REMARKABLE: RIS-Enabled Mobile Beamforming through Kernalized Bandit Learning

Kubra Alemdar*, Arnob Ghosh†, Vini Chaudhary‡, Ness Shroff†, Kaushik R. Chowdhury+
*Institute for the Wireless Internet of Things, Northeastern University, †The Ohio State University, ‡Mississipi State University, +Wireless Networking and Communications Group, The University of Texas Austin

Abstract

Mobile Robots (MRs), typically equipped with single-antenna radios, face many challenges in maintaining reliable connectivity established by multiple wireless access points (APs). These challenges include the absence of direct line-of-sight (LoS), ineffective beam searching due to the time-varying channel, and interference constraints. This paper presents REMARKABLE, an online learning based adaptive beam selection strategy for robot connectivity that trains kernelized bandit model directly in real-world settings of a factory floor. REMARKABLE employs *reconfigurable intelligent surfaces* (RISs) with passive reflective elements to create beamforming toward target robots, eliminating the need for multiple APs. We develop a method to create a beamforming codebook, reducing the search space complexity. We also develop a *reconfigurable rotational mechanism* to expand RIS coverage by rotating its projection plane. To address non-stationary conditions, we adopt the *bandit over bandit* idea that employs adaptive restarts, allowing the system to forget outdated observations and safely relearn the optimal interference-constrained beam. We show that our approach achieves a dynamic regret and the violation bound of $\tilde{O}(T^{3/4}B^{1/4})$ where T is the total time, and B is the total variation budget which captures the total changes in the environment without even assuming the knowledge of B. Finally, experimental validation with custom-designed RIS hardware and mobile robots demonstrates 46.8% faster beam selection and 94.2% accuracy, outperforming classical methods across diverse mobility settings.

CCS Concepts

• **Hardware** → *Analysis and design of emerging devices and systems*;
• **Theory of computation** → *Online learning algorithms.*

Keywords

reconfigurable intelligent surfaces, beam selection, online bandit learning, time-varying channels, mobile robot networks

ACM Reference Format:
Kubra Alemdar*, Arnob Ghosh†, Vini Chaudhary‡, Ness Shroff†, Kaushik R. Chowdhury+. 2025. REMARKABLE: RIS-Enabled Mobile Beamforming through Kernalized Bandit Learning. In *The Twenty-sixth International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing (MobiHoc '25), October 27–30, 2025, Houston, TX, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3704413.3764443

ACM ISBN 979-8-4007-1353-8/2025/10
https://doi.org/10.1145/3704413.3764443

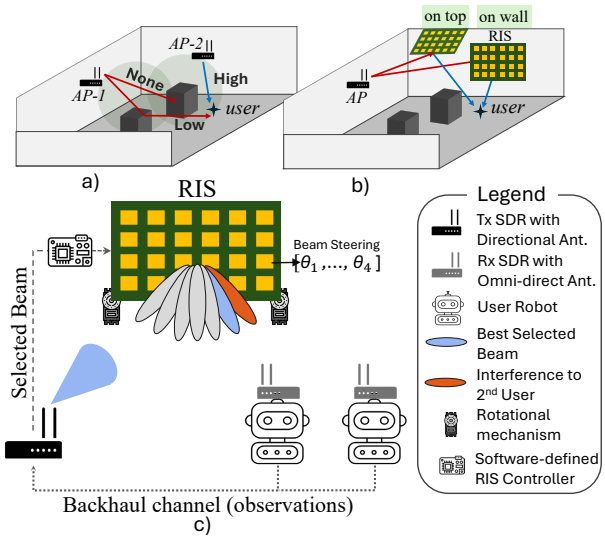

Figure 1: Illustration of mobile robot connectivity in a factory floor when; a) multiple APs deployed to mitigate the issue of blockages, but creating interference among APs, b) solution to the issue in (a) via RIS deployment on the ceiling/wall, and c) REMARKABLE system architecture presenting beam selection for MRs through RIS.

1 Introduction

Industry 4.0 is set to revolutionize manufacturing and industrial services through the digital transformation of the field, enabling real-time decision-making and automation [1]. Technologies such as the Internet of Things (IoT), artificial intelligence (AI), cloud connectivity, large-scale machine-to-machine communication (M2M), and networked, mobile robotics are central to the future of manufacturing [2]. These robots can collect and analyze data, making autonomous decisions with minimal human intervention. However, such a robot-enabled network architecture demands ultra-fast data transfer speeds, exceptional reliability, and minimal latency [3]. While network densification is a possible path to achieve these goals, it involves a significant cost overhead [4], and thus, network designers need to trade-off permanent infrastructure installations with reconfigurable platforms that can adapt to robot mobility over time. Aside of mobility, the harsh propagation environment within the factory floors increases blockages, results in limited coverage, and significant path loss. To overcome these challenges, recent studies have shown that programmable wireless environments that enable reconfigurability by shaping signal reflections can improve signal-to-noise ratio (SNR) [5] and expand coverage [6].

In REMARKABLE, we realize such a network architecture with low-mobility MRs and multiple wireless APs deployed in a robotic factory floor within a rich-scattering environment. Network reconfigurability for enhancing the APs' coverage is achieved by

controlling the propagation environment using software-controlled reconfigurable intelligent surfaces (RISs) [7–9]. This ensures reliable connectivity for low-mobility MRs by creating a radio environment that adapts with the MR location, as in Fig. 1a-b. Yet, this requires a solution to the problem of adaptive beam selection in dynamic channel conditions between the AP and the MR, as depicted in Fig. 1c.

1.1 Factory-Floor Networking Challenges

• **Problem 1** ***(blockage and coverage)***: Consider Fig. 1a, showing a factory floor where the LoS signal is blocked by obstacles. In absence of LoS conditions, the MR relies on the strongest non-line-of-sight (NLoS) reflection to establish a communication link with the AP. NLoS multipath components can cause destructive interference due to uncontrollable phase reflections, leading to significant communication disruption. This results in significant received signal strength (RSS) fluctuations with small robot movements. This issue is exacerbated in single-antenna equipped MRs. Deploying multiple APs as in Fig. 1a, ensures LoS links and expands coverage but increases communication overhead and infrastructure complexity.

• **Problem 2** ***(mobility and beam searching)***: Narrow beams formed via phased antenna arrays can mitigate propagation loss, as well as improve signal reception through increased SNR. Typically, these beams are formed by adjusting antenna element weights, with steering directivity and beamwidth defined through a codebook. The APs equipped with such capabilities exhaustively sweep over the beams in a codebook to discover the optimal beam with the highest signal strength. However, exhaustive beam searches create significant overhead, and MR mobility requires repeated searches to maintain connectivity.

• **Problem 3** ***(interference)***: Even with APs forming highly directional beams towards MRs, in a heterogeneous environment with multiple APs, the close proximity of MRs can cause excessive interference , degrading network performance [10]. Therefore, beam selection must be judiciously performed, as some beam candidates may not be suitable for data transmission.

1.2 Proposed Approach

Our approach aims to tackle problems 1, 2, and 3 for MR connectivity in factory floor settings by achieving the following steps:

1) We design a passive beamformer using an RIS, a planar array of passive reflective elements, each configured to adjust the amplitude and phase of incident signals. This allows us to create various beam patterns, which are then used to form a beam codebook for the beam steering (see Fig. 1c). To address *Problem 1* (see Fig. 1b), several practical challenges must be considered: (i) Instead of relying on Channel State Information (CSI), our approach uses a predefined codebook where each codeword corresponds to the weights of RIS elements. This is necessary because RIS elements are passive and lack radio chains, making traditional channel estimation impractical. Estimating each channel component—would be proportional to the number of reflective elements, would create extreme overhead. Therefore, we develop a method for creating the desired reflection beam pattern by using a non-uniform phase sampling technique, optimizing each element's reflection gain while considering incident and reflection signals. ii) In a planar RIS, edge elements of RIS contribute less to beamforming, limiting overall gain—especially in dynamic settings like mobile robots. To address this challenge, we propose a novel *reconfigurable rotation mechanism* that adjusts the pitch and roll angles along the RIS's local coordinate axes, effectively enhancing beam coverage and improving performance. Given a fixed AP location, this method requires reflective beam pattern synthesis w.r.t the new angular domain of RIS. Consequently, we generate a multi-level codebook, each level corresponding to a specific pair of rotational angles.

2) We study beam selection using codebooks derived from reflective beam-pattern synthesis. Our goal is to learn online the optimal beam from the RIS to the MR by casting the task as a kernelized multi-armed bandit (MAB), with each codeword as an arm. To address *Problem 2* and *Problem 3*, we impose an interference constraint at a neighboring MR. The objective is to select beams that maximize RSS at the target MR subject to this constraint over a time-varying channel. We model cross-beam correlations with a Gaussian Process (GP) bandit and propose a primal–dual GP-Upper Confidence Bound (UCB) algorithm to balance exploration and exploitation while enforcing the interference constraint. To handle non-stationarity, we add an *adaptive restart* mechanism inspired by the *bandit-over-bandit* framework, which dynamically tunes the restart interval from feedback. REMARKABLE is theoretically grounded and validated on a real RIS-enabled robotic testbed—unlike prior theoretical works [11, 12], which remain untested in practice, and existing RIS implementations [5, 13, 14], which predominantly target static scenarios.

1.3 Summary of Contributions

(1) We create an RIS codebook with beam patterns in multiple directions, enabling the online learning algorithm to find the best beam without channel estimation. Additionally, we introduce a reconfigurable rotational mechanism to expand RIS coverage.
(2) We formulate beam selection for an MR as a primal-dual GP-UCB framework to maximize signal strength while avoiding interference. To address the *time-varying* or *non-stationarity*, we adopt *"bandit over bandit"* concept restart strategy, which adaptively forgets past data by tuning the restart interval via an adversarial bandit. Our method achieves sub-linear dynamic regret and constraint violation without prior knowledge of budget variations, *safely learning* beam selection even under a time-varying channel.
(3) We show that we achieve $\tilde{O}(B^{1/4}T^{3/4})$ regret and $\tilde{O}(B^{1/4}T^{3/4})$ violation bound where B indicates the total change in the environment (i.e., the change in the reward and the constraint, defined in Sec. 6.3.1) over T time steps. We improve the existing bound of $\tilde{O}(BT^{3/4})$ dynamic regret and the dynamic violation bound achieved in [12].
(4) We demonstrate REMARKABLE in a real-world setting using USRP X310-B210 SDRs (Software Defined Radios), with MRs and a PCB-fabricated RIS, as shown in Fig. 1c. Our results show that REMARKABLE achieves 46.8% improved performance over classical methods with 94.2% selection accuracy.
(5) We release the software pipeline for the online learning framework and the RIS configuration-orchestration software [15].

2 Related Work

• **RIS & Smart Surface:** RIS technology and similar concepts like metasurfaces have recently been proposed to enhance applications

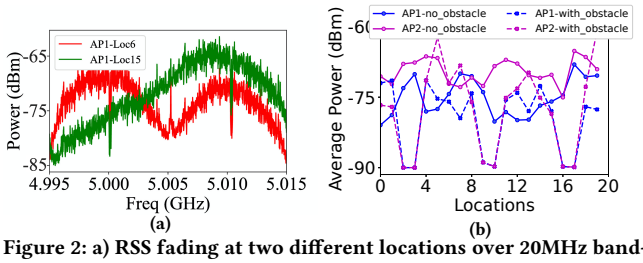

Figure 2: a) RSS fading at two different locations over 20MHz bandwidth; b) Average RSS measured per location where there are obstacles or not.

such as security [16], virtual reality [17], localization and sensing [18], beamforming [5, 19], and over-the-air aggregation [20]. Recent research has focused on optimizing transceivers and RIS phase-shifts to minimize signal distortion [19, 20], especially with imperfect CSI estimation [21]. However, these methods assume knowledge of wireless channels, which is challenging due to RIS's limited processing capabilities. Additionally, some practical works [5, 13, 14, 18, 22] rely on real-time channel estimation, causing overhead proportional to system size and requiring fast feedback. In contrast, REMARKABLE uses a predefined RIS codebook, avoiding such overhead. Similar to our approach, other works consider configuring RIS elements with pre-defined coding patterns [20, 23] and leveraging an extra degree of freedom by optimizing rotation of RIS plane/elements to improve system performance [24, 25]. In fact, the work [11] proposes a hierarchical codebook-generating method using pattern synthesis, followed by a beam training method using the two-mainlobe codewords from the designed codebook for beam sweeping. Unlike these stationary setups that use exhaustive beam searching, REMARKABLE offers a novel method for faster beam selection, even considering mobility.

• **Beam Selection with Bandit:** Online Learning (OL), particularly MAB frameworks, has become prominent for beam selection due to its inherent ability to balance exploration and exploitation. Standard MAB frameworks utilized in beam selection cannot capture the correlation among beam directions. The authors in [26–29] leverage *contextual* information to exploit such correlations. However, these papers do not consider the time-varying channel and interference constraints that we considered, assuming quasi-static channels; thereby, driven models cannot capture time-varying channels, most likely mapped to real-world settings. A recent kernelized MAB approach [12] addresses time-varying, interference-constrained channels but neglects RIS settings and lacks real-world implementation. Our work explicitly incorporates RIS, demonstrates improved theoretical bounds compared to [12], and provides experimental validation in practical scenarios.

3 Motivation for Designing REMARKABLE

Before designing REMARKABLE, we conduct preliminary experiments in a factory floor use-case to investigate *Problem 1.*

• **Experimental Setup:** Consider a scenario where low-mobility MRs roam a factory floor to complete assigned tasks (see Fig. 5). We use a Turtlebot2 robot, which navigates the floor and stops at target locations to collect data. The data collection part is obtained at the 5GHz band with 20MHz bandwidth signal by SDR X310 radios equipped with omni-directional VERT2450 tx-rx antennas, where one of them is mounted on the Thurtlebot, while the other two are placed in designated areas in the environment as APs to communicate with the MR.

• **Observation:** We conducted two factory-floor experiments—one with obstacles and one without—while the MR navigated and RSS was measured from each AP at target locations. As shown in Fig. 2a, the AP1–MR channel exhibits frequency-selective fading that varies with MR position, whereas Fig. 2b shows AP2 providing better coverage in regions where AP1 is weak, even without obstacles. However, with obstacles, neither AP ensures reliable coverage, indicating the need for additional APs. This, in turn, introduces interference management challenges and increases overhead in terms of coordination and communication resources (e.g., bandwidth).

4 REMARKABLE Codebook Design

We aim to create a codebook of beam patterns by optimizing the phases of RIS's reflective elements to achieve the desired reflections. We start by looking at a scenario with a single AP and a single RIS. The RIS is a planar array with $N \times N$ passive reflective elements that can be configured for complex-valued amplitude and phase changes. Moreover, each generated codebook should consist of beams with predefined beam resolution and cover desired angular space.

4.1 Beam Steering Design

With $N \times N$ layout RIS, we can derive the far-field reflection gain pattern of the surface w.r.t a specific target angle as:

$$F(\phi_r, \theta_r) = \sum_{m,n=0}^{(N-1)} G_{mn}(\phi_i, \theta_i)\Gamma_{mn} e^{j((mu_r+nv_r)-(mu_i+nv_i))} \quad (1)$$

where $u_i = kd_x cos\phi_i sin\theta_i$ and $v_i = kd_y sin\phi_i sin\theta_i$ are u-v space coordinates when the source signal contacts on different reflective elements from AP with azimuth and elevation angle of ϕ_i and θ_i. Similarly, the term $u_r = kd_x cos\phi_r sin\theta_r$ and $v_r = kd_y sin\phi_r sin\theta_r$ represent when the signal reflects from the surface towards the target with the azimuth and elevation angle, ϕ_r and θ_r, respectively. The reflective elements are placed in half wavelengths along the x and y directions, $d_x = d_y = \lambda/2$, also $k = 2\pi/\lambda$, and λ is the wavelength of the operational frequency. Eventually, the signal path follows the incident angle and impacts the surface. Then, on the surface, it will be perturbed by the configurations of reflective elements. This is represented by the term of *reflection coefficient*, $\Gamma_{mn} = |\Gamma_{mn}|e^{j\Phi_{mn}}$, wherein the complex-valued amplitude and phase changes are applied to the incident signal. Assuming the reflection magnitude of all reflective elements is unity, i.e., $|\Gamma_{mn}| = 1$. Additionally, we denote $G_{mn}(\phi_i, \theta_i)$ as radiated gain per reflective element defined as $G_{mn}(\phi_i, \theta_i) = (cos^2\phi_i cos^2\theta_i + sin^2\theta_i)|F_e(\phi_i, \theta_i)|^2$ [30]. Here, F_e is obtained from estimated full-wave simulation in the Ansys HFSS 3D electromagnetic simulator. By re-forming Eq.1, we can transform reflection pattern to:

$$F(u_r, v_r) = \sum_{m,n=0}^{(N-1)} A_{mn}(\phi_i, \theta_i) e^{j(mu_r+nv_r)} \quad (2)$$

$$A_{mn}(\phi_i, \theta_i) = \underbrace{G_{mn}(\phi_i, \theta_i)|\Gamma_{mn}|}_{a_{mn}} \underbrace{e^{-j(mu_i+nv_i-\Phi_{mn})}}_{e^{j\Phi^s_{mn}}} \quad (3)$$

where $\Phi^s_{mn} = \Phi_{mn} - (mu_i + nv_i)$ is constant due to fixed locations of AP and the RIS. From the sampling theory for 2D periodic functions, the reflection array complex weights, A_{mn} can be obtained from the samples of its radiation pattern $|F(u_r, v_r)|$ as follows:

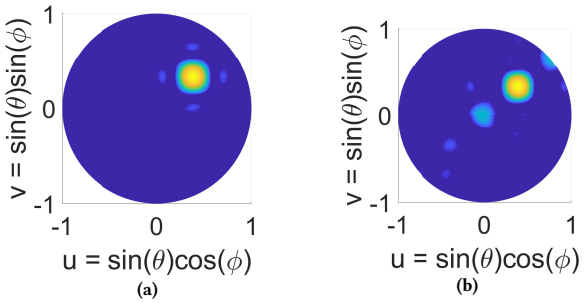

Figure 3: a) Desired and b) generated beam patterns in u-v coordinates

$$A_{mn}(\phi_i, \theta_i) = \sum_{p,q=-(N-1)/2}^{(N-1)/2} F(u_p, v_q) e^{-j(m' u_p + n' v_q)} \quad (4)$$

In Eq.4, u_p, v_q are sampling points for the RIS, where $u_p = 2\pi p/N$ and $v_q = 2\pi q/N$. Also, $m' = m-(N-1)/2$ and $n' = n-(N-1)/2$ for $m, n \in [0, N-1]$. In this method, we assign nonuniform phases to the radiation patterns of RIS at the sampling points via $F(u_p, v_q) = |F(u_p, v_q)|e^{j\Phi_{pq}}$[31], where Φ_{pq} is the phase assigned to sampling points (u_p, v_q). Then, we can find optimized phase values needed to steer the beam in the intended direction by minimizing the Mean Square Error (MSE) function between desired, $\hat{I}_{mn}$, and generated, I_{mn}, power of reflection array weights such as $I_{mn} = a_{mn}^2$, that follows $MSE = \frac{1}{N^2}\sum_{m=0}^{(N-1)}\sum_{n=0}^{(N-1)}(I_{mn} - \hat{I}_{mn})^2$. We apply Gradient Descent (GD)[32] optimization method to minimize MSE. First, we define the gradient of MSE, $\partial MSE/\partial\Phi_{pq}$ w.r.t non-uniform sampling points of the regenerated beam pattern, which is calculated by chain rule as in $\frac{\partial MSE}{\partial \Phi_{pq}} = \sum_{m=0}^{N-1}\sum_{n=0}^{N-1}\frac{\partial MSE}{\partial I_{mn}}\frac{\partial I_{mn}}{\partial \Phi_{pq}}$. Here, both derivatives in the chain are derived independently, then we can have:

$$\frac{\partial MSE}{\partial \Phi_{pq}} = \frac{4}{N^2}\Im\left\{e^{-j\phi_{pq}}\sum_{m,n=0}^{N-1} M_{mn} e^{j\left\{\Phi_{mn}^s + \Phi_f\right\}}\right\} \quad (5)$$

where $M_{mn} = a_{mn}(a_{mn}^2 - \hat{I}_{mn})$ and $\Phi_f = m'u_p + n'v_p$. Since Eq.5 is a form of Fourier Transform, we utilize the Fast Fourier Transform (FFT) techniques to calculate gradients. Then at each iteration, new non-uniform phase samples are found as:

$$\Phi_{pq,r+1} = \Phi_{pq,r} + \nabla\Phi_{pq,r} = \Phi_{pq,r} - \eta_g \frac{\partial MSE}{\partial \Phi_{pq}} \quad (6)$$

where η_g is the learning rate of the GD optimizer, determining the step size to converge it to the optimal point. After obtaining optimized non-uniform phase samples, we can find the phase distribution of the surface, Φ_{mn}, by considering the amplitude and assigned phases, Φ_{pq}, of the radiation pattern $F(u_p, v_q)$ through Eq.4. Fig. 3 compares the desired and generated pattern at steered angles of (40°,30°), showing that desired pattern can be achieved by our method, albeit with some increased side slopes due to the effect of quantization.

4.2 Beam Coverage Design

Here, we address RIS's limitation in terms of its angular coverage and propose a solution to mitigate this by integrating a reconfigurable rotational mechanism.

4.2.1 Addressing the Coverage Problem of RIS. The steering capabilities of a planar array RIS are limited by target direction since not all reflective elements contribute equally to the beam's reflective gain, especially if the direction is near the edge of the RIS coverage area. Additionally, RIS coverage depends on its relative size, making it challenging to cover the entire angular space. To illustrate, we simulate a 9×9 inset-fed patch antenna array, measuring beamforming gain at two angular locations. The RIS is placed along the x-y plane with the mn^{th} element at $(mdx, ndy, 0)$, and measurements are taken along the z-axis for azimuth, ϕ_r, and elevation, θ_r. By manipulating the phase values of each antenna, we form two distinct beams toward the targets at $(3.3m, 20°, 40°)$ and $(3.3m, 66°, 40°)$, as in (r, ϕ, θ). Fig. 4a-4b show that the larger-azimuth target suffers a 25% gain loss compared to the one closer to the center, and the boresight beamwidth narrows undesirably (see Fig. 4b). We propose enhancing RIS coverage by adding a reconfigurable rotational mechanism to adjust its orientation through *pitch* and *roll* angles. The *yaw* angle is not used as the z-axis is the projection axis of RIS. Implementing this method necessitates rebuilding the codebook structure and reformulating the beam shaping and steering process.

4.2.2 Beam Synthesis for Rotated RIS. In our scenario, with the location of the AP fixed, we only need to rotate the RIS to cover different sectors in the work zone. To achieve this, we first define the rotation matrices, $R_x(\alpha)$ and $R_y(\beta)$, which rotate the vector positions by an angle of roll α and pitch β around the x-axis and y-axis. These matrices are $R_x(\alpha) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos(\alpha) & -sin(\alpha) \\ 0 & sin(\alpha) & cos(\alpha) \end{bmatrix}$ and $R_y(\beta) = \begin{bmatrix} cos(\beta) & 0 & sin(\beta) \\ 0 & 1 & 0 \\ -sin(\beta) & 0 & cos(\beta) \end{bmatrix}$. Notably, we assume the placement of AP satisfies the far-field conditions [33]. Consequently, the incident angle on each reflective element is given as ϕ_i and θ_i for a planar array surface. Each element position vector defined as $r_{mn} = [md_x, nd_y, 0]'$, $r_{mn} \in R^3$ and $m = [0, 1, ..., N-1]$, $n = [0, 1, ..., N-1]$, does not change in terms of its position w.r.t the RIS's local coordiantes. However, we must calculate the new incident angles, $\phi_i^{'}$ and $\theta_i^{'}$, after rotating the RIS with the predetermined roll and pitch angles. The derivations for the rotation with α over y-z plane are, ($\mathcal{Z} = cos\phi_i sin\theta_i$):

$$\theta_i^{'}(\alpha) = arccos(sin\phi_i sin\theta_i sin(\alpha) + cos\theta_i cos(\alpha))$$

$$\phi_i^{'}(\alpha) = arg(\mathcal{Z} + j(sin\phi_i sin\theta_i cos(\alpha) - cos\phi_i sin(\alpha))$$

The derivations for the rotation with β over x-z plane can be obtained in a similar manner. Hence, the required u-v plane coordinates for generating beams for different directions change to $u_i^{'} = kd_x cos\phi_i^{'} sin\theta_i^{'}$ and $v_i^{'} = kd_x sin\phi_i^{'} sin\theta_i^{'}$. Similar definitions apply to $u_r^{'}$ and $v_r^{'}$. These terms are used in Eq.1, 2, and 3 to generate the rest of the codebook. In the final step, the phase distribution of the RIS, Φ_{mn}, needs to be obtained from Eq.4 by taking the phasor form of $A_{mn}(\phi_i^{'}, \theta_i^{'})$ with the subtraction of $u_i^{'}$ and $v_i^{'}$ related terms from it. Fig. 4c shows %13 improvements in terms of gain after the rotation angle of α is applied for the same target, shown in Fig. 4b.

4.3 Hierarchy of RIS Codebook

4.3.1 2-bit Phase Quantization. Following the Sec. 4.1, to generate a beam toward (ϕ_r, θ_r), phase correction $\Phi_{mn} = \Phi_{mn}^s + (mu_i + nv_i)$ is required for the $(m, n)^{th}$ reflective element. These obtained phase correction values are continuous, and they must be mapped to the

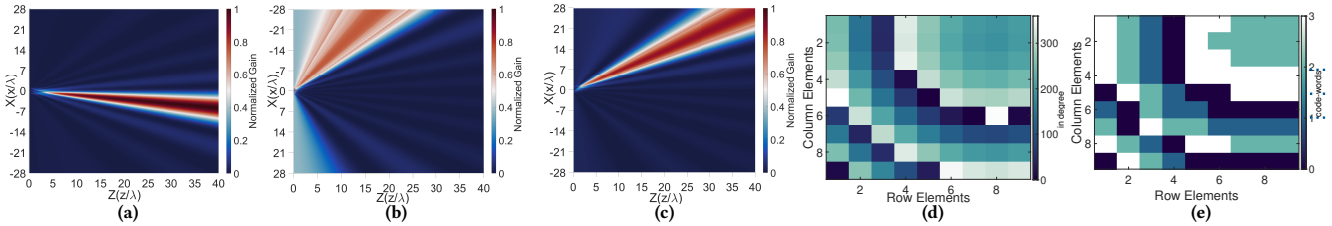

Figure 4: Normalized power gain distribution over x-z plane of RIS; while a) the target is close to its center, and b) the edge of its coverage area, c) after rotational angle is applied for the target at (b), also phase coding patterns at different ϕ_r and θ_r illustrating; d) the continuous phase shifts, e) corresponding quantized 2-bit phase values where $\phi_r = 60°$, $\theta_r = 20°$, $\alpha = 0°$ and $\beta = 0°$

predetermined discrete RIS configurations via configurable states of each reflective element. Although this limitation is necessary to decrease complexity, it introduces quantization errors. Yang et al. [34] demonstrate that the reflective gain loss is highly substantial when the quantization error exceeds $[-\frac{\pi}{4}, \frac{\pi}{4}]$, equivalent to at least 2-bit phase-shift quantization. Hence, we select the phase configurations from $\Omega_c = [\frac{jc\pi}{2^{b-1}}; c = 0, 1, ..., 2^b - 1]$ which are the nearest values to Φ_{mn}, keeping error $[-\frac{\pi}{4}, \frac{\pi}{4}]$, when $b = 2$,

$$\Phi^q_{mn} = \left\lfloor \frac{\Phi_{mn}}{\sigma} \right\rfloor \sigma, \quad \sigma = \frac{\pi}{2^{b-1}} \tag{7}$$

Fig. 4e shows the output of quantization corresponding to Fig.4d.

4.3.2 Codebook Design. For R rotational angles, the size of the codebook S_r, $r \in R$ is given by $S_r = i\lfloor \frac{\kappa_\phi \kappa_\theta}{BW_\phi BW_\theta} \rfloor$, $i = [1, 2, ..., I]$, considering a 3D beam scanning. Here, i is a correlation constant that increases correlation between neighboring beams in C_r, while κ denotes RIS angular coverage (e.g., $2\pi/3$ within $[-\pi/3, \pi/3]$ for our design). Each beam code in C_r consists of $N \times N$ codewords representing the required phase configurations of reflective elements. Implementing mechanical rotation introduces a controlled overhead (2.4 ms/angle). In our experiments, rotations occur infrequently—primarily when robot mobility substantially alters the optimal RIS orientation and demands broader angular coverage—thus balancing the overhead against the achieved coverage gains.

5 REMARKABLE Model

5.1 System Model

5.1.1 RIS-based Channel Model. In REMARKABLE, we consider the scenario in Fig. 1c, involving an AP with K antennas, a single-antenna MR, and an RIS with $A = N \times N$ passive reflective elements. Each element adjusts the amplitude and phase of the incident signal. Before data transmission, optimal beam selection through AP-RIS and RIS-MR channels is needed for array gain and high throughput, assuming no direct AP-MR link due to the cluttered factory floor. Let $w_a \in \mathbb{C}$ represent the effect of RIS element a on the reflected signal; $W \in C$ be the beamforming weight vector from the predefined codebook C. The received signal at MR through the RIS for the transmitted pilot signal x_t at the t^{th} time slot is:

$$y_t = \boldsymbol{h'W_t h''} x_t + n_t, \tag{8}$$

Let $\boldsymbol{W_t} = \text{diag}[w_1, ..., w_a, ..., w_A] \in \mathbb{C}^{A \times A}$, $\boldsymbol{h'} \in \mathbb{C}^{K \times A}$, and $\boldsymbol{h''} \in \mathbb{C}^{A \times 1}$ be complex-valued matrices and vectors, with elements representing channel coefficients between the AP and RIS element a and between a and MR, respectively. $n_t \in \mathbb{C}^{K \times 1}$ is the additive white Gaussian noise (AWGN), with $n_t \sim \mathcal{CN}(0, \sigma^2 \boldsymbol{I}_K)$.

5.1.2 Time-varying Channel Model. In our setup, the AP and RIS are fixed, while the robots are mobile. Thus, h' is quasi-static channel with coherence time T_S, and h'' is time-varying channel with T_M, where $T_S \gg T_M$. Here, we adopt the time-varying geometric channel model [35] with L multipath components between the RIS and MR. The multipath time-varying channel model at the p^{th} time instance in T_M is:

$$\boldsymbol{h''_p} = \sqrt{\frac{A}{L}} \sum_{l=1}^{L} \alpha_l e^{j2\pi v_l p T_M} \boldsymbol{a}(\phi_l, \theta_l) \tag{9}$$

where $\alpha_l \sim \mathcal{CN}(0, 1)$ is the complex channel gain of the l^{th} path, v_l is the Doppler shift, and $\boldsymbol{a}(\phi_l, \theta_l)$ is the reflection steering vector in the direction of ϕ_l and θ_l, respectively. We can derive the received signal at p^{th} time instance as $Y_p = \boldsymbol{h'} diag(\boldsymbol{h''_p}) \boldsymbol{\omega_p} \boldsymbol{X} + \boldsymbol{N_p}$, $\boldsymbol{X} \in \mathbb{C}^{1 \times T_M}$ is transmitted signal sequence and $\boldsymbol{\omega_p} \in \mathbb{C}^{A \times 1}$. We assume MRs cannot estimate the cascaded channel $\boldsymbol{H_p} = \boldsymbol{h'} diag(\boldsymbol{h''_p})$ from AP to MR over RIS; thereby, they only observe received signal power when the AP selects a beam from the RIS codebook C and transmits a pilot signal. if the pilot signal X is set to be 1, then the received signal power (RSP) can be expressed as $\mathcal{F}_p(\boldsymbol{\omega_p}) = |\sqrt{s}\boldsymbol{H_p}\boldsymbol{\omega_p} + \boldsymbol{N_p}|^2$.

Note that we convert the received signal power to the RSS by $10log(\mathcal{F}_p(\boldsymbol{\omega_p}))$ for our experiments. The target user is MR $\mathcal{M}$, but other MRs, denoted as $i \in \mathcal{I}$, should not experience interference from MR $\mathcal{M}$. Received signal power $\mathcal{F}_{p,i}(\boldsymbol{\omega_p})$ for i^{th} MR is also effected by time-varying channel $\boldsymbol{H_{p,i}}$.

5.2 Problem Formulation

We aim to control the beamforming weight vector $\boldsymbol{\omega_p}$ to find the optimal beam that achieves the largest expected RSP, $\mathbb{E}[\mathcal{F}_{p,m}(\boldsymbol{\omega_p})] = s|\boldsymbol{H_{p,m}}\boldsymbol{\omega_p}|^2$, for MR $\mathcal{M}$ while maintaining the expected maximum RSP, $\mathbb{E}[\max_{i \in \mathcal{I}}\{\mathcal{F}_{p,i}(\boldsymbol{\omega_p})\}] = \max_{i \in \mathcal{I}}\{s|\boldsymbol{H_{p,i}}\boldsymbol{\omega_p}|^2\}$, for MRs in $\mathcal{I}$ less than a threshold ρ. Given that $H_{p,m}$ and $H_{p,i}$ are unknown and environment-dependent, we formulate the beam selection as an online constrained stochastic optimization problem. Let T denote the time slots of equal duration for beam selection before transmitting data. In time slot $p \in T$, the AP selects a beamforming weight vector, $\boldsymbol{\omega_p}$, from a set of candidate beams (arms in the bandit), and observes RSP, $\mathcal{F}_p(\boldsymbol{\omega_p})$, from all the MRs. Here, we define $r_p(\boldsymbol{w_p}) = \mathcal{F}_{p,m}(\boldsymbol{\omega_p})$ as observed reward and $g_p(\boldsymbol{w_p}) = \max_{i \in \mathcal{I}}\{\mathcal{F}_{p,i}(\boldsymbol{\omega_p})\}$ as observed utility function, and both are time-varying. Sequentially selected beams and corresponding sequential rewards with utilities are presented as $\{\boldsymbol{w}_1, \boldsymbol{w}_2, ..., \boldsymbol{w}_p\}$ and $\{(r_1(\boldsymbol{w}_1), g_1(\boldsymbol{w}_1)), ..., (r_p(\boldsymbol{w}_p), g_p(\boldsymbol{w}_p))\}$, respectively. Our objective is to find a policy, $\pi \in \Pi$, that maximizes the expected cumulative reward, i.e., expected RSP, while satisfying

a constraint on the expected utility:

$$\max_{\boldsymbol{\pi}_p \in \Pi} \mathbb{E}[r_p(\boldsymbol{\omega}_p)] \quad \text{s.t. } \mathbb{E}[g_p(\boldsymbol{\omega}_p)] \leq \rho \tag{10}$$

Here, the AP selects the beam vector based on selection probability through policy π_p. Note that such a policy can depend on the historical information.

6 Adaptive Beam Selection with REMARKABLE

We adopt GP bandits due to their proven effectiveness in beam alignment tasks [12], particularly over traditional multi-armed bandits. GP bandits capture spatial correlations among beams and adapt to time-varying channels—key for mobile scenarios with dynamic channel states. With this motivation, first, we introduce Gaussian process (GP) kernel to represent the reward and the constraints. Then, we define a constrained GP-bandit problem to determine the beamforming vector (cf,(Eq.10)). We next describe the base algorithm to address the static robot case, followed by our novel modification addressing the non-stationarity.

6.1 Designing Solution with Gaussian Processes

6.1.1 Gaussian Processes. Our RIS codebook contains beamforming vectors for various rotational angles; different beamforming vectors $\boldsymbol{\omega}$, from different codebooks $C_r \in C$ can produce similar beams as the coverage sector of each codebook can partially overlap. This results in a high correlation between beamforming vectors (ω, ω'). Since our reward r_p and utility functions g_p are unknown and non-linear, we use Gaussian Processes and their Reproducing Kernel Hilbert Space (RKHS) to model this correlation, as inspired by [12]. *Note that MAB problem, where each beamforming vector is an orthogonal arm, cannot model such correlation.*

We define a $\mathcal{GP}$ over C as $\mathcal{GP}_C(\mu(\cdot), k(\cdot,\cdot))$ that is completely specified by a mean function μ and covariance function (kernel) $k : \forall \omega \in C$. We assume that the reward function without noise f_p and constrained utility function without noise z_p come from a $\mathcal{GP}$, and perturbed with Gaussian noise: $r_p = f_p(\omega) + n_p$, with $n_p \sim \mathcal{N}(0, \sigma^2)$ and $f_p(\cdot) \sim \mathcal{GP}(\mu_p(\cdot), k(\cdot,\cdot))$. Hence, if a beam vector ω is selected then $r_p \sim \mathcal{GP}(\mu_p(\omega), k(\omega,\cdot) + \sigma^2)$. Similar argument holds for z_p and $g_p = z_p(\omega) + n_p$, with kernel $\hat{k}$. We use $\mathcal{GP}(0_c, k(\cdot,\cdot))$ as a prior distribution over f_p. Given a set of sampling points $A_T = [\omega_1, ..., \omega_T]$ within C, observed rewards $r_p = [r_1, ..., r_T]^T$, the posterior distribution of f_p is $\mathcal{GP}(\mu_p(\cdot), \sigma_p^2(\cdot))$, where the mean and variance are:

$$\mu_p(\omega) = k_p(x)^T (K_p + \sigma^2 \mathcal{I})^{-1} r_{1:p} \tag{11}$$

$$\sigma_p^2(\omega) = k(\omega, \omega') - k_p(\omega)^T (K_p + \sigma^2 I)^{-1} k_p(\omega') \tag{12}$$

with $k_p(\omega) = [k(\omega_1, \omega), ..., k(\omega_T, \omega)]^T$, $K_p = [k(\omega, \omega')]_{\omega,\omega' \in A_p}$, and $\mathcal{I}$ is the identity matrix. Similarly, for the constraint z_p, we consider the posterior $\mathcal{GP}(\tilde{\mu}_p, \tilde{\sigma}_p^2)$ where $r_{1:p}$ is replaced by $g_{1:p}$, and the kernel k is replaced by $\tilde{k}$.

6.1.2 Reproducing Kernel Hilbert Space (RKHS). We assume that f_p belongs to RKHS $\mathbb{H}_k$. In particular, $\mathbb{H}_k$ is equipped with the kernel k such that $f_p(\omega) = \langle f_p(\cdot), k(\omega,\cdot)\rangle_{\mathbb{H}_k}$. Similarly, we assume that the constraint function z_p also belongs to $\mathbb{H}_{\tilde{k}}$, i.e., $z_p(\omega) = \langle z_p(\cdot), \tilde{k}(\omega,\cdot)\rangle_{\mathbb{H}_{\tilde{k}}}$. Some examples of kernel functions are square exponential, Matern etc., [36].

Throughout the rest of this paper, we assume that the functions are bounded, i.e., $||f_p(x)||_{\mathbb{H}_k} \leq F$, and $||g_p(x)||_{\mathbb{H}_{\tilde{k}}} \leq G$ for all p. Such assumptions are also common in practice in the wireless communication [12, 37].

6.1.3 Kernel Selection. We employ the Matern kernel to specify the RKHS in $\mathcal{GP}$ [36] as it shows the best performance:

$$k_{Matern}(\omega, \omega') = \frac{2^{1-v}}{\Gamma(v)} \left(\frac{s\sqrt{2v}}{l} \right) B_v \left(\frac{s\sqrt{2v}}{l} \right) \tag{13}$$

Here, $v > 0$ is the hyperparameter that controls the smoothness of the output, $s = d(\omega, \omega')$ encodes the similarities between two arms with the Euclidean distance, $B(v)$ and $\Gamma(v)$ are the modified Bessel function and the gamma function, respectively.

6.2 Base Algorithm

We now discuss the base algorithm (inspired from [12]) which we use to find the beamforming vector in the static case. This algorithm also forms the basis in the non-stationary case as well where the robots are mobile.

6.2.1 GP-Upper confidence bound (GP-UCB). Since we do not know f_p and z_p, rather we are learning. We only get feedback (noisy) corresponding to selecting the beamforming vector ω; we need to balance between the *exploration* and *exploitation* carefully. For the unconstrained GP-bandit, GP-UCB algorithm is proposed [36] where the idea is to select the points that have a higher mean estimate reward (exploitation) or have a higher posterior variance (exploration as it does not have enough information). Similarly, we maintain the upper confidence of the f_p (at time p) as the following term $\hat{f}_p(\omega) = \mu_{p-1}(\omega) + \beta_{p-1}\sigma_{p-1}(\omega)$. β_{p-1} is the weight factor: $F + \frac{1}{\sigma^2}\sqrt{2\log(1/\delta) + 2\gamma_{p-1}}$ where γ is the information gain[1] and the δ is the confidence parameter. Please see [12, 36] for details.

Primal-Dual: Unlike the unconstrained version [36], we consider a constrained optimization problem. Similar to [12], we consider the Lagrangian of Eq.10 as $\mathbb{E}[r_p(\omega_p)] - \phi(\mathbb{E}[g_p(\omega_p)] - \rho)$. We, then seek to solve for the Lagrangian:

$$\max_{\omega} \min_{\phi} \mathbb{E}[r_p(\omega)] - \phi(\mathbb{E}[g_p(\omega)] - \rho) \tag{14}$$

Hence, unlike the unconstrained version, we have to develop the UCB for the Lagrangian for a given dual variable ϕ. Since noise is zero-mean, $\mathbb{E}[r_p(\omega)] = \mathbb{E}[f_p(\omega)]$, $\mathbb{E}[g_p(\omega)] = \mathbb{E}[z_p(\omega)]$. We, thus, only need to find the lower confidence bound of z_p as we already obtained UCB for f_p, for which, we use $\hat{z}_p(\omega) = \tilde{\mu}_{p-1}(\omega) - \tilde{\beta}_{p-1}\tilde{\sigma}_{p-1}(x)$ where $\tilde{\beta}_{p-1} = G + \frac{1}{\sigma^2}\sqrt{2\log(1/\delta) + 2\hat{\gamma}_{p-1}}$. With probability $1-\delta$, $f_p(\omega) \leq \hat{f}_p(\omega)$, and $z_p(\omega) \geq \hat{z}_p(\omega)$ (from [12]) ensuring that if we use $\hat{f}_p$ and $\hat{z}_p$, it will be indeed UCB. For the static channel f_p and z_p are drawn from the GP with time invariant mean. We decide to choose the solution at time p as:

$$\omega_p = \arg\max_{\omega \in C} (\hat{f}_p(\omega) - \phi \hat{z_p}(\omega)) \tag{15}$$

After selecting a beamforming vector based on Eq.15, all the posterior mean and variance, $\mu_p(x)$, $\sigma_p(x)$ and $\hat{\mu}_p(x)$, $\hat{\sigma}_p(x)$ are updated through Eq.12 based on the received value. Finally, we update dual variable ϕ with the gradient descent in the dual domain $\phi = \max\{\phi + \eta(\hat{z}_p(\omega_p) - \rho), 0\}$, where η is the learning rate $\frac{\rho}{G\sqrt{T}}$.

[1] For matern kernel, $\gamma_T = T^{d(d+1)/(2v+d(d+1))}\log(T)$

Algorithm 1 GP kernel bandit for non-stationary

Input: T (the total time steps), η' (learning rate of dual variable) $=Y_{max}\sqrt{I/T}$, $\zeta = \min\left\{\sqrt{\frac{J\log(J)}{(e-1)(T/I)}}, 1\right\}$

Initialization: Number of arms $J = \lfloor \frac{1}{2}\log(T) \rfloor$, weight for each arm $j = 1, \ldots, J$, $w(j) = 1$, arm selection interval $I = \sqrt{T}$.

for $k = 1, \ldots, \lfloor T/I \rfloor$ **do**

  Initialization: Reward $R(k) = 0$, constraint value $G(k) = 0$.

  Set the probability for arms $j = 1, \ldots, J$ as

$$p(j) = (1-\zeta)\frac{w(j)}{\sum_{l=1}^{J} w(l)} + \frac{\zeta}{J} \tag{17}$$

  Choose arm j according to the probability $p(j)$.

  Set the restart interval $W = 2^j$.

  Run the base algorithm with the restart interval W.

  Collect the total reward $R(k)$ and the constraint value $G(k)$ across the interval I.

  for $j = 1, \ldots, J$ **do**

$$w(j) = \begin{cases} w(j)\exp\left(\frac{\zeta(R(k) - YG(k))}{(1+Y_{max})GIJp(j)}\right) \text{if } j \text{ is selected} \\ w(j) \text{otherwise .} \end{cases} \tag{18}$$

  Update $Y = \min(\max(Y + \eta'(G(k)/I - \rho), 0), Y_{max})$

We also clip the dual variable at ϕ_{max}. Please see Algorithm 1 in [12] for details of the Base algorithm.

6.2.2 Learning Metric. : For the static-case, we are interested in minimizing the regret and the violation:

$$R(T) = \sum_{p=1}^{T} \mathbb{E}[r_p^{\pi^*}] - \mathbb{E}[r_p^{\pi_p}], V(T) = \sum_{p=1}^{T} (\mathbb{E}[g_p^{\pi}] - \rho) \tag{16}$$

The regret measures the sub-optimality gap between the reward following the optimal policy π^*, and the reward following the policy π_p at time p. The violation measures the constraint violation at time p. Here, we seek to have sub-linear growth of $R(T)$ and $V(T)$, i.e., $R(T)/T \to 0$ as $T \to \infty$ as it will ensure that in most of the episodes, the policy is *feasible* and optimal. The following result signifies that the base algorithm indeed achieves the sub-linear regret and violation.

Proposition 1 ([12]). *With probability* $1-\delta$*, the base algorithm achieves* $R(T) \leq \tilde{O}(T^{1/2}), \quad V(T) \leq \tilde{O}(T^{1/2})$.

6.3 Addressing Non-Stationary Conditions

We now discuss how we modify the base algorithm to address the non-stationarity. Our key contribution is adapting the '*bandit over bandit*' approach inspired from [38] to the constrained GP-bandit. We first quantify non-stationarity, discuss existing metric-dependent methods, and finally introduce our novel approach that removes this dependency.

6.3.1 Time-varying Budget. Since our scenario is mobile, the channel conditions are time-varying. These time-varying channels affect reward/utility; thereby, f_p and z_p also vary over time. We assume that these variations are bounded by B_f and B_z. In particular, $B_f := \sum_{p=1}^{T-1} \max_x \left\|f_p - f_{p+1}\right\|_{\mathcal{H}_k}$ and $B_z := \sum_{p=1}^{T-1} \max_x \left\|z_p - z_{p+1}\right\|_{\mathcal{H}_k}$. The total combined variation budget would be $B = \max\{B_f, B_z\}$.

6.3.2 Restart Strategy. Similar to [12], to combat non-stationary conditions, we adopt the restart strategy, which resets the kernels and forgets previous observations, no longer useful for deciding the new beamforming vector as perhaps the environment has changed. Restart enables efficient adaptation by discarding outdated observations in non-stationary environments. Note that instead of restart, one can employ sliding window, or weight-based algorithm. The key algorithmic contribution is selecting the restart interval W which we explain in the following.

6.3.3 Unknown Variation Budget. Estimating the variation budgets B_f and B_z in real time is challenging and channel-dependent; thus, especially in mobile settings, the restart window W should be learned adaptively. We propose Algorithm 1, a bandit-over-bandit scheme that treats a candidate set $J = \{2^j\}_{j=1}^{\lfloor 0.5\log_2 T \rfloor}$ of restart intervals as outer-loop arms. We partition the horizon T into epochs of length $I = \sqrt{T}$. At epoch k, EXP3 [39] selects an arm j from weights $w_j(k)$ (updated via Eq. (18)), yielding $W = 2^j$; the base algorithm (Sec. 6.2) then runs for I steps with restarts every W. We aggregate reward $R(k)$ and interference $G(k)$ over the epoch, update arm probabilities using Eq. (17), and update the dual variable via $Y \leftarrow \Pi_{[0,Y_{\max}]}(Y + \eta'(G(k)/I - \rho))$. Arms that induce low reward and/or higher constraint violation (see, (Eq.18)) receive lower weight and are selected less often in subsequent epochs.

6.4 Performance Metrics

•**Regret and Violation:** Since the optimal policy can change over time in a non-stationary environment, we evaluate our algorithm using dynamic regret Dy − R(T) and constraint violation $V(T)$. We define the dynamic regret as:

$$\text{DynR}(T) = \mathbb{E}\left[\sum_{p=1}^{T}\left(r_p^{\pi_p^*}(\boldsymbol{\omega}_p) - r_p^{\pi_p}(\boldsymbol{\omega}_p)\right)\right] \tag{19}$$

The constraint violation metric $V(T)$ is still given by Eq.16. Compared to regret, bounding dynamic-regret is fundamentally more challenging. Nevertheless, we obtain sub-linear dynamic-regret and violation bound.

Theorem 1. *Algorithm 1 achieves with probability* $1-\delta$*,*

$$\text{DynR}(T) \leq \tilde{O}(B^{1/4}T^{3/4}), \quad V(T) \leq \tilde{O}(B^{1/4}T^{3/4})$$

The proof is in our technical report [40] owing to space constraint.

[12] achieves dynamic regret and violation bound of $\tilde{O}(BT^{3/4})$ (Theorem 1 there) where one can choose the worst estimated budget if the budget is unknown. In contrast, our bound achieves $\tilde{O}(B^{1/4}T^{3/4})$. Hence, we improve the dependency of the budget. Note that [12] indicates that even if $B = o(T^{1/4})$, one can have linear regret and violation. In contrast, our result still achieves sub-linear regret and sub-liner violation bound as long as the time-varying budget grows sub-linearly. [12] has also achieved $\tilde{O}(B^{1/4}T^{3/4})$ dynamic regret and violation, however, it requires the knowledge of the budget B (Corollary 2 there). Our result achieves the same bound without the information of B.

For an online setting, it is norm to assume that T is known. If T is unknown, one can use the doubling 'trick' [41], which scales the regret and violation bound by $\log(T)$. In particular, one can choose $T = 2^0, 2^1, 2^2, \ldots$, and run the algorithm until reaching the T.

•**Selection Accuracy:** We also evaluate the selection accuracy by defining a prediction metric as $P = \frac{1}{T}\sum_{n=1}^{T} \frac{\log(1+RSS^t(n))}{\log(1+RSS^{t^*})}$ for total

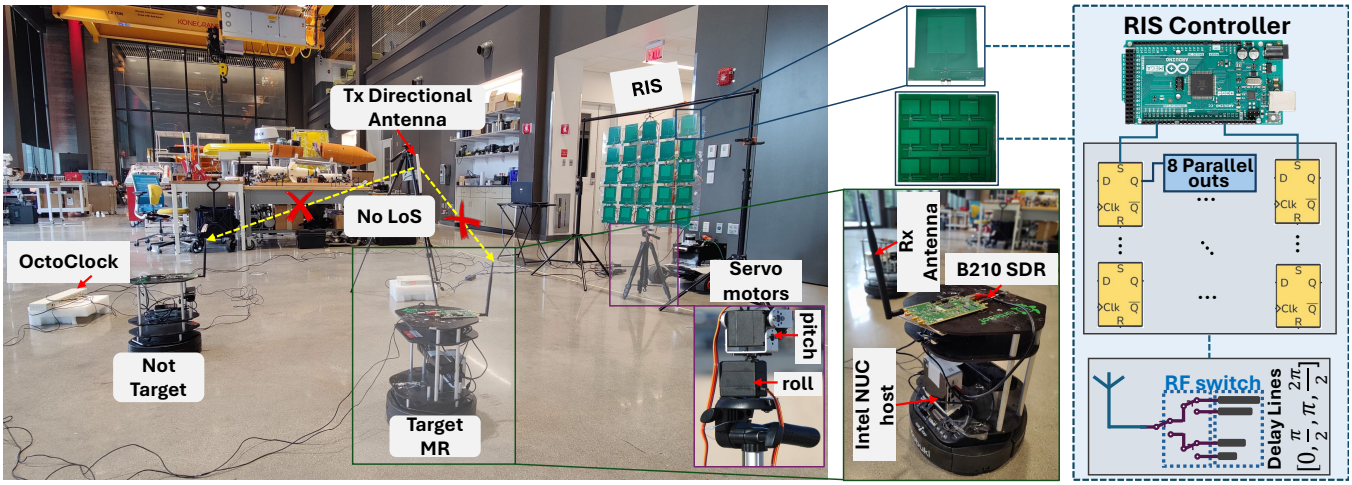

Figure 5: Experimental testbed with MRs in a lab emulating a factory floor (left). Closeup view of two different reflective elements configured through the control unit (middle). The schematic of an individual reflective element with multiple delay lines (right).

given time T, which measures the throughput ratio between the policy chosen and the optimal policy.

7 System Implementation of REMARKABLE

7.1 Experimental Testbed

7.1.1 REMARKABLE Setup. Fig. 5 shows two Thurtlebot robots functioning as MRs. Each includes an Intel NUC running the DSP application in a Linux-based system and a B210 SDR operating in the ISM band, equipped with omni-directional antennas. An x310 radio with a directional antenna serving as an AP, facing the RIS, with a distance between AP-RIS meeting far-field conditions ($2D^2/\lambda$, D represents surface size). All radios are connected to OctoClock-G CDA-2990 for frequency and time synchronization. There is a feedback (backhaul) channel between MRs and AP for sending current observations (RSS) to the AP. The AP's host machine coordinates signal transmission and issues directives to the RIS control unit, including the selected codeword of $\boldsymbol{\omega}$ for RIS and rotational angles (α and β) for servo motors attached to the surface. Note that we tested only *roll* rotation, but both *pitch* and *roll* rotations are feasible for a ceiling-mounted RIS.

7.1.2 RIS Implementation. The loss-based transmission line concept for phase shifting, previously implemented in [9, 42], was modified to realize our surface. Our fabricated RISs, shown in Fig. 5, consist of switchable patch-type antennas designed with the inset feeding technique [33] operating at the ISM bands of 900MHz and 2.4GHz. The element size can be significantly reduced as operating at higher frequencies. We deploy two RISs: one with 25 reflective elements in a 5×5 layout and another with 81 elements in a 9×9 layout, with antenna spacing of $\lambda/2$ to minimize mutual coupling and grating lobes. Each reflective element connects to four lossless transmission lines via a single RF switch, allowing phase shifts of 0, $\pi/2$, π, and $3\pi/2$ with 2-bit quantization. By selecting the length of the transmission line, we can alter the impedance of each reflective element by changing its reflective coefficient, Γ_{mn}, thereby, introducing a phase shift to the reflected signal. In control unit, each element attached to MASWSS0204 RF switches is controlled by an arduino Mega2560 μ-controller who orchestrates the configurations of elements parallel via SN74HC595 shift registers.

8 Performance Evaluation of REMARKABLE

8.1 Validation of RIS Codebook

We first verify the performance of the RIS codebook by measuring the RSS at various locations with equal distances from RIS. We set up a receiver antenna on a tripod at varying heights to cover angular space in the elevation plane. At each location, we scan the entire angular range by switching between beams from the codebook, containing codewords for angular directions in the range of $[-\pi/2, \pi/2]$ for ϕ and $[-\pi/4, \pi/4]$ for θ, with a resolution of 1°. An example of the collected data is shown in Fig. 7, comparing two beam patterns in two different directions. We observe that the 9×9 surface achieves an 11dB gain as shown in Fig. 7a, while the gain drops to 7dB when the beam direction approaches the edge of coverage. As a result, we determine the coverage area of RIS to be $[-\pi/3, \pi/3]$. Moreover, we observe a deviation of 2.3% from the desired direction due to quantization, still within the beamwidth of the pattern, around $15^\circ - 20^\circ$. However, this error increases for the direction in Fig. 7b. Lastly, we can enhance the gain for the corresponding direction at the edge of the coverage by adjusting the orientation of RIS as shown in Fig. 7b.

8.2 Stationary Scenario

Here, we employ the testbed illustrated in Fig. 5 (see Sec. 7.1) to assess the performance of the proposed beam selection algorithm while the robots are stationary at different locations. One robot is designated as the target, while the other is constrained by interference threshold, ρ. ρ from Eq.10 is determined through multiple measurements in the environment. Two different sizes of codebooks are employed in this experiment: the first one includes two rotation angles with $S_r = 16$ and a total of $S = 32$ for codebook $CB1 = \{C_1, C_2\}$, and the other has $S_r = 32$ with a total of $S = 64$ for codebook $CB2 = \{C_3, C_4\}$ (see Sec. 4.3). For each codebook, $CB1$ and $CB2$, we interchange the roles of robots, such as swapping the second one as the target, and repeat the experiment 200 times. We employ the base algorithm, *GP-UCB-C*, (see Sec. 6.2) and plot the regret and reward, as in Fig. 6a-6b. Our findings demonstrate that the algorithm finds the optimum policy for both codebook sizes and rapidly converges to the best policy without violating $V(T)$.

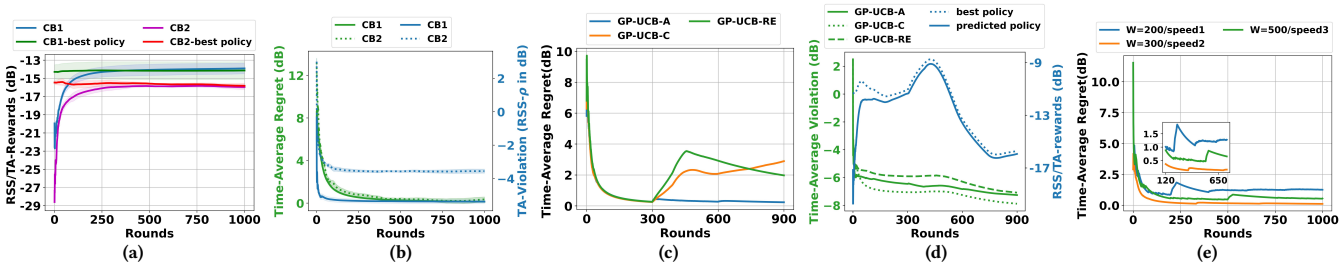

Figure 6: Performance evaluation of proposed approaches with different experiment settings: at stationary conditions a) Time-average (TA) rewards with variation, from multiple measurements, b)TA regret and TA violation performances; at non-stationary conditions c) Comparison of three different methods (see Sec. 6.3), and d)TA violation and TA rewards, e) TA regret under different mobility settings.

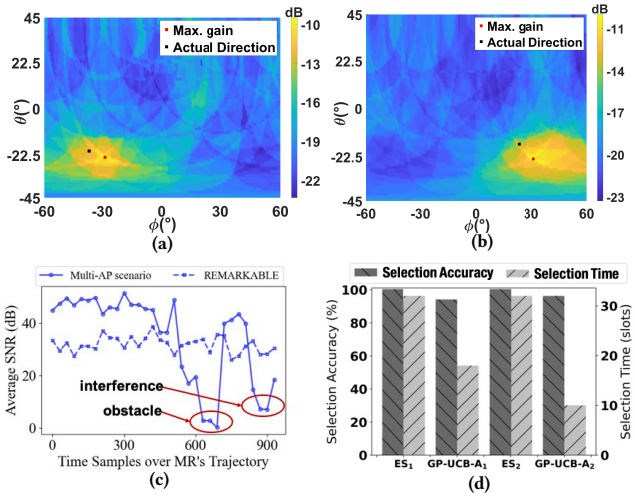

Figure 7: Validation of codebook design; a) measured beam pattern at (-35°,-19°), b) after changing the orientation of RIS (MR was outside of the coverage at (65°,-13°)); c) Comparison of REMARKABLE's SNR performance against multi-AP scneraio; and d) Comparing the performance of REMARKABLE and exhaustive search (ES) w.r.t selection accuracy and beam selection time.

As this is a real-time experiment, obtaining the best policy while running the algorithm is not feasible. Consequently, we conduct two consecutive epochs: one for beam selection using the proposed approach and another utilizing an exhaustive search technique to find the best policy.

8.3 Non-stationary Scenario

In this study, the MRs travel along a pre-determined path and send their observations (RSS) to the AP. We first conduct the experiment with our base algorithm (*GP-UCB-C*) (without restart strategy). The results (Fig. 6c-6d) show that *GP-UCB-C* has a significant increase in TA-regret (DynR$(T)/T$) due to changes in the channel, specifically starting around 300, even if with non-violated constraint. We then apply the restart strategy *GP-UCB-RE* proposed in [12] with an upper estimate on the budget. While this restart strategy fails but still beats the baseline as it ultimately restarts and the TA-regret starts decreasing. Finally, our proposed approach Algorithm 1 (GP-UCB-A) has the best performance (Fig. 6c-6d) compared to other two baselines as it can quickly adapt to the change in the environment. In particular, our proposed approach has the lowest TA-regret even when the environment changes while also finding feasible solution. The last experiment corresponds to a time-varying channel with different channel coherence times as we adjust the speed of the MR from min to the allowed max speed as in $[0.15m/s, 0.3m/s, 0.45m/s]$. Fig. 6e presents that our approach Alg.1 adaptively selects the best optimum beams over time-varying channel quickly which is evident from its TA-regret performance.

8.4 Comparison with Classical Methods

Here, we first compare REMARKABLE's SNR performance with the classical multi-AP deployment, and then evaluate REMARKABLE against classical methods in terms of the beam selection time, and selection accuracy from Sec. 6.4. For benchmarking, REMARKABLE is compared with exhaustive search (ES)—the widely adopted baseline in mmWave protocols (e.g., IEEE 802.11ad/ay)—to highlight the efficiency of our approach without resorting to full codebook scans, rather than against complex learning methods such as reinforcement learning, which demands extensive training. As shown in Fig. 7c, the MR suffers an SNR drop when encountering obstacles or interference, whereas REMARKABLE maintains reliable links with consistent SNR. Despite RIS-induced cascaded path loss, REMARKABLE consistently achieves high SNR levels, underscoring the robustness of the design. Furthermore, we define a term of *time slot* as the end-to-end latency of each selected beam, comprising: (i) the beam searching latency of the employed algorithm, including both selection and real-time execution; (ii) feedback latency between the MR and AP, where the MR reports the observed RSS to the AP; and (iii) control latency between the AP and the RIS, where the AP transmits the selected codewords to the RIS. The control latency for transmitting a single codeword to the RIS is 208 µs, the average feedback latency over the wireless backhaul is 8.1 ms, and the beam searching latency is approximately 23.8 ms for codebook CB1 (higher for ES). When using ES, 32 time slots are needed to examine all beams from $CB1$ and select the one with the highest RSS, resulting in a guaranteed overhead delay. In contrast, REMARKABLE averages 18 time slots (based on multiple experiments) to find the best beam with our adaptive algorithm *GP-UCB-*A_1. After first initiated restart, *GP-UCB-*A_2 achieves adaptation in only 10 slots, whereas ES_2 must re-scan. Finally, Fig. 7d shows that REMARKABLE achieves a 46.8% improvement in beam selection time while maintaining 94.2% accuracy. Note that this study primarily focuses on low-mobility scenarios, while high-mobility cases are left for future work with further latency optimization.

Time Complexity Analysis: The codebook design (see Sec 4) is performed offline, which significantly supports scalability of the hardware. As we use gradient-descent, its time complexity is $O(SIN^2 \log N)$, where S is the codebook size, N^2 the number of RIS elements, and I the number of gradient descent iterations. At

runtime, the codebook size affects the online complexity of the bandit algorithm. Specifically, computing the mean and variance requires matrix inversion with worst-case complexity $O(W^2)$, where W is the restart interval. Evaluating GP-UCB over S codebook cardinality results in overall complexity $O(SW^2)$. Since $W \leq \sqrt{T}$, the runtime remains linear in both S and T. Through empirical RIS measurements, the runtime is 25% fraction of the total end-to-end latency reported in Sec 8.4. This implies that the practical runtime overhead is suitable for real-world environments. Designing faster algorithms and potentially hardware-accelerated solutions remains an important future direction, especially in high-mobility settings.

9 Conclusions and Future Work

We demonstrate optimal beam selection for robot connectivity under interference-constrained, time-varying channels. Our RIS codebook enables selection without channel estimation, and a reconfigurable mechanism extends coverage. Using an adaptive bandit-over-bandit restart strategy, REMARKABLE safely learns optimal beams in dynamic conditions, achieving 46.8% faster selection and 94.2% accuracy—outperforming classical methods. Characterizing the lower bound on the dynamic regret and violation is an important future work. Extending this framework to multiple robots and multiple RIS also constitutes an important future research direction.

Acknowledgments

This work has been supported in part by the U.S. National Science Foundation under the grants: NSF AI Institute (AIEDGE) CNS-2112471, and CNS-2312836, and the Army Research Laboratory under Cooperative Agreement Number W911NF- 23-2-0225. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S.Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

[1] C. Bai, P. Dallasega, G. Orzes, and J. Sarkis, "Industry 4.0 technologies assessment: A sustainability perspective," *International journal of production economics*, vol. 229, p. 107776, 2020.
[2] Y. Liu, X. Liu, X. Gao, X. Mu, X. Zhou, O. A. Dobre, and H. V. Poor, "Robotic communications for 5g and beyond: Challenges and research opportunities," *IEEE Communications Magazine*, vol. 59, no. 10, pp. 92–98, 2021.
[3] L. Qiao, Y. Li, D. Chen, S. Serikawa, M. Guizani, and Z. Lv, "A survey on 5g/6g, ai, and robotics," *Computers and Electrical Engineering*, vol. 95, p. 107372, 2021.
[4] W. Hong, Z. H. Jiang, C. Yu, D. Hou, H. Wang, C. Guo, Y. Hu, L. Kuai, Y. Yu, Z. Jiang *et al.*, "The role of millimeter-wave technologies in 5g/6g wireless communications," *IEEE Journal of Microwaves*, vol. 1, no. 1, pp. 101–122, 2021.
[5] V. Arun and H. Balakrishnan, "{RFocus}: Beamforming using thousands of passive antennas," in *17th USENIX symposium on NSDI*, 2020, pp. 1047–1061.
[6] M. Nemati, J. Park, and J. Choi, "Ris-assisted coverage enhancement in millimeter-wave cellular networks," *IEEE Access*, vol. 8, pp. 188 171–188 185, 2020.
[7] K. Li, Y. Naderi, U. Muncuk, and K. R. Chowdhury, "Isurface: Self-powered reconfigurable intelligent surfaces with wireless power transfer," *IEEE Communications Magazine*, vol. 59, no. 11, pp. 109–115, 2021.
[8] L. Zhang, X. Q. Chen, S. Liu, Q. Zhang, J. Zhao, J. Y. Dai, G. D. Bai, X. Wan, Q. Cheng, G. Castaldi *et al.*, "Space-time-coding digital metasurfaces," *Nature communications*, vol. 9, no. 1, p. 4334, 2018.
[9] M. Dunna, C. Zhang, D. Sievenpiper, and D. Bharadia, "Scattermimo: Enabling virtual mimo with smart surfaces," in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, 2020, pp. 1–14.
[10] Y. Ren and V. Friderikos, "Interference aware path planning for mobile robots in mmwave multi cell networks," in *2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall)*. IEEE, 2022, pp. 1–6.
[11] J. Wang, W. Tang, S. Jin, C.-K. Wen, X. Li, and X. Hou, "Hierarchical codebook-based beam training for ris-assisted mmwave communication systems," *IEEE Transactions on Communications*, vol. 71, no. 6, pp. 3650–3662, 2023.
[12] Y. Deng, X. Zhou, A. Ghosh, A. Gupta, and N. B. Shroff, "Interference constrained beam alignment for time-varying channels via kernelized bandits," WiOpt, 2022.
[13] C. Feng, X. Li, Y. Zhang, X. Wang, L. Chang, F. Wang, X. Zhang, and X. Chen, "Rflens: metasurface-enabled beamforming for iot communication and sensing," in *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, 2021, pp. 587–600.
[14] Z. Li, Y. Xie, L. Shangguan, R. I. Zelaya, J. Gummeson, W. Hu, and K. Jamieson, "Towards programming the radio environment with large arrays of inexpensive antennas," in *16th USENIX Symposium on NSDI*, 2019, pp. 285–300.
[15] [Online]. Available: https://github.com/dDreamCatcher/RIS_Bandit_impl
[16] X. Li, C. Feng, F. Song, C. Jiang, Y. Zhang, K. Li, X. Zhang, and X. Chen, "Protego: securing wireless communication via programmable metasurface," in *MobiCOM*, 2022, pp. 55–68.
[17] O. Abari, D. Bharadia, A. Duffield, and D. Katabi, "Enabling {high-quality} untethered virtual reality," in *14th USENIX Symposium on NSDI 17*, 2017.
[18] D. Xie, X. Wang, and A. Tang, "Metasight: localizing blocked rfid objects by modulating nlos signals via metasurfaces," in *MobiSys*, 2022, pp. 504–516.
[19] X. Zhai, G. Han, Y. Cai, and L. Hanzo, "Beamforming design based on two-stage stochastic optimization for ris-assisted over-the-air computation systems," *IEEE Internet of Things Journal*, vol. 9, no. 7, pp. 5474–5488, 2021.
[20] W. Fang, Y. Jiang, Y. Shi, Y. Zhou, W. Chen, and K. B. Letaief, "Over-the-air computation via reconfigurable intelligent surface," *IEEE transactions on communications*, vol. 69, no. 12, pp. 8612–8626, 2021.
[21] Q. An, Y. Zhou, and Y. Shi, "Robust design for reconfigurable intelligent surface assisted over-the-air computation," in *2021 IEEE Wireless Communications and Networking Conference (WCNC)*, 2021, pp. 1–6.
[22] X. Tan, Z. Sun, D. Koutsonikolas, and J. M. Jornet, "Enabling indoor mobile millimeter-wave networks based on smart reflect-arrays," in *IEEE INFOCOM*. IEEE, 2018, pp. 270–278.
[23] K. Qian, L. Yao, X. Zhang, and T. N. Ng, "Millimirror: 3d printed reflecting surface for millimeter-wave coverage expansion," in *MobiCOM*, 2022, pp. 15–28.
[24] Y. Cheng, W. Peng, C. Huang, G. C. Alexandropoulos, C. Yuen, and M. Debbah, "Ris-aided wireless communications: Extra degrees of freedom via rotation and location optimization," *IEEE Transactions on Wireless Communications*, vol. 21, no. 8, pp. 6656–6671, 2022.
[25] G. Hu, Q. Wu, D. Xu, K. Xu, J. Si, Y. Cai, and N. Al-Dhahir, "Intelligent reflecting surface-aided wireless communication with movable elements," *IEEE Wireless Communications Letters*, vol. 13, no. 4, pp. 1173–1177, 2024.
[26] Y. Wei, Z. Zhong, and V. Y. Tan, "Fast beam alignment via pure exploration in multi-armed bandits," *IEEE Transactions on Wireless Communications*, vol. 22, no. 5, pp. 3264–3279, 2022.
[27] W. Wu, N. Cheng, N. Zhang, P. Yang, W. Zhuang, and X. Shen, "Fast mmwave beam alignment via correlated bandit learning," *IEEE Transactions on Wireless Communications*, vol. 18, no. 12, pp. 5894–5908, 2019.
[28] J. Zhang, Y. Huang, Y. Zhou, and X. You, "Beam alignment and tracking for millimeter wave communications via bandit learning," *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5519–5533, 2020.
[29] M. Hashemi, A. Sabharwal, C. E. Koksal, and N. B. Shroff, "Efficient beam alignment in millimeter wave systems using contextual bandits," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE, 2018, pp. 2393–2401.
[30] S. J. Orfanidis, *Electromagnetic waves and antennas*, ser. International series of monographs on physics. Rutgers University, 2016.
[31] M. Khalaj-Amirhosseini, "Phase-only synthesis of antenna arrays using nonuniform phased sampling method." *Iranian Journal of Electrical & Electronic Engineering*, vol. 17, no. 2, 2021.
[32] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.
[33] C. A. Balanis, *Antenna theory: analysis and design*. John wiley & sons, 2016.
[34] J. Yang, Y. Chen, Y. Cui, Q. Wu, J. Dou, and Y. Wang, "How practical phase-shift errors affect beamforming of reconfigurable intelligent surface?" *IEEE TCOM*, 2023.
[35] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave mimo systems," *IEEE journal of selected topics in signal processing*, vol. 10, no. 3, pp. 436–453, 2016.
[36] N. Srinivas, A. Krause, S. M. Kakade, and M. Seeger, "Gaussian process optimization in the bandit setting: No regret and experimental design," *arXiv preprint arXiv:0912.3995*, 2009.
[37] A. M. Girgis, J. Park, M. Bennis, and M. Debbah, "Predictive control and communication co-design via two-way gaussian process regression and aoi-aware scheduling," *IEEE Transactions on Communications*, 2021.
[38] H. Wei, A. Ghosh, N. Shroff, L. Ying, and X. Zhou, "Provably efficient model-free algorithms for non-stationary cmdps," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2023, pp. 6527–6570.
[39] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire, "The nonstochastic multiarmed bandit problem," *SIAM journal on computing*, vol. 32, pp. 48–77, 2002.
[40] [Online]. Available: https://github.com/arnobghosh-njit/Kernel_banidt.git
[41] L. Besson and E. Kaufmann, "What doubling tricks can and can't do for multi-armed bandits," *arXiv preprint arXiv:1803.06971*, 2018.
[42] S. G. Sanchez, K. Alemdar, V. Chaudhary, and K. Chowdhury, "Ris-star: Ris-based spatio-temporal channel hardening for single-antenna receivers," in *IEEE INFOCOM*. IEEE, 2023.