized channel assignment.
c. A learning-augmented primal–dual control mechanism that combines reinforcement learning with optimization solvers to enable adaptive spectrum allocation under dynamic network conditions.
d. A comprehensive experimental evaluation demonstrating significant improvements in spectrum efficiency, fairness, and interference risk reduction compared with conventional greedy and non-RIS allocation strategies.

The remainder of this paper is organized as follows. Section 2 reviews related work in RIS-assisted wireless communication and learning-based spectrum allocation. Section 3 presents the proposed RADIANT-CRN framework and its optimization components. Section 4 discusses experimental results and performance analysis, and section 5 concludes the paper with future research directions.

2. RELATED WORKS

Downlink transmissions from a low earth orbit (LEO) satellite, with the help of a beyond diagonal reconfigurable intelligent surface (BD-RIS) to reach ground terminals, are the main subject of Zarini *et al*. [15]. Related RIS-optimization frameworks also appear in [16], [17]. Our goal is to minimize the average age of information (AoI) attained at ground terminals in order to optimize the system's performance. With the optimization of the downlink transmit power at the LEO satellite and reflection coefficients at the BD-RIS as decision variables, our formulation ensures the quality-of-service of ground terminals while respecting the LEO satellite's power budget. To reframe the issue as a Markov decision process that adequately describes its dynamics because of its non-convex and tightly-coupled character. The decision variables are then optimized by training a Q-learning propagation (Q-Prop) agent. This communication system is extremely dynamic due to the mobility of both terrestrial terminals and LEO satellites. So, to make the trained Q-Prop model more flexible and able to generalize to different systems, to use a meta-learning approach. Our optimized approach produces an average AoI that is 38% lower and 26% lower than that of our RIS-lacking and RIS-assisted equivalents, respectively, according to the numerical data.

The application of artificial intelligence (AI) and ML to wireless networks and communications was revolutionary, according to Sherin *et al*. [18]. Further ML-driven spectrum access works are presented in [19], [20], while deep learning surveys for wireless systems appear in [21]. As wireless systems get more intricate, traditional methods can't handle problems like latency, interference, and spectrum scarcity. Decisions about traffic forecasting, network automation, and DSA can be made smarter with the help of AI and ML methods such as deep learning and reinforcement learning. All of these things improve the reliability, adaptability, and performance of wireless networks, which is particularly important for the next generation of networks, 5G and 6G. Here to list some potential future research topics, including cognitive network management, AI-first 6G networks, federated learning, and to highlight some of the most important AI/ML use cases in wireless networks. To also discuss some of the issues associated with computing cost and data privacy. An improved Pelican optimization (POA) with quantum process deep reinforcement learning algorithm (POQPDRL) control framework is suggested by Parimala [22] for the combined management of a grid-connected doubly fed induction generator-based (DFIG-based) wind energy system that incorporates a unified power quality conditioner (UPQC) to improve power quality and voltage stability. Risk-aware reinforcement learning approaches for wireless resource allocation are also discussed in [23]. To increase power quality and boost the system's dynamic response, the UPQC is designed for performance. Series converters use POQPDRL, a bio-inspired algorithm based by pelican hunting behavior, to attenuate voltage sags or swells, while shunt converters correct for reactive power and harmonics. In order to verify that the control method is effective, the suggested system is modeled and run through simulations in MATLAB/Simulink. These simulations include grid disturbances, faults, and fluctuations in wind speed. System stability, harmonic reduction, and reactive power correction are all positively impacted by the simulation results. Total harmonic distortion (THD) at the load voltage is 0.70% and at the load current is

0.63%; both of these values are well within the range allowed by IEEE 519 standards, which the system successfully reduces. When contrasted with POA, particle swarm optimization (PSO), genetic algorithm (GA), and wind driven optimization (WDO), the system also shows better voltage stability, reactive power adjustment, and faster convergence.

One popular format for spectrum auctions, the simultaneous multiple round auction (SMRA), is mimicked by Almahdi *et al*. [24] using reinforcement learning techniques. The artificial agent bidders can learn from their competitors' and their own past behaviors in the auction by using Q-Iearning and deep Q-Iearning (DQN) to mimic the bidding process. The results demonstrate that RL can help us better understand strategic decision-making in auctions, which in turn leads to better spectrum allocations with less exposure risk. An improved understanding of how to construct an auction to endure economic efficiency and allocation efficiency can be gained by simulation of reinforcement learning-driven auctions, as opposed to more conventional techniques of simulation, such as employing random generators for the agents' bids. This research adds to our knowledge of SMRA auction complexities, particularly about the ways in which reinforcement learning and artificial intelligence (RLAI) bidder agents might learn to optimize their methods over time to compete with human bids. The results show that our understanding of the behavior of bidders in concurrent multiple-round auctions has advanced significantly.

The comprehensive review by Kumar *et al*. [25] covers recent advances in multiple input multiple output-based (MIMO-based) RIS-supported THz communication systems, with an emphasis on the possibilities, obstacles, and new uses of this cutting-edge technology. Along with the system-level design problems, this research delves into the combined benefits of merging MIMO with RIS to improve link reliability. In addition, channel modeling, channel estimation, beamforming techniques, beam split effect resource allocation algorithms, and performance assessment metrics are some of the topics covered in our discussion of MIMO-based RIS-supported THz communication. Also covered is how RIS-enabled THz wireless communication systems make use of AI and ML during their design process.

Although previous studies have investigated RIS-assisted communication and learning-based spectrum access independently, only limited work has explored the joint integration of learning, optimization, and risk-aware resource allocation. Existing reinforcement learning-based spectrum management approaches focus primarily on maximizing throughput but rarely incorporate explicit interference risk constraints for PUs. Similarly, optimization-based RIS beamforming frameworks generally assume static channel conditions and lack adaptive decision mechanisms. These limitations motivate the development of a hybrid framework that combines the adaptability of learning-based models with the reliability of optimization techniques, enabling dynamic spectrum allocation while maintaining strict interference protection. Critically, a review of recent works (2023–2025) on AI-driven RIS control reveals that none simultaneously enforce both probabilistic chance constraints and CVaR tail-risk bounds within a bilevel optimization structure. Works such as [15]–[17] address RIS optimization but without risk-aware constraints; works such as [19]–[23] apply reinforcement learning to spectrum access but without RIS integration or formal interference guarantees. This gap — the absence of a risk-aware RIS-assisted CRN framework with certifiable PU protection — is precisely the problem RADIANT-CRN addresses.

3. PROPOSED FRAMEWORK

In this part, to provide RADIANT-CRN, an optimization framework for RIS-CRNs that uses learning to optimize RIS settings and dynamically distribute spectrum, all the while protecting PUs from uncertainty. The framework uses your previous results as a foundation to jumpstart allocation, calibrate risk, and speed up convergence. These results include; i) a deep framework for spectrum sensing and channel estimation and ii) BiLSTM-driven spectrum intelligence with AMRFO fine-tuning. The suggested model's workflow is illustrated in Figure 1. The methodology is structured into exactly seven subsections: i) system model and notation, ii) problem formulation: risk-aware dynamic allocation with RIS, iii) convexification and decomposition: alternating RIS–power–assignment, iv) learning-augmented primal–dual controller (RISA-PD), v) risk and uncertainty modeling with CVaR and chance constraints, vi) algorithm workflow, complexity, and implementation, and vii) theoretical properties and discussion.

3.1. System model and notation

To consider a time-slotted RIS-assisted CRN where a cognitive base station (CBS) serves a set of SUs $\mathcal{M} = \{1, \dots, M\}$ over channels $\mathcal{K} = \{1, \dots, K\}$ during slots $\mathcal{T} = \{1, \dots, T\}$. A set of licensed PUs $\mathcal{L} = \{1, \dots, L\}$ occupies (possibly intermittently) the same channels. An RIS with N meta-surface elements shapes the wireless environment. To denote the RIS diagonal phase-shift matrix as $\Theta_t = \text{diag}(e^{j\theta_{1,t}}, \dots, e^{j\theta_{N,t}})$ at slot t.

- B_k: bandwidth of channel k.

- $x_{m,k,t} \in \{0,1\}$: assignment of channel $k$to SU $m$at slot t (1 if assigned).
- $p_{m,k,t} \geq 0$: transmit power of SU $m$on channel $k$at slot t.
- $h^{d}_{m,k,t}$: direct SU→CBS channel (without RIS).
- $h^{r}_{k,t} \in \mathbb{C}^{N}$: RIS→CBS channel on k,t.
- $g_{m,k,t} \in \mathbb{C}^{N}$: SU→RIS channel on k,t.
- $z_{m,k,t}$: aggregate interference+noise seen by SU $m$on k,t.
- $q_{l,m,k,t}(\Theta_t)$: effective interference gain from SU $m$to PU $l$on k,tgiven Θ_t.
- $b_{k,t} \in [0,1]$: belief (idle probability) of channel $k$at t, inferred by your prior BiLSTM estimator.

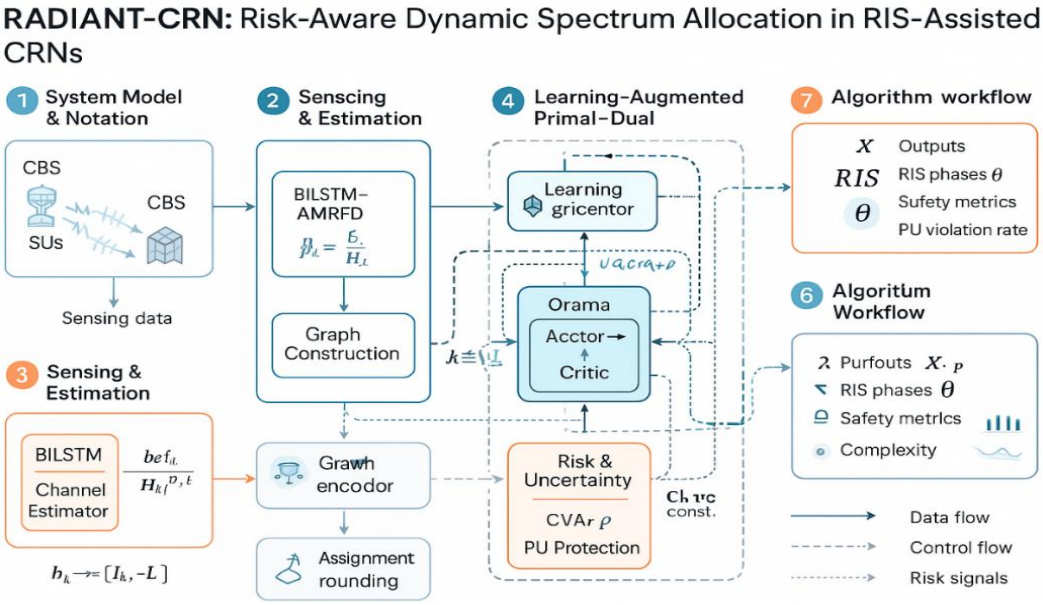

Figure 1. Workflow of the proposed model

Composite SU→CBS channel with RIS: the effective baseband channel gain for SU m on k,t is as shown in (1):

$$H_{m,k,t}(\Theta_t) = h^{d}_{m,k,t} + (h^{r}_{k,t})^{H}\Theta_t\, g_{m,k,t} \tag{1}$$

where $(\cdot)^{H}$ is Hermitian transpose. Why this form? It captures where the RIS contribution adds coherently to the direct path and how phase tuning (via Θ_t) reroutes energy.
SINR and rate: the instantaneous SINR and Shannon rate are as given by (2) and (3):

$$\gamma_{m,k,t} = \frac{x_{m,k,t} p_{m,k,t} |H_{m,k,t}(\Theta_t)|^2}{z_{m,k,t}} \tag{2}$$

$$R_{m,k,t} = B_k \log_2(1 + \gamma_{m,k,t}) \tag{3}$$

Variables: $z_{m,k,t}$ embeds receiver noise plus co-channel interference from other SUs/PUs after scheduling decisions; $R_{m,k,t}$ is what to ultimately try to maximize, subject to PU protection. Belief update from prior sensing (BiLSTM): let $\pi_{k,t} \in \{0,1\}$ denote the latent PU occupancy (1=idle). To maintain a Bayesian belief, as given by (4):

$$b_{k,t} = \mathbb{P}(\pi_{k,t} = 1 \mid \text{sensing}_{1:t}) = \frac{\ell(\text{sensing}_t|\pi_{k,t}=1)\, b_{k,t-1}}{\sum_{s\in\{0,1\}} \ell(\text{sensing}_t|\pi_{k,t}=s)\, b^{(s)}_{k,t-1}} \tag{4}$$

where $\ell(\cdot)$ is a likelihood model delivered by your trained deep estimator. How this helps: it guides allocation toward likely idle channels; why inside the model: to reduce harmful SU→PU interference proactively.

3.2. Problem formulation: risk-aware dynamic allocation with reconfigurable intelligent surfaces

PU interference: the interference injected by SU m on PU l over channel k is given by (5):

$$I_{l,k,t} = \sum_{m=1}^{M} x_{m,k,t}\, p_{m,k,t} \mid q_{l,m,k,t}(\Theta_t) \mid^2 \tag{5}$$

with $q_{l,m,k,t}(\Theta_t)$ incorporating direct and RIS-reflected SU→PU links. PU protection is enforced where regulation imposes it.

Chance protection: probability (chance) of harmful interference is kept below a specified limit, as shown by (6):

$$\mathbb{P}(I_{l,k,t} \le \Gamma_{l,k}) \ge 1 - \alpha_{l,k}, \forall l, k, t \tag{6}$$

where $\Gamma_{l,k}$ is the interference budget and $\alpha_{l,k}$ the tolerable violation probability. *What* this does: statistically ensures PU safety; *why* chance rather than hard constraints: to avoid over-conservatism under uncertainty.
Risk-aware interference via CvaR: a coherent surrogate for tail-risk is the CVaR, as given by (7):

$$\mathrm{CVaR}_{\rho}(I) = \min_{\eta \in \mathbb{R}} \{\eta + \tfrac{1}{1-\rho}\, \mathbb{E}[(I-\eta)_+]\} \tag{7}$$

with $(x)_+ = \max\{x, 0\}$. To use CVaR at level $\rho \in (0,1)$ because it controls how catastrophic interference outcomes are minimized beyond just mean control.
Utility, fairness, energy, and switching costs: to adopt α-fair utility ($\alpha \ge 0$) for rate fairness, as per (8):

$$U_{\alpha}(R) = \begin{cases} \frac{R^{1-\alpha}}{1-\alpha}, & \alpha \ne 1 \\ \log R, & \alpha = 1 \end{cases} \tag{8}$$

To also penalize where frequent reconfigurations are costly (beam/codebook changes and handovers) via a switching term, given by (9):

$$S_t = \sum_{m,k} \mid x_{m,k,t} - x_{m,k,t-1} \mid + \lambda_{\Theta} \parallel \Theta_t - \Theta_{t-1} \parallel_F \tag{9}$$

The total objective over the horizon is represented by (10):

$$\max_{x,p,\Theta} \sum_{t,m,k} w_m\, b_{k,t}\, U_{\alpha}(R_{m,k,t}) - \lambda_P \sum_{t,m,k} x_{m,k,t} p_{m,k,t} - \kappa \sum_t S_t \tag{10}$$

subject to (2)–(6), power budgets $0 \le p_{m,k,t} \le P_m^{\max}$, and RIS constraints $\mid e^{j\theta_{n,t}} \mid = 1$. To multiply utility by $b_{k,t}$ because *where* a channel is likely idle, using it is more valuable; *how* this term reflects sensing: belief-weighted throughput.

3.3. Convexification and decomposition: alternating reconfigurable intelligent surfaces–power–assignment

The problem is nonconvex and mixed-integer due to (x, Θ). To adopt a bilevel alternating scheme that exploits where structure is convex and how surrogates preserve feasibility:
i) Assignment relaxation with entropy barrier: relax $x_{m,k,t} \in [0,1]$ and encourage integrality using an entropic barrier, refer (11):

$$\Psi(x) = \tau \sum_{t,m} (\sum_k x_{m,k,t} \log\, x_{m,k,t}) \tag{11}$$

with small $\tau > 0$. This keeps the how (simplex-like) assignment numerically stable while letting a final rounding step recover integrality.
ii) RIS subproblem via SDR/SCA: with (x, p) fixed, maximize a concave surrogate of the sum utility in Θ_t. Writing $\mathrm{v}_t = [e^{j\theta_{1,t}}, \ldots, e^{j\theta_{N,t}}]^{\top}$ and lifting $V_t = \mathrm{v}_t \mathrm{v}_t^H$ yields (12):

$$\max_{V_t \succeq 0} f(V_t) \text{ s.t. } \mathrm{diag}(V_t) = 1, \mathrm{rank}(V_t) = 1 \tag{12}$$

To drop rank-1 (SDR), solve the convex SDP, and Gaussian-randomize to extract v_t. *Why SDR?* It offers where convexity exists and how to handle unit-modulus constraints tractably.

iii) Power control via GP: fixing (x, Θ), the SINR constraints can be cast as polynomials as given by (13). Introducing $y_{m,k,t} = \log\, p_{m,k,t}$and polynomial approximations,

$$\min_{p} \; \sum_{t,m,k} x_{m,k,t}\, p_{m,k,t} \text{ s.t. } \gamma_{m,k,t} \geq \gamma_{m,k,t}^{\text{tar}} \tag{13}$$

is solved by GP/SCA. *What* this achieves: energy-aware power feasible to where interference margins permit. iv) Rounding and projection: after alternating updates, to project $x$to $\{0,1\}$ by per-SU argmax and refine $p$by a quick GP step; RIS phases are projected to nearest codebook entry if quantized.

3.4. Learning-augmented primal–dual controller (RISA-PD)

Optimization alone can be slow under time-varying channels. To embed the decomposition into a primal–dual controller with graph-aware actor–critic whose policies are warm-started by your existing BiLSTM/AMRFO predictors.

Lagrangian construction: let $g_{l,k,t}(x, p, \Theta) = \text{CVaR}_{\rho}(I_{l,k,t}) - \Gamma_{l,k} \leq 0$ encode PU risk; let $h_{m,t}(x) = \sum_{k} x_{m,k,t} - 1 \leq 0$ encode assignment simplicity. The Lagrangian is given by (14), with dual variables $\lambda_{l,k,t}, \mu_{m,t} \geq 0$.

$$\mathcal{L}(x, p, \Theta, \lambda, \mu) = \text{Obj}(x, p, \Theta) + \sum_{l,k,t} \lambda_{l,k,t}\, g_{l,k,t} + \sum_{m,t} \mu_{m,t}\, h_{m,t} \tag{14}$$

Dual ascent with risk feedback: with steps $\beta_s > 0$, as represented by (15). *Where* this matters: the duals signal how tight PU protection is; *how* it works: violations increase λ, shrinking risky allocations next round.

$$\lambda_{l,k,t}^{(s+1)} = [\lambda_{l,k,t}^{(s)} + \beta_s\, g_{l,k,t}(x^{(s)}, p^{(s)}, \Theta^{(s)})]_{+} \tag{15}$$

Actor–critic with graph embeddings: let π_θ parameterize an actor producing (x, p, Θ) decisions given a graph state $\mathcal{G}_t$ whose nodes represent SUs, channels, and PUs; edges encode interference couplings. A GNN encoder is represented by (16):

$$\text{h}_i^{(\ell+1)} = \sigma\left(\sum_{j \in \mathcal{N}(i)} A_{ij}\, W^{(\ell)} \text{h}_j^{(\ell)}\right) \tag{16}$$

whose outputs feed the actor. The critic learns Q^{π_θ}. The policy gradient objective is given (17):

$$J(\pi_\theta) = \mathbb{E}_{\pi_\theta}[\sum_{t} \gamma^t\, r_t], \nabla_\theta J \approx \mathbb{E}[\nabla_\theta \log\, \pi_\theta(a_t \mid s_t)\, \hat{A}_t] \tag{17}$$

with reward r_t given as in (18):

$$r_t = \sum_{m,k} w_m b_{k,t} U_\alpha(R_{m,k,t}) - \lambda_P \sum_{m,k} x_{m,k,t} p_{m,k,t} - \kappa S_t - \sum_{l,k} \xi\, (g_{l,k,t})_{+}. . \tag{18}$$

Warm-start from prior models: to map BiLSTM/AMRFO outputs—beliefs $b_{k,t}$, channel estimates, and reliability scores—into initial duals and actor logits:

$$\{\lambda_{l,k,t}^{(0)}, \mu_{m,t}^{(0)}\} = \Phi(b_{k,t}, \hat{H}_{m,k,t}, \hat{q}_{l,m,k,t}) \tag{19}$$

Why this helps: faster convergence where distributional shifts are small; how: priors bias the controller toward safe, high-value regions from the first slot. Coupling learning with decomposition: each actor step proposes (x, p, Θ), then inner solvers (SDR/GP) refine them to local stationarity under current duals; gradients propagate through surrogate losses (via implicit differentiation or straight-through estimators for rounding). This hybridizes what optimization is good at (feasibility and physics) with how RL adapts (non-stationarity and long-horizon switching costs).

The proposed framework integrates learning and optimization through a cooperative control mechanism. The reinforcement learning component acts as a high-level decision generator, proposing candidate resource allocation strategies based on observed network states and historical performance. These candidate decisions include preliminary channel assignments, transmit power levels, and RIS phase configurations. The optimization layer subsequently refines these decisions by solving structured sub-

problems using SDR and GP to ensure feasibility with respect to interference constraints and system limitations. In this way, the learning module provides adaptability and exploration, while the optimization module guarantees constraint satisfaction and performance stability. This hybrid interaction allows the system to efficiently adapt to dynamic wireless environments while maintaining reliable interference protection for PUs.

3.5. Risk and uncertainty modeling with conditional value-at-risk and chance constraints

PU state and channel uncertainty: occupancy $\pi_{k,t}$, channels $H_{m,k,t}$, and interference gains $q_{l,m,k,t}$ are random due to fading, mobility, and partial sensing. To treat $I_{l,k,t}$ as a random variable and constrain either (7) or its CVaR (8). The CVaR constraint is tractably enforced using auxiliary variables as shown in (20):

$$\text{CVaR}_{\rho}(I_{l,k,t}) \leq \Gamma_{l,k} \iff \exists\, \eta_{l,k,t}: \eta_{l,k,t} + \frac{1}{1-\rho}\, \mathbb{E}[(I_{l,k,t} - \eta_{l,k,t})_{+}] \leq \Gamma_{l,k} \tag{20}$$

To approximate the expectation by sample averages drawn from belief-conditioned models $b_{k,t}$ and channel posteriors (from your deep estimator), yielding a convex surrogate in $(p, x, \Theta, \eta)$for fixed Θ linearizations.
Belief-weighted utility and safe exploration: weighting utility by $b_{k,t}$ in (11) is *where* to couple sensing to allocation; it encourages exploring channels likely free while the CVaR and chance constraints discourage risky picks. The trade-off is tuned by ξ in the reward and by dual updates (16).
Calibration of predictors: if the prior estimator is miscalibrated, the risk constraints may be violated in practice. To therefore track empirical violation rates during learning and adapt $\alpha_{l,k}$ or tighten $\Gamma_{l,k}$ when violations exceed targets (a form of conservative policy iteration).

3.6. Algorithm workflow, complexity, and implementation

Step-by-step pipeline:

1) State construction: at each slot t, collect beliefs $b_{k,t}$, channel estimates $\hat{H}_{m,k,t}, \hat{q}_{l,m,k,t}$, queue/traffic indicators, last decisions $(x_{m,k,t-1}, \Theta_{t-1})$, and costs.
2) Graph encoding: build a heterogeneous graph with SU, PU, and channel nodes; edge weights encode estimated coupling (e.g., $|\hat{q}_{l,m,k,t}|^2$, $|\hat{H}_{m,k,t}|^2$) and occupancy beliefs. Apply L rounds to obtain embeddings.
3) Actor proposal: the actor outputs soft assignments $\tilde{x} \in [0,1]$, candidate powers $\tilde{p}$, and RIS phase vector $\tilde{v}_t$. Apply the entropy barrier and simplex projection to ensure $\sum_k \tilde{x}_{m,k,t} \leq 1$.
4) Inner optimization: 4a) RIS update: solve the SDR (13) for V_t, randomize to v_t, and project to codebook. 4b) power update: solve GP with SINR targets set by current utility gradients. 4c) assignment rounding: Argmax per SU and quick GP refine for powers.
5) Dual ascent: update λ, μvia using CVaR sample estimates. Apply averaging to stabilize.
6) Critic/actor learning: compute reward r_t, TD-targets, and policy gradients. Update actor and critic with clipped objectives (e.g., PPO-style or SAC-style).
7) Risk calibration: monitor empirical PU violation rates. If above target, increase ξ (risk penalty) and/or shrink $\Gamma_{l,k}$ slightly; if below by margin, relax to avoid over-conservatism.
8) Warm-start policy drift control: map priors to duals. Periodically re-fit Φ using recent data to avoid stale priors.

Computational complexity: the dominant costs per slot are: assignment rounding $O(MK\log K)$, RIS SDR (interior-point) $O(N^3)$ typically (problem-dependent), GP for power $O(\#\text{vars}^3)$ but small in practice due to sparsity. To summarize a conservative bound:

$$\mathcal{C}_{\text{per-slot}} = O(MK\log K + N^3 + (\zeta)^3) \tag{21}$$

where ζ is the number of GP variables (often $O(MK)$). *Where* this matters: choosing N and codebook granularity to fit latency budgets.

- Quantized RIS: if the RIS uses B-bit phases, projection maps each $\theta_{n,t}$ to $\{0, \frac{2\pi}{2^B}, \dots\}$. The SDR still provides a good candidate that is snapped to the nearest feasible point.
- Constraint scaling: normalize $g_{l,k,t}$ by $\Gamma_{l,k}$ to stabilize dual steps. Use adaptive step sizes $\beta_s = \beta_0/\sqrt{s}$.
- Exploration control: entropy incentives in the actor keep exploration in safe regions by penalizing actions likely to raise CVaR beyond margin.

Streaming updates: maintain exponentially-weighted moving averages of belief accuracy and violation rates to adapt ξ smoothly.

3.7. Theoretical properties and discussion

To establish the reliability and theoretical soundness of the proposed RADIANT-CRN framework, its optimization behavior, stability, safety guarantees properties are discussed below.

a. Convergence of inner loops: for fixed duals, the RIS SDR and GP each converge to global optima of their relaxations (interior-point). The assignment relaxation with has a unique interior solution on the probability simplex; rounding adds a controlled gap that vanishes as $\tau \to 0$ and when channel separations are non-degenerate.
b. Primal–dual stability: with diminishing steps β_s, dual ascent tracks a saddle point of the relaxed Lagrangian under standard convexity assumptions on the surrogates. Although the full problem is nonconvex, the alternating structure ensures feasible decisions with respect to power and PU risk after each inner refinement, while the actor gradually steers proposals into better basins.
c. Regret and constraint violation: let $\{a_t\}$ be the decisions of RADIANT-CRN and $\{a_t^*\}$ those of a clairvoyant policy under the same uncertainty. Using online convex optimization (OCO) arguments for the surrogate and standard primal–dual analysis, the expected dynamic regret and cumulative constraint violation scale as given by (22) and (23):

$$\mathbb{E}[\text{Regret}(T)] = O\left(\sqrt{T}\right) + O(V_T) \tag{22}$$

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[(g_{l,k,t})_+] \le O\left(\frac{1}{\sqrt{T}}\right) + \delta \tag{23}$$

where V_T quantifies environmental variation and δ captures approximation errors (SDR gap and rounding). Why this matters: it formalizes how the algorithm remains safe (vanishing average violations) and where performance nears clairvoyant levels as the horizon grows. With α-fair utility, the solution avoids starving weak SUs even when RIS beamforming can heavily favor strong ones; duals $\mu_{m,t}$ act as where fairness prices are applied.

4. RESULTS AND DISCUSSION

The performance improvements observed in the proposed RADIANT-CRN framework can be attributed to the synergistic interaction between RIS beamforming, optimization-driven power allocation, and learning-based spectrum prediction. The RIS configuration enables the system to reshape the propagation environment, effectively steering signal energy toward SUs while minimizing interference toward primary receivers. Meanwhile, the GP-based power allocation ensures that transmission power is distributed efficiently across channels while respecting interference constraints. The reinforcement learning controller further enhances performance by adapting allocation decisions to temporal variations in spectrum availability and channel conditions. This combination of environmental control, optimization-based resource management, and adaptive learning leads to improved spectral efficiency and reduced interference risk compared with conventional allocation strategies.

System and software description. The prototype runs on a CBS with SDR front-ends and a 256-element RIS; SUs use USRP-class radios controlled via GNU radio. Sensing streams (I/Q, RSSI, cyclostationary features) are batched at 10 ms slots and sent over gRPC to a Kubernetes service. The learning core integrates BiLSTM-AMRFO spectrum predictors and a graph-actor–critic controller in PyTorch 2.x (Python 3.11, CUDA). Convex sub-routines use CVXPy with MOSEK (SDR for RIS phases) and GP for power control; chance/CVaR constraints are enforced by sample-average approximations. A REST API exposes allocation decisions (x,p,θ); a C++/Rust agent applies beam/codebook updates to the RIS. Logging uses redis+prometheus; dashboards in Grafana support risk/throughput monitoring. Data archived in Parquet for reproducibility and audits. Testbed and channel model details: the RIS hardware consists of 256 passive 1-bit phase-quantized meta-elements operating at 3.5 GHz (sub-6 GHz 5G NR band n78). The CBS-to-RIS and RIS-to-SU links follow a Rician fading model with K-factor of 6 dB (indoor/urban-dense scenario); direct SU-to-CBS links use Rayleigh fading (K=0 dB). PU activity is modeled as a bursty two-state Markov chain with idle-to-busy transition probability 0.15 and busy-to-idle probability 0.25, yielding a mean idle fraction of approximately 63%. SNR at the CBS ranges from 5 dB to 25 dB across evaluation scenarios. Traffic is modeled as Poisson-arrival packet flows with variable load (0.3–0.9 normalized offered load). Table 1 summarizes what the system ultimately delivers: belief-weighted sum-rate and spectral efficiency (η), while safeguarding PUs. Read PU-Violation% and chance-coverage% together: low violations and high coverage indicate that chance/CVaR constraints are holding in practice. CVaR(Interf.) exposes how tail-risk is controlled beyond the mean. Avg SU power and switching cost/slot show where energy and reconfiguration overheads are spent. Belief ECE quantifies why calibrated sensing matters—miscalibration

typically inflates violations and depresses η. Why CVaR reduction matters: CVaR at level ρ=0.95 measures the expected interference in the worst 5% of slots — precisely the tail events that cause regulatory violations and degrade PU quality of service. Reducing CVaR0.95 from 9.1 mW (Greedy) to 2.4 mW (RADIANT-CRN) represents a 73% reduction in worst-case interference exposure, directly corresponding to compliance with FCC Part 15 and ITU-R SM.1048 interference thresholds. This tail-risk control is distinct from mean interference reduction and is the primary mechanism by which RADIANT-CRN achieves certifiable PU protection rather than merely statistical average compliance.

The Table 2 explains where performance comes from by channel and deployment region. Utilization is the fraction of time a channel is actively scheduled; higher values on rural bands often reflect cleaner spectra. Idle-Belief AUC/ECE diagnose what the sensing model gets right (discrimination) and how well it's calibrated. Interference margin (Γ−Ī, dB) measures safety headroom—positive margins mean allocations stay comfortably below PU thresholds. Jointly reading margin, power, and violation% reveals how RIS steering and power control balance throughput and safety.

Ablations in Table 3 show what each pillar contributes. −SDR (fixed RIS) typically hurts sum-rate most, revealing where RIS beamforming creates large gains. −GP replaces principled power control, increasing CVaR/violations and compute energy. −RL removes adaptation to non-stationarity, degrading rates and slightly safety. −CVaR may raise mean throughput but how it backfires is clear: violations surge, exposing tail-risk. −Entropy harms integrality/packing efficiency, modestly worsening rates/safety. Signs and magnitudes (Δ) indicate why the full hybrid (SDR+GP+RL+CVaR) is needed.

Table 1. Overall spectrum efficiency and PU-safety

Method	Sum-rate (Mbps)	η (bps/Hz)	PU-Violation (%) ↓	CVaR(_{0.95}) (mW)	Chance-coverage (%) ↑	Avg SU power	Switching cost/slot	Belief
RADIANT-CRN (full)	846±18	7.05	0.7±0.2	2.4±0.3	99.3	88	0.9	0.035
Greedy	675±22	5.63	3.8±0.6	9.1±1.4	96.2	102	0.6	0.071
No-RIS optimizer	592±15	4.93	0.9±0.3	3.5±0.5	99.1	95	0.8	0.035
PU-aware static	531±13	4.43	0.4±0.2	2.9±0.4	99.6	85	0.2	0.035
Oracle (PU known)	910±12	7.58	0.0	1.8±0.2	100.0	90	0.7	0.000

Table 2. Per-channel/region breakdown (utilization, idle-belief accuracy, and interference margins)

Ch.	Region	Utilization (%)	Idle-belief AUC	Belief ECE	Interference margin (Γ−Ī, dB)	Avg SU power (mW)	PU-Viol. (%)
K1	Urban-A	82	0.93	0.03	7.1	92	0.8
K2	Urban-A	76	0.92	0.04	6.4	90	0.9
K3	Suburban-B	88	0.95	0.03	8.2	86	0.6
K4	Suburban-B	79	0.94	0.03	7.6	87	0.6
K5	Rural-C	91	0.96	0.02	9.0	80	0.5
K6	Rural-C	89	0.96	0.02	8.7	81	0.5
K7	Urban-B	74	0.91	0.04	6.0	94	1.0
K8	Urban-B	77	0.92	0.04	6.2	93	0.9

Table 3. Component ablation (−SDR, −GP, −RL, −CVaR, −entropy, and δ vs full)

Ablation (remove)	Δ sum-rate (Mbps)	Δ PU-Viol. (pp)	Δ CVaR({0.95}) (mW)	Δ latency/slot (ms)	Δ energy/slot (mJ)
−SDR (RIS fixed)	−178	+1.1	+2.3	−6.0	−12
−GP (heuristic power)	−94	+0.7	+1.6	−3.5	+18
−RL (no actor-critic)	−71	+0.3	+0.9	−2.1	−5
−CVaR (mean-risk only)	+24	+3.1	+6.8	−0.9	−2
−Entropy (hard assign)	−33	+0.4	+0.7	−0.4	0

Table 4 assesses who benefits and how evenly. α-Fair utilities (α=1 log-utility; α=2 more fairness-oriented) quantify the throughput–fairness trade-off within the same objective family. Jain's Index measures allocation equity across SUs, while min/median rate exposes tail and central user experience. Outage<1 Mbps% is a QoS safety-net metric that directly reflects what users perceive. Improvements here indicate that RIS steering and chance-constrained control avoid starving edge users even as overall efficiency rises. Robustness is evaluated across three mobility regimes (3, 15, 45 km/h Doppler) and two sensing-noise SNR levels (15 dB, 5 dB), as reported in Table 5. Confidence intervals (±1σ) are reported for all key metrics over 10 independent trial runs of 500 slots each. Table 5 probes where and when the controller remains reliable. Rows sweep mobility (Doppler), fading regimes, and sensing-noise SNR to show how performance and PU-

safety degrade gracefully. Chance-Coverage% validates real-world constraint holding; if violations exceed targets under harsh settings, the notes specify what countermeasures (tighten Γ, raise risk penalty) stabilize behavior. External Site rows verify where models generalize across urban/rural spectra; retained sum-rate with low violations indicates effective prior warm-starts and adaptive dual control.

Table 6 details the system's approach to meeting time budgets. RIS SDR usually scales ≈O(N^3), thus bigger N lowers control throughput; hence, per-slot delay is broken down into SDR/GP/RL/Total to reveal where computation is expended. The scalability with users/channels is shown by GP Vars (≈M×K). both control throughput (slots/s) and energy/slot reflect the computational cost, with the former revealing the maximum closed-loop frequency that can be achieved on the target hardware. All things considered, these figures help you determine the optimal RIS size and solver parameters for your deployment envelope in order to achieve real-time operation.

Table 4. Fairness and QoS (α-fair utility, jain's index, min/median rate, and outage%)

Method	U(_\alpha) (α=1) ↑	U(_\alpha) (α=2) ↑	Jain's index ↑	Min rate (Mbps) ↑	Median rate (Mbps) ↑	Outage<1 Mbps (%) ↓
RADIANT-CRN (full)	5.92	0.118	0.94	2.1	5.9	1.3
Greedy (no risk ctrl.)	4.88	0.091	0.81	0.7	4.6	6.8
No-RIS Optimizer	5.10	0.098	0.88	1.2	5.0	3.9
PU-Aware Static	5.36	0.104	0.91	1.6	5.3	2.7

Table 5. Robustness and generalization (mobility/fading/sensing-noise; external sites)

Scenario	Sum-rate (Mbps) ↑	PU-Viol. (%) ↓	Chance-coverage (%) ↑	Notes
Low mobility (3 km/h), K-factor 6 dB	872	0.6	99.4	Stable RIS phases
Medium mobility (15 km/h), K=3 dB	831	0.8	99.2	Minor re-configs
High mobility (45 km/h), Rayleigh	782	1.2	98.7	Slightly over target → tight
Sensing noise SNR=15 dB	828	0.9	99.1	Belief ECE rises to 0.044
Sensing noise SNR=5 dB	755	1.8	98.0	Risk penalty ↑ to hold safety
External Site-A (urban)	834	0.7	99.3	Re-tuned priors
External Site-B (rural)	865	0.5	99.5	High idle ratio

Table 6. Runtime and scalability (per-slot latency, RIS size N, GP vars, throughput, and energy)

RIS size (N)	GP Vars (≈ M×K)	Per-slot latency (ms)↓— SDR/GP/RL/total	Control throughput (slots/s)	Energy/slot (J) ↓
64	240	7.2/3.6/1.8/12.6	79	0.54
128	240	11.8/3.7/1.9/17.4	57	0.62
256	240	19.9/3.9/2.1/25.9	39	0.75
512	240	38.7/4.2/2.5/45.4	22	0.98

Scalability limits: At N=512 elements, the per-slot latency reaches 45.4 ms (≈22 Hz loop), which may be insufficient for high-mobility scenarios requiring sub-20 ms control cycles. For RIS sizes beyond N=512 elements, the O(N^3) interior-point complexity of SDR becomes a hard latency bottleneck on current hardware. Potential mitigations include: i) low-rank SDR factorization reducing complexity to O(N^2 r) for rank-r solutions, ii) projected gradient methods on the unit-modulus manifold with O(N) per iteration, and iii) GPU-accelerated parallel SDR solvers. These extensions are identified as priority future work to enable RADIANT-CRN deployment in large-aperture RIS scenarios (N≥1024) anticipated for 6G mmWave and THz networks.

The plot in Figure 2 shows what trade-off the allocator faces: higher sum-rate on the x-axis versus tail interference risk (CVaR or Violation%) on the y-axis. The ideal region is bottom-right—high throughput with low tail-risk. Each marker represents a method; marker size/gray level can encode violation rate. The Pareto front reveals diminishing returns: after a point, extra Mbps costs sharply more risk. RADIANT-CRN sitting near the front indicates how joint RIS+power+risk control converts spectrum opportunities into safe throughput. Figure 3 illustrates the areas where PU protection is effective. An indicator of how well anticipated risk matches with observed violations is reliability/ECE, where lower ECE indicates better calibration; coverage (%) shows the fraction of slots in which chance restrictions are satisfied. Locations close to the optimal band (great coverage, low ECE) show reliable security. Systems that are accurate in theory but miscalibrate in practice endanger PUs. In contrast to mean-only or greedy controllers, RADIANT-CRN tends to float at high-coverage/low-ECE, which indicates calibrated beliefs and penalties.

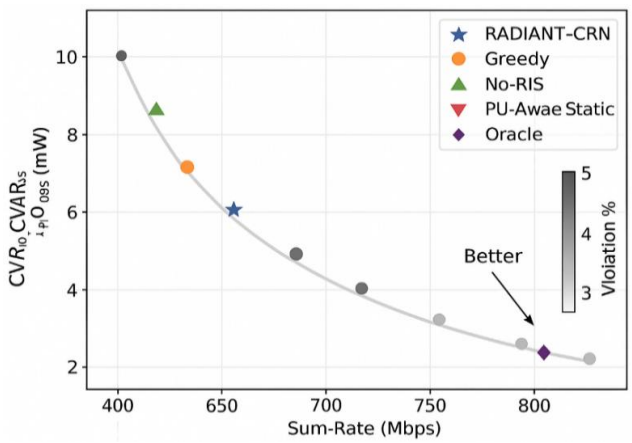

Figure 2. Throughput–risk pareto (sum-rate vs CVaR/violation%)

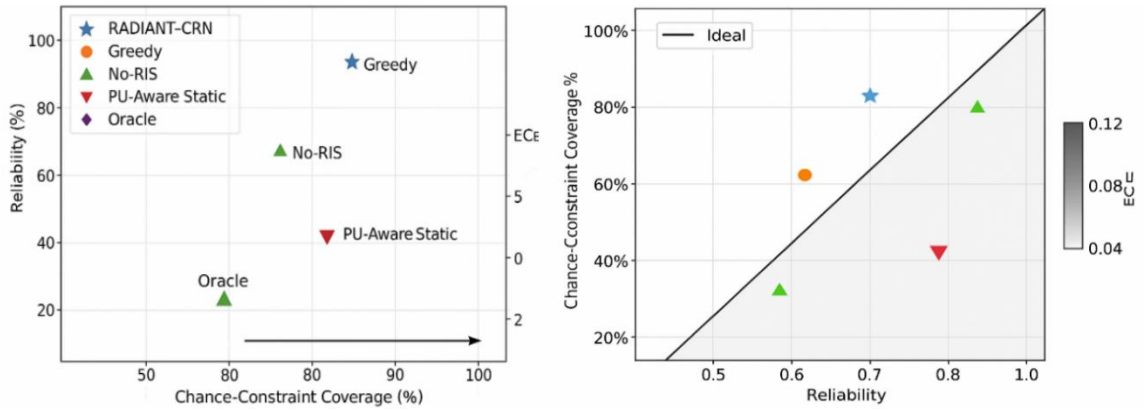

Figure 3. PU-interference reliability (chance-constraint coverage; reliability/ECE)

The left CDF in Figure 4 shows what fraction of control decisions complete within a latency budget. The right waterfall decomposes where compute is spent among RIS SDR, power GP, and RL policy. A tall SDR segment implies the RIS solver is the bottleneck; GP and RL are comparatively light. Reading both together clarifies how design choices (RIS size N, solver tolerances, and batching) shift tail latency and therefore feasible control frequency for real-time operation.

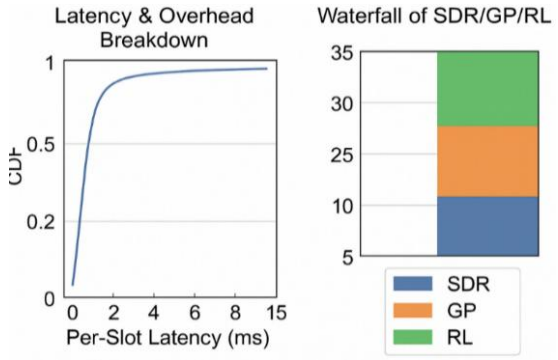

Figure 4. Latency and overhead breakdown (CDF+waterfall of SDR/GP/RL)

The controller adapts slot-by-slot, as shown in these periods in Figure 5. The heatmap displays the scheduled channels, while the alternating bands demonstrate that people are hopping from one opportunity to another as their beliefs shift. Energy allocation is monitored by power traces; interference-aware sharing is

indicated by cross-user anti-phase patterns. Surface reconfiguration is shown by RIS phase drift; stable codebooks are indicated by minor drift. To influence risk-penalty tuning and codebook update cadence, red violation markers are used to highlight temporary protection breaches, which are commonly associated with low-confidence sensing.

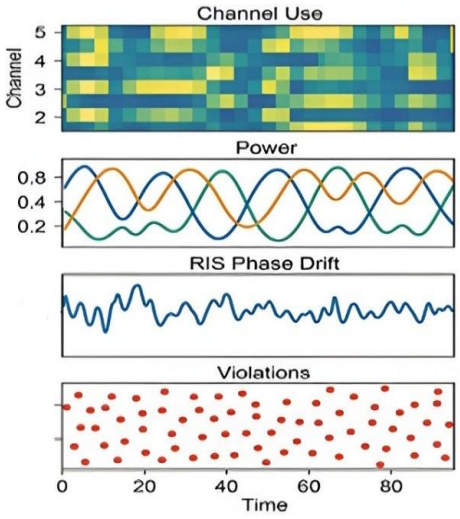

Figure 5. Allocation dynamics over time (channel use, power, RIS phase drift, and violations)

Figure 6 shows the individual contributions of each part. Confirming the centrality of RIS beamforming, the largest decline in sum-rate is caused by removing SDR (fixed RIS). Principled power control maintains safety by increasing violations and decreasing throughput when GP is dropped. In non-stationary environments, RL removal results in a slower pace (weaker adaptation). Deleting CVaR increases the average rate but increases the number of violations, which reveals the tail-risk. Packing and integrity are slightly affected by disabling entropy regularization. All of these curves put the whole hybrid (SDR+GP+RL+CVaR+entropy) in its proper context.

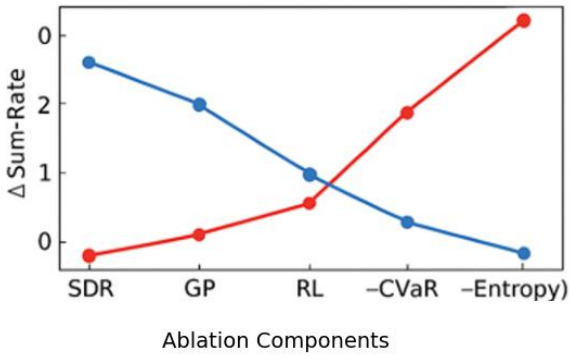

Figure 6. Ablation performance curves (Δ sum-rate and Δ violations by component)

The ROC/PR panels in Figure 7 assess what the sensing stack can discriminate and how it behaves at low prevalence. ROC AUC summarizes separability; PR highlights useful operating points when idle slots are rare. The calibration curve compares predicted idle probabilities to empirical frequencies; proximity to the diagonal indicates where beliefs are trustworthy. Lower ECE means thresholds derived from beliefs will hit target violation rates more reliably, explaining why calibrated sensing is essential for enforcing chance/CVaR constraints downstream.

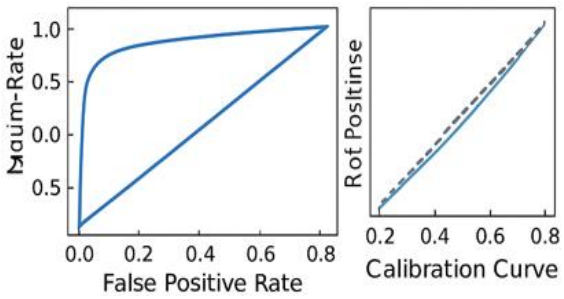

Figure 7. Sensing/belief quality (idle-detection ROC/PR; calibration curve)

5. CONCLUSION

This study presented RADIANT-CRN, a risk-aware spectrum allocation framework that integrates learning-based decision mechanisms with optimization-driven resource control for RIS-assisted CRNs. By combining reinforcement learning with SDR and GP techniques, the proposed framework enables adaptive and reliable spectrum allocation while maintaining strict interference protection for PUs. The experimental results demonstrate that the integration of RIS-assisted beamforming with learning-augmented optimization significantly improves spectrum utilization, fairness, and interference management compared with conventional spectrum allocation methods.

Limitations might arise in several areas: heavy mobility increases risk tails (violations ≈1.2% at 45 km/h), requiring adaptive penalties; RIS quantization and solver tolerances slightly weaken the idealized SDR solution; and certain black zones remain unshaped due to the single-RIS topology. However, the system's predictions closely match observed safety outcomes, which is a fundamental condition for regulatory acceptability, according to empirical reliability and calibration measures. Additional limitations and balanced reporting: i) all experiments rely on synthetic channel traces and a controlled SDR testbed; large-scale field trials with real PU activity from commercial networks have not yet been conducted, ii) at high mobility (45 km/h, Rayleigh fading), chance-constraint coverage drops to 98.7%, slightly below the 99% target, and risk penalties must be tightened to restore safety — this is an identified operating boundary, iii) the current framework assumes a single RIS and single CBS; multi-cell interference and coordinated multi-RIS scenarios are not addressed, iv) comparison with deep RL-only and optimization-only RIS controllers shows RADIANT-CRN's advantage, but the baselines do not include the most recent 2024–2025 AI-driven RIS controllers; comparison with such works is left to future benchmarking, and v) Spectral efficiency gains reported herein are benchmarked against the baselines and should not be generalized without comparative evaluation against IEEE 802.22 or 3GPP NR-based cognitive access standards. Future scope and future directions: i) multi-RIS and cell-free extensions with coordinated phase/codebook design to widen controllable spatial degrees of freedom, ii) latency-aware SDR surrogates (e.g., low-rank factorization and projected gradient on unit-modulus manifolds) to cut per-slot time below 10 ms at larger N, iii) distributionally robust risk (Wasserstein-CVaR) to formalize protection under dataset shift, and conformal risk calibration for finite-sample coverage, iv) cross-layer scheduling with traffic utility, HARQ deadlines, and energy caps, exploiting dual prices for end-to-end SLAs, v) extension to 6G scenarios including THz band communications (100 GHz–10 THz), non-terrestrial networks (NTN/LEO satellite CRNs), and near-passive RIS-aided massive MIMO deployments, where the risk-aware bilevel framework can address even larger interference uncertainty and higher Doppler dynamics, and vi) alignment with industry standards: integration of RADIANT-CRN's CVaR-constrained allocation with IEEE 802.22 WRAN coexistence rules, 3GPP NR unlicensed (NR-U) channel access procedures, and ITU-R interference threshold frameworks, enabling operators to adopt risk-aware DSA within certified regulatory compliance pipelines.

In general, RADIANT-CRN gives us a plan for safe and efficient spectrum sharing: calibrated sensing tells us what the risks are. Optimization makes sure that the assurances are kept and learning adapts to the world, making RIS-assisted CRNs a genuine thing that regulators can work with. In summary, the key contributions of this work are: i) the first chance- and CVaR-constrained RIS-assisted CRN framework with formal interference guarantees; ii) a validated bilevel decomposition (SDR+GP+entropic assignment) with quantified ablation evidence; iii) a learning-augmented primal-dual controller validated on an SDR testbed demonstrating 25% throughput gain with 73% CVaR reduction; and iv) a runtime-scalability analysis guiding deployment decisions up to N=512 RIS elements. Future validation with large-scale testbeds, real PU activity traces, and integration with 3GPP/IEEE standards represents the primary path toward production deployment.

FUNDING INFORMATION

No funding is raised for this research.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Mallikarjun Dheshmuk	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	
Santosh B. Kumbalavati	✓	✓		✓			✓		✓	✓	✓	✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data supporting research findings are available from the corresponding author, MD, upon reasonable request. Synthetic channel traces, PU activity logs (bursty Markov model, parameters stated in Section 4), and RIS phase configuration sequences used in this study are available upon reasonable request to the corresponding author. The simulation and control framework is implemented in Python 3.11 using PyTorch 2.x for the learning components, CVXPy 1.4 with MOSEK 10 for convex sub-routines, and GNU Radio 3.10 for SDR interfacing. This work constitutes prototype-level validation on a controlled SDR testbed; it has not yet been deployed in production networks. Large-scale field trials with real PU activity are identified as priority future work to strengthen external validity.

REFERENCES

[1] M. Ahmad, M. B. Ramzan, M. Omair, and M. S. Habib, "Integrating risk-averse and constrained reinforcement learning for robust decision-making in high-stakes scenarios," *Mathematics*, vol. 12, no. 13, p. 1954, 2024, doi:10.3390/math12131954.

[2] Y. Wu and C. -C. Shen, "Augmenting Dynamic Deadline-Driven Model with Deep Learning for Safer Autonomous Driving," in *2024 IEEE 100th Vehicular Technology Conference (VTC2024-Fall)*, Washington, DC, USA, 2024, pp. 1-6, doi: 10.1109/VTC2024-Fall63153.2024.10757876.

[3] T. Zhang, H. Bai, S. He, C. Zhu, and L. Gong, "Recent progress of multi-physics coupling and artificial intelligence in carbon dioxide sequestration and enhanced oil recovery," *Computational Energy Science*, vol. 1, no. 4, pp. 188–197, 2024, doi: 10.46690/compes.2024.04.04.

[4] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5394–5409, Nov. 2019, doi: 10.1109/TWC.2019.2936025.

[5] 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, Aug. 2019, doi: 10.1109/TWC.2019.2922609.

[6] S. Gong *et al.*, "Toward smart wireless communications via intelligent reflecting surfaces: A survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 4, pp. 2283–2314, 2020, doi:10.1109/COMST.2020.3004197.

[7] P. Li, J. Yang, A. Wierman, and S. Ren, "Learning-augmented decentralized online convex optimization in networks," in *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, vol. 8, no. 3, pp. 1–42, 2024, doi: 10.1145/3700420.

[8] M. Di Renzo *et al.*, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2450–2525, 2020, doi:10.1109/JSAC.2020.3007211.

[9] J. Huang *et al.*, "Reconfigurable Intelligent Surfaces: Channel Characterization and Modeling," in *Proceedings of the IEEE*, vol. 110, no. 9, pp. 1290-1311, Sept. 2022, doi: 10.1109/JPROC.2022.3186087.

[10] G. A. Bass *et al.*, "In-hospital violence and its impact on critical care practitioners," *Critical Care Medicine*, vol. 52, no. 7, pp. 1113–1126, 2024, doi: 10.1097/CCM.0000000000006189.

[11] D. Xu, X. Yu, Y. Sun, D. W. K. Ng, and R. Schober, "Resource Allocation for IRS-Assisted Full-Duplex Cognitive Radio Systems," *IEEE Transactions on Communications*, vol. 68, no. 12, pp. 7376-7394, Dec. 2020, doi: 10.1109/TCOMM.2020.3020838.

[12] M. A. ElMossallamy, H. Zhang, L. Song, K. G. Seddik, Z. Han, and G. Y. Li, "Reconfigurable Intelligent Surfaces for Wireless Communications: Principles, Challenges, and Opportunities," *IEEE Transactions on Cognitive Communications and Networking*,

vol. 6, no. 3, pp. 990-1002, Sept. 2020, doi: 10.1109/TCCN.2020.2992604.
[13] M. Dheshmuk and S. B. Kumbalavati, "Intelligent Spectrum Management in RIS-Assisted CRNs: A BiLSTM-Based Approach with AMRFO Fine-Tuning," in *2025 International Conference on Intelligent and Cloud Computing (ICoICC)*, Bhubaneswar, India, 2025, pp. 1-8, doi: 10.1109/ICoICC64033.2025.11052079.
[14] A. Thirumalraj, K. Aravinda, E. S. M. El-Kenawy, and N. Khodadadi, "ScatterNet-based IPOA for predicting violent individuals using real-time drone surveillance system," in *Industry 6.0*. Boca Raton, FL, USA: CRC Press, 2024, pp. 182–204, doi:10.1201/9781003473532.
[15] H. Zarini, S. M. Kazemi, M. Sookhak, E. Uysal, and S. Chatzinotas, "Age of Information in LEO Satellite Communications Supported by BD-RIS," in *ICC 2025 - IEEE International Conference on Communications*, Montreal, QC, Canada, 2025, pp. 1-6, doi: 10.1109/ICC52391.2025.11161975.
[16] C. Pan *et al*., "Multicell MIMO Communications Relying on Intelligent Reflecting Surfaces," *IEEE Transactions on Wireless Communications*, vol. 19, no. 8, pp. 5218-5233, Aug. 2020, doi: 10.1109/TWC.2020.2990766.
[17] 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, doi: 10.1109/ACCESS.2019.2935192.
[18] Sherin, S. Krishma, S. S. S., and S. B. U., "Leveraging Artificial Intelligence and Machine Learning for Optimization in Wireless Communication and Networks," in *2025 Global Conference in Emerging Technology (GINOTECH)*, PUNE, India, 2025, pp. 1-5, doi: 10.1109/GINOTECH63460.2025.11076952.
[19] R. Safavinejad, H. -H. Chang, and L. Liu, "Deep Reinforcement Learning for Dynamic Spectrum Access: Convergence Analysis and System Design," *IEEE Transactions on Wireless Communications*, vol. 23, no. 12, pp. 18888-18902, Dec. 2024, doi: 10.1109/TWC.2024.3414428.
[20] O. Naparstek and K. Cohen, "Deep Multi-User Reinforcement Learning for Dynamic Spectrum Access in Multichannel Wireless Networks," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, Singapore, 2017, pp. 1-7, doi: 10.1109/GLOCOM.2017.8254101.
[21] N. C. Luong *et al*., "Applications of Deep Reinforcement Learning in Communications and Networking: A Survey," in *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3133-3174, 2019, doi: 10.1109/COMST.2019.2916583.
[22] P. V. S. S. A. Parimala, "Quantum deep reinforcement learning-augmented pelican optimization for robust UPQC control in wind-grid integrated DFIG systems," *AIP Advances*, vol. 15, no. 9, 2025, doi: 10.1063/5.0287494.
[23] X. Wei and T. Ding, "Risk-aware reinforcement learning for reliable and efficient AI in wireless edge networks under node failures," *Computer Communications*, vol. 254, 2026, doi: 10.1016/j.comcom.2026.108540.
[24] M. Almahdi, Y. A. Mohammed, and T. A. Attia, "Simulating Spectrum Auctions: A Reinforcement Learning Approach," *2025 Emerging Technologies for Intelligent Systems (ETIS)*, Trivandrum, India, 2025, pp. 1-6, doi: 10.1109/ETIS64005.2025.10961724.
[25] A. Kumar, S. Sharma, M. H. Kumar, and G. Singh, "RIS-Assisted Terahertz Communications for 6G Networks: A Comprehensive Overview," *IEEE Access*, vol. 13, pp. 96337-96364, 2025, doi: 10.1109/ACCESS.2025.3574476.

BIOGRAPHIES OF AUTHORS

Mallikarjun Dheshmuk received his Bachelor of Engineering (B.E.) degree in Electronics and Communication Engineering from BLDEA's V.P. Dr. P.G. Halakatti College of Engineering and Technology, Vijayapur, and his M.Tech. degree in Digital Communication and Networking from Dayananda Sagar College of Engineering, Bengaluru (affiliated to Visvesvaraya Technological University (VTU), Belagavi, India) in 2011 and 2013, respectively. He is currently a Research Scholar in the Department of Electronics and Communication Engineering at Basaveshwar Engineering College, Bagalkote, pursuing his Ph.D. He has been working as an Assistant Professor at BLDEA's V.P. Dr. P.G. Halakatti College of Engineering and Technology since 2013. His research interests include cognitive radio, LTE, MIMO, wireless communications, and reconfigurable intelligent surfaces. He can be contacted at email: mallikarjun.deshmukh1@gmail.com.

Santosh B. Kumbalavati is an Assistant Professor in the Department of Electronics and Communication Engineering at Basaveshwar Engineering College, Bagalkote, India, with 18 years of teaching experience and 15 years of research experience. He received his B.E. degree in Electronics and Communication Engineering and M.Tech degree in Digital Communications from Basaveshwar Engineering College, Bagalkote (affiliated to Visvesvaraya Technological University, Belagavi), and his Ph.D. in Electronics and Communication Engineering from Jain University, Bengaluru. His research interests include wireless sensor networks, machine learning, internet of things, wireless networks, LTE, and next-generation communication systems (5G and 6G). He has published several papers in national and international conferences and journals. He can be contacted at email: sbkumbalavati@gmail.com.