ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)
https://doi.org/10.17559/TV-20230908000926
Original scientific paper

Enhancing Spectrum Sharing Efficiency in Large-Scale MIMO Systems over Integration of Cognitive Radio and Reinforcement Learning

Madhan KRISHNAMURTHI, Vinoth Kumar KALIMUTHU*

Abstract: Cognitive Radio Networks (CRNs) aim to optimize the limited frequency spectrum by enabling spectrum sharing among different networks and making use of unoccupied frequency bands. The combination of massive multiple-input multiple-output (mMIMO) and CRNs has the potential to greatly improve the efficiency of upcoming wireless communication networks. In our research, we introduce an innovative approach to spectrum sharing in cognitive radio, utilizing 3D spatial data acquisition and Deep Learning for learning and decision-making. We incorporate mMIMO structures into cognitive radio base stations (CRBS) to extract angular information from user equipment (UE) and estimate Direction of Arrival (DoA) using Iterative Hard Thresholding (IHT). Our method involves deploying two base stations per cell for comprehensive 3D spatial spectrum coverage during spectrum prediction. We employ advanced Deep Learning techniques for spectrum sensing instead of reinforcement learning, enhancing CRN performance. Our approach includes a two-fold spectrum scheduling strategy, one focused on maximizing CR coverage and the other on optimizing transmission rates in CRN mMIMO scenarios. By fine-tuning Deep Learning parameters, our model achieves significantly higher Average Aggregate Sum Rate (AASR) compared to previous CRN spectrum sharing methods, without relying on reinforcement learning for spectrum sensing. Our research underscores the effectiveness of integrating Deep Learning into cognitive radio networks, offering the potential for enhanced spectrum utilization and network performance. Additionally, we address energy efficiency using the Nakagami fading channel model and evaluate key metrics, including channel occupancy and energy efficiency, through experimental analysis.
Keywords: cognitive radio networks; direction of arrival; iterative hard thresholding; multiple-input multiple-output; user equipment

1 INTRODUCTION

Cognitive Radio (CR) has emerged as a promising solution to efficiently utilize the limited radio spectrum resources, particularly in the face of the rapid proliferation of wireless applications and devices. Initially rooted in the concept of Software Defined Radio (SDR), CR was conceived to expand the use of available spectrum [1]. Regulatory bodies like the Federal Communications Commission (FCC) are tasked with overseeing spectrum administration due to the growing demand driven by emerging wireless technologies. To address this demand, the FCC has explored dynamic spectrum access, allowing unlicensed users to detect and use unoccupied spectrum portions, optimizing spectrum utilization. Future wireless networks, termed Cognitive Radio Linkages (CRNs), align with Haykin's CR concept and integrate seamlessly into modern wireless systems by dynamically adapting attributes like uplink and downlink frequencies in real-time to distribute spectrum efficiently and minimize disruptions to users [2-3]. This adaptability draws insights from various aspects of the radio environment, such as radio frequency (RF) transmissions and device-level interferences. Effective and equitable spectrum distribution across geographically dispersed devices requires a cooperative approach, leading to ongoing research on collaborative strategies for spectrum sharing within CR networks. Similar to CR nodes, a Compound Representative Structure (MAS) comprises numerous self-governing agents that dynamically respond to user needs, working individually or collectively to achieve goals. This concept optimizes radio spectrum utilization in the face of expanding wireless applications, enhancing efficiency through real-time parameter adjustments and cooperative interactions [4-7]. CR systems introduce a dynamic and intelligent approach to wireless communication by efficiently using the radio spectrum [8]. These systems involve two main user types: Primary Users (PUs) and Secondary Users (SUs). PUs are licensed spectrum users with exclusive rights to specific frequency bands, while SUs are unlicensed, cognitive users. In CR systems, Orthogonal Frequency Division Multiplexing (OFDM) is a crucial modulation technique employed to achieve spectral efficiency. Primary Users (PUs) are incumbents in the spectrum, representing authorized radio and communication services. Their primary services must be protected from interference to ensure reliable communication. Secondary Users (SUs) are cognitive or opportunistic users without primary spectrum allocations, typically unlicensed. SUs opportunistically access spectrum bands not used by PUs, with cognitive capabilities to sense, detect, and adapt their transmissions to prevent interference with PUs. Orthogonal Frequency Division Multiplexing (OFDM) is essential in CR systems due to its ability to manage frequency-selective fading, reduce interference, and support flexible spectrum allocation. OFDM allows Secondary Users (SUs) to access underutilized frequency bands while avoiding disruption to Primary Users (PUs). The combination of OFDM with advanced signal processing techniques like beamforming and Multiple-Input Multiple-Output (MIMO) enhances CR systems' communication capabilities [9-10]. MIMO-OFDM configurations, utilizing multiple transmitting and receiving antennas, facilitate simultaneous signal transmission through spatial multiplexing, introduce diversity benefits, and reduce interference, increasing the overall capacity and reliability of CR networks. OFDM's robustness, adaptability, and synergy with advanced technologies make it a crucial enabler for efficient and intelligent Cognitive Radio networks. Adaptive Resource Allocation (RA) has been a focal point in the study of OFDM systems for over a decade, particularly in the context of OFDM-based Cognitive Radio (CR) systems. In scenarios with a single Secondary User (SU), RA simplifies to power distribution. The integration of MIMO and OFDM in CR networks with multiple transmitting antennas optimizes resource utilization and communication performance in dynamic spectrum environments [11-15].

2 RELATED WORKS

In the dynamic world of wireless communication, effectively managing spectrum resources is a persistent issue due to increasing device numbers and connectivity demands. Large-Scale Multiple-Input Multiple-Output (MIMO) systems, when combined with Cognitive Radio (CR) and Reinforcement Learning (RL), offer promising solutions to this challenge. This essay conducts a thorough literature review on the synergy between CR, RL, and Large-Scale MIMO for spectrum sharing optimization. The escalating demand for wireless communication has resulted in spectrum scarcity. Large-Scale MIMO systems, equipped with multiple antennas at both ends, promise increased capacity and enhanced spectral efficiency [24]. Nevertheless, the efficient use of available spectrum remains a complex issue. Cognitive Radio, which enables opportunistic spectrum access, and Reinforcement Learning, a potent mechanism for dynamic decision-making, offer promising solutions to this challenge. Cognitive Radio (CR) empowers secondary users (SUs) to opportunistically access underutilized spectrum bands without causing interference to primary users. Spectrum sensing, a fundamental aspect of CR, enables SUs to detect available channels. The study conducted by Mitola et al. [25] emphasized the critical role of spectrum sensing in achieving dynamic spectrum access. Spectrum sensing involves CR devices monitoring and detecting primary user transmissions within a frequency band. Reinforcement Learning (RL) has gained prominence for enabling autonomous decision-making in dynamic and uncertain environments. In spectrum sharing, RL algorithms facilitate the learning of optimal strategies for spectrum access. A seminal work by Q. Zhao and B. M. Sadler [26] explores RL's application in CR systems, showcasing its potential for adaptively exploiting available spectrum opportunities.

The integration of Cognitive Radio and Reinforcement Learning holds promise for enhancing spectrum sharing efficiency. A study by C. Sengul et al. [27] presents a framework that combines CR and RL to optimize channel selection in Large-Scale MIMO systems, achieving improved spectrum utilization and interference mitigation. This research highlights the synergy between CR and RL in addressing spectrum sharing challenges.

Large-Scale MIMO (Massive Multiple-Input Multiple-Output) systems provide improved capacity and signal quality through spatial multiplexing and diversity. Research by M. Chen et al. [28] introduces an integrated optimization strategy for beamforming and power allocation in Large-Scale MIMO networks, incorporating Cognitive Radio (CR) and Reinforcement Learning (RL) techniques. This approach efficiently allocates resources, leading to enhanced capacity and spectral efficiency. Numerous studies have explored the fusion of CR and RL in Large-Scale MIMO systems, showcasing their potential to optimize spectrum sharing, channel selection, and resource allocation. This integration presents a promising solution to address spectrum efficiency challenges in the face of increasing demand for wireless communication resources.

3 THE PROPOSED MODEL

The envisioned system is a Cognitive Radio Network (CRN) comprising a Primary Network (PN) with massive Compound Input and Output (mMIMO) technology and Cognitive Base Stations (CBS). The network topology includes mMIMO antennas in CBS and single antennas in Cognitive Radios (CRs), with half-duplex communication constraints. Angular information is acquired using Iterative Hard Thresholding (IHT), and a novel deep reinforcement learning (RL) architecture aids in spectrum sharing. Energy efficiency is considered through the Nakagami fading channel model, ultimately optimizing the CRN's efficiency and energy utilization. The Fig. 1 and Fig. 2 show the proposed systems

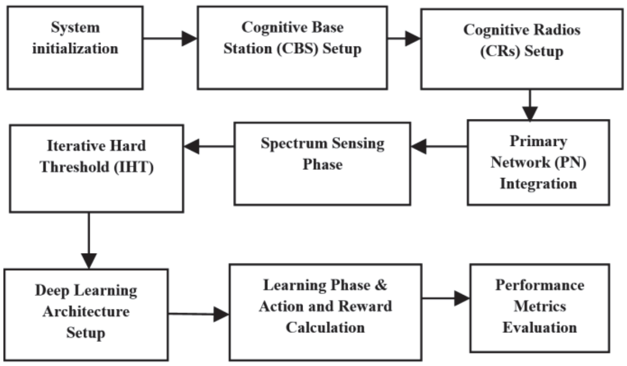

Figure 1 Block diagram of couple CBS based 3D space spectrum sharing

In a 3D cognitive radio network, there are two Cognitive Base Stations (CBS) and N randomly positioned Cognitive Radio Users (CR Users). CBS employ a uniform rectangular array (URA) of TXR transmit antennas and RXR receive antennas, while CR Users use a single antenna setup with a uniform linear array (ULA) configuration. Additionally, there are intermittent half-duplex transceivers for Primary Users (PUs), denoted as PU1 and PU2.

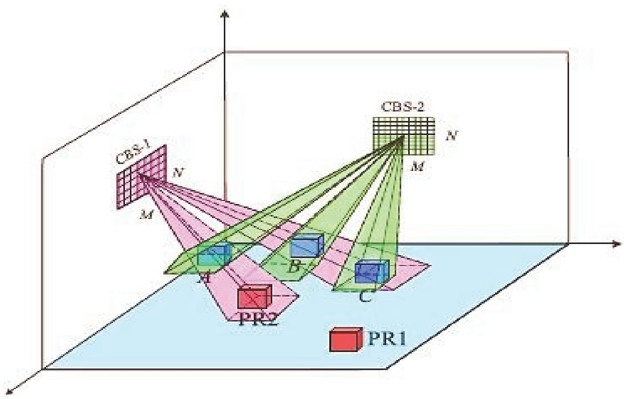

Figure 2 Cognitive base station (CBS) system in (3D) spatial environment

Moreover, we define key parameters: the wavelength (λ) and inter-antenna spacing (D_s), which is set to half of λ, as shown in Fig. 3. Our assumptions consider signal transmission from CR/PUs to CBS, which undergoes propagation through Pi clusters, each consisting of multiple independent resolvable paths. These paths are associated with elevation and azimuth Direction of Arrival (DoA) angles. Specifically, the p-th cluster is characterized by the average (θ_p) and angular spread ($\Delta\theta_p$) of the elevation DoA angle, and similarly, the mean (ϕ_p) and angular spread ($\Delta\phi_p$) of the azimuth DoA angle for the P-th cluster. This description outlines a 3D cognitive radio network scenario with two CBS, varying CR User counts, and a half-duplex PU transceiver pair. The network's distinctive features include antenna configurations, propagation clusters, and angular parameters, all influencing system behavior significantly.

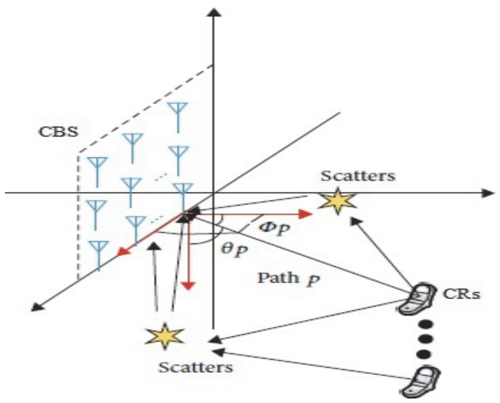

Figure 3 Formation of uniform rectangular array (URA) at cognitive base station (CBS)

As the signal traverses through the P-th cluster, the channel characteristics can be expressed by performing an integration across the angular domain, yielding the following formulation.

$$C_p = \int^{\theta_p+\Delta\theta_p} \cdots \int^{\phi_p+\Delta\phi_p} \beta\cdot(\theta,\phi)\,a\cdot(\theta)\,b^T\cdot(\theta,\phi)\,d\theta d\phi \tag{1}$$

The channel matrix, denoted as Cp for the Pth cluster, corresponds to specific angular information. In this context, the symbol (θ, ϕ) signifies the channel gain associated with a particular Direction of Arrival (DoA). θ represents the elevation angle, while ϕ signifies the azimuth angle. This channel gain (θ, ϕ) reveals the magnitude of the received signal's strength at a specific DoA, offering insights into signal propagation in the three-dimensional space, covering both vertical and horizontal dimensions. Eq. (2) precisely governs this representation.

$$(\theta,\phi) = |(\theta,\phi)|\cdot exp\{-j\phi(\theta,\phi)\} \tag{2}$$

where $|(\theta, \phi)|$ represents the magnitude of $\alpha(\theta, \phi)$, $\phi(\theta, \phi)$ is the phase of $\alpha(\theta, \phi)$, j is the imaginary unit. The channel gain $\beta(\theta, \phi)$ is influenced by both the magnitude and phase of the complex quantity $\alpha(\theta, \phi)$, which captures the behaviour of the signal propagation in terms of its amplitude and phase changes as it arrives from different directions in the angular domain (elevation angle θ and azimuth angle ϕ).

(θ) and (θ, ϕ) denote the array manifold vectors (AMVs), expressed as follows:

$$(\theta) = \left[1, e^{(jXC(\theta))}, e^{(jX(M-1))}\cdot\cos(\theta)\ \right]T \tag{3}$$

In the Eq. (3) (θ) is the array manifold vector for a specific angle θ, $e^{(jXC(\theta))}$ represents the complex exponential term raised to the power of the cosine of θ, where $X\cos(\theta)$ represents a constant value. $e^{(jX(M-1))}$ $\cos(\theta)$ corresponds to the complex exponential term raised to the power of $X(M-1)$ times the cosine of θ. M is the number of antenna elements in the array. T indicates the transpose operation. The vector (θ) encapsulates the antenna array's behavior across diverse angles θ, encompassing the inherent spatial receptivity of individual antenna elements to signals originating from different directions.

$$bT(\theta,\phi) = \left[1, e^{(jX\cos(\theta))}, \ldots, e^{(jX(M-1))}\cdot\cos(\theta)\ \right]T \tag{4}$$

The vector (θ, ϕ) characterizes the array response in terms of the given pair of angles (θ, ϕ), signifying the spatial sensitivity of the antenna elements to signals arriving from specific. We establish a steering matrix represented as $S(\theta, \phi) = a(\theta)bT(\theta, \phi) \in M \times N$, where $a(\theta)$ and $b(\theta, \phi)$ denote the antenna mode vectors (AMVs). This formulation leverages the characteristics of AMVs to construct the steering matrix, yielding a complex-valued matrix of dimensions $M \times N$. It encapsulates the spatial interactions and propagation effects within the communication environment. The dimensions of the steering matrix depend on the number of antennas and users in the system.The entries of the steering matrix are determined by the channel conditions, which include the path losses, channel gains, phases, and spatial signatures. The structure of the steering matrix can be mathematically expressed as:

$$\boldsymbol{A} = \begin{matrix} \begin{bmatrix} a(\theta_1,\phi_1)\cdot b(\theta_1,\phi_1)\cdot a(\theta_1,\phi_2)\cdot \\ \cdot b(\theta_1,\phi_2), \ldots, a(\theta_1,\phi_N)\cdot b(\theta_1,\phi_N) \end{bmatrix} \\ \begin{bmatrix} a(\theta_2,\phi_1)\cdot b(\theta_2,\phi_1)\cdot a(\theta_2,\phi_2)\cdot \\ \cdot b(\theta_2,\phi_2), \ldots, a(\theta_2,\phi_N)\cdot b(\theta_2,\phi_N) \end{bmatrix} \\ \ldots \\ \begin{bmatrix} a(\theta_M,\phi)\cdot b(\theta_M,\phi)\cdot a(\theta_M,\phi)\cdot \\ \cdot b(\theta_M,\phi), \ldots, a(\theta_M,\phi_N)\cdot b(\theta_M,\phi_N) \end{bmatrix} \end{matrix} \tag{5}$$

The symbol (θ, ϕ) represents the AMV for the receive antennas, capturing the spatial response of the channel at a specific angle (θ, ϕ), $b(\theta, \phi)$ represents the AMV for the transmit antennas, reflecting the spatial characteristics of the transmitted signals at the given angle (θ, ϕ). This structure accounts for the complex interplay of antennas, propagation, and user positions within the CR system. The steering matrix serves as a fundamental component for various CR tasks, such as beam forming, channel estimation, and interference management, enabling efficient and adaptive spectrum utilization.

The channel for the P-th cluster in terms of steering variables within Cognitive Radio (CR) is represented using a steering matrix that captures the spatial characteristics of the propagation environment. The steering matrix connects the transmitted signals from antennas to the received signals at different clusters, incorporating the spatial signatures associated with each cluster. Mathematically, the channel representation for the P-th cluster in terms of steering variables can be expressed as in Eq. (6):

$$C_p = ADB^H \tag{6}$$

The steering matrix $\boldsymbol{A}$ describes the response of receive antennas in the P-th cluster to signals from transmit antennas, while the diagonal matrix $\boldsymbol{D}$ incorporates the channel gains or path losses specific to that cluster. The conjugate transpose of the steering matrix $\boldsymbol{B^H}$ accounts for the spatial characteristics of the transmitted signals from the P-th cluster.

Indeed, each cluster within the Cognitive Radio (CR) system can be treated as an independent propagation path, with each path comprising both line-of-sight (LOS) and first-order reflection components. This perspective allows us to express the channel representation for the P-th cluster in a simplified form. Accordingly, the channel matrix C_p can be defined in Eq. (7).

$$C_p = \left(\theta_p, \phi_p\right) \cdot \boldsymbol{A}\left(\theta_p, \phi_p\right) \tag{7}$$

The variables (θ_p, ϕ_p) represent the combined effect of the channel gains associated with the LOS and first-order reflection components for the P-th cluster. This factor captures the signal strength and attenuation due to the propagation path characteristics. $\boldsymbol{A}(\theta_p, \phi_p)$ denotes the steering matrix associated with the P-th cluster, representing the spatial response of the channel.

The received signal strength at the Base Station (BS) of the Cognitive Radio (CR) system is expressed as in Eq. (8).

$$P_r = P_t \cdot G_t \cdot G_r \left[\lambda^2 \middle/ \left(4\pi d^2\right) \right] \tag{8}$$

This expression quantifies the received signal strength at the Base Station and considers factors such as transmitted power, antenna gains, wavelength, and the distance between the transmitter and receiver. It forms a fundamental metric for assessing the quality and effectiveness of signal transmission in a Cognitive Radio system.

3.1 Method of Extending Spatial Foundations

The innovative DFT-based Spatial Basis Expansion Model (SBEM) technique represents a paradigm shift in the realm of multi-user Massive MIMO systems. It revolutionizes the approach to uplink and downlink communication channels by capitalizing on the strengths of DFT and spatial domain representation. By adeptly harnessing the characteristics of uniform linear arrays, the DFT transformation matrix is strategically employed to create sparse channel representations in the spatial domain. This intelligent utilization of DFT aligns seamlessly with the array's physical properties, revealing a pattern of sparsity that enhances efficiency and reduces complexity. Through skilful management of user-specific spatial characteristics, SBEM orchestrates harmonious signal coexistence, effectively mitigating pilot interference issues. This approach's reduced reliance on complex channel statistics distinguishes it from established channel estimation methods, focusing on channel sparsity. The DFT-based SBEM technique for multi-user Massive MIMO systems offers heightened efficiency, reduced complexity, and increased adaptability by expertly organizing spatial characteristics and circumventing pilot interference, heralding a new era in next-generation wireless communication system design and optimization.

3.2 Iterative Hard Thresholding based Signal Arrival Estimation

The output of a massive Multiple-Input Multiple-Output (MIMO) antenna array at a given time index t in Eq. (9).

$$y(t) = A(\theta) \cdot s(t) + n(t),\ t = 1, 2, 3, ..., N \tag{9}$$

The equation itself indicates that the received signal at time index t, denoted as (t), is a linear combination of two main components:

The first component, (θ) $s(t)$, represents the contribution of the transmitted signals. It is the result of the interaction between the array response matrix and the transmitted signals. This part depends on the angle of arrival (θ) and the transmitted signals. The second component, (t), represents the noise or interference present in the received signals. It is typically modelled as additive white Gaussian noise (AWGN) and is independent of the transmitted signals. The Iterative Hard Thresholding (IHT) algorithm, which is used to address the sparse localization problem. This algorithm is characterized by its simplicity and iterative nature. It initializes with [0] = 0 and proceeds to follow a specific scheme for complex matrix operations. The scheme iterates as At each iteration i, compute Si + 1 using the formula given in Eq. (10).

$$Si + 1 = H' \cdot k \cdot \left(\left[i\right] + A' \cdot H' \cdot \left(Y - A'\left[i\right]\right)\right) \tag{10}$$

were $[i]$ denotes estimation of the sparse signal at iteration i, A' describes a matrix related to the measurement process or the system. It plays a role in mapping the signal to the observed data. The hard thresholding operator H'. It acts to retain the largest elements of a matrix while setting the rest to zero. This is often used to enforce by discarding small values with the observed data or measurements YSi + 1: The updated estimate of the sparse signal at iteration i + 1. To enhance the recovery performance, a modified version of the Iterative Hard Thresholding (IHT) algorithm is employed, utilizing a technique known as coherence-inhibiting. The coherence-inhibiting approach focuses on reducing the correlation between adjacent support atoms. This is achieved by removing certain support atoms, leaving behind those that have lower correlation with their neighbours. The objective is to create more favourable conditions for the sparse reconstruction of the signal S. This step is crucial because high correlation between atoms corresponding to different directions of arrival (DOAs) can lead to inaccurate estimations when sources impinge on the antenna array at closely spaced angles. Hence the manipulation of angles and parameters related to the CR system in defined in Eq. (11).

$$\left[\left(\alpha+\Delta\right)\right] \text{with } \Delta > 0° \tag{11}$$

where α: This represents an angle, possibly associated with the direction of arrival (DOA) of a signal or some other angular parameter in the CR system. Δ: This is another angle, likely representing a small positive angular increment or separation. $[(\alpha + \Delta)]$: This expression indicates a mathematical operation where the angle α is multiplied by the quantity $(\alpha + \Delta)$. In other words, it is the product of α and $(\alpha + \Delta)$.

A fundamental representation in array signal processing, often used in fields like radar, sonar, wireless communication, and sensor networks to understand and analyse the behaviour of antenna arrays in capturing and processing signals from various directions. It represents the array output

$$(t) = S(t)a(\alpha) + S(t)a(\alpha+\Delta) + n(t) \tag{12}$$

The first term, $S(t)(\alpha)$, represents the contribution of signals arriving from the direction α. It is the product of the signal vector $S(t)$ and the array response vector (α). The second term, $S(t)(\alpha + \Delta)$, represents the contribution of signals arriving from the direction $\alpha + \Delta$. It is the product of the same signal vector $S(t)$ and the array response vector $(\alpha + \Delta)$, which corresponds to signals slightly offset from the angle α. The last term, $n(t)$, represents the noise in the received signals.

3.3 Reinforcement Learning (RL) Based Rewards and Consequences Methodology

Implementing Reinforcement Learning (RL) in Cognitive Radio (CR) algorithms can offer an intelligent and adaptable approach to optimize spectrum utilization and decision-making. RL involves training an agent to learn optimal actions in an environment through trial and error. Here's a high-level overview of how RL can be implemented in a CR algorithm:

Step 1. Define the Problem and Environment: Identify the specific problem within the CR domain that you want to address using RL. Define the states, actions, rewards, and possible transitions in the environment. In CR, states might represent different spectrum bands, actions could be selecting a channel or transmission power, and rewards might reflect the quality of communication or interference levels.

Step 2. Choose an RL Algorithm: Select a suitable RL algorithm for your CR problem. Common algorithms include Q-learning, Deep Q Networks (DQN), Proximal Policy Optimization (PPO), and more advanced methods like Deep Deterministic Policy Gradient (DDPG). Choose an algorithm that fits the complexity and requirements of your problem. For the proposed system *Deep Q-learning* will be used for training the environment

Step 3. Design the RL Agent: Create an RL agent that interacts with theenvironment. The agent's task is to learn the best actions to take in each state to maximize cumulative rewards. Design the agent's neural network (if using deep RL), define the exploration strategy, and set hyper parameters.

Step 4. Training: Run the RL agent through training episodes in the CR environment. During each episode, the agent takes actions, receives rewards, and updates its policy or value functions based on the chosen RL algorithm. Training may involve thousands of iterations.

Step 5. Explore and Exploit: The agent learns to balance exploration (trying new actions) and exploitation (choosing actions with the highest expected rewards). This is crucial in CR, where the agent needs to discover optimal spectrum access strategies.

Step 6. Evaluation and Testing: After training, evaluate the trained RL agent's performance on a separate set of test scenarios. This helps ensure that the agent generalizes well to new situations and performs effectively in real-world CR environments.

Step 7. Fine-Tuning and Optimization: Adjust hyper parameters, network architectures, and other settings to optimize the RL agent's performance. You may need to experiment with different configurations to achieve desired results.

Step 8. Deployment and Integration: Once the RL agent demonstrates satisfactory performance, integrate it into the CR algorithm or system. The RL agent can make decisions about channel selection, power allocation, spectrum sensing, or other actions to optimize CR performance.

Step 9. Continual Learning and Adaptation: Implement mechanisms for the RL agent to continue learning and adapting in response to changing environmental conditions, ensuring the CR system remains adaptive and effective over time.

3.3.1 Defining Environment Variables

Defining the environment phase in the context of Cognitive Radio (CR) involves creating a model that represents the behaviour of the CR system, the available channels, primary users, noise, interference, and the interactions between these elements. This model serves as the basis for simulating and testing various CR algorithms and policies.

Primary Users (PUs):

- Model the behaviour of primary users, which are the licensed users of the spectrum.
- Define the activities and transmission patterns of primary users in terms of time, frequency, and power.

Secondary Users (SUs):

- Represent the secondary users, which are the unlicensed users seeking to utilize the available spectrum.
- Define the behaviour of secondary users, including channel selection, power control, and interference avoidance.

Let's say you have N channels, and at each time t, the occupancy status of channel n is denoted as $On(t)$, where $O(t) \in \{0, 1\}$. $On(t) = 1$ indicates channel n is occupied by a primary user at time t, and $On(t) = 0$ indicates it's available for secondary users. The channel occupancy can be define as follows in Eq. (13).

$$O(t) = \begin{cases} 1, \text{if a primary user is on channel } n \text{ at time } t \text{ 0,} \\ \text{otherwise} \end{cases} \quad (13)$$

These occupancy statuses affect the availability of channels for secondary users and impact their decisions on channel selection and transmission. Overall, building a comprehensive environment model for CR involves integrating these components and considering various mathematical equations that describe the behaviours and interactions of primary and secondary users, channel availability, interference, and noise. It is a critical step for evaluating and optimizing CR algorithms and policies before deployment.

Then noises in a CR environment are defined as Noise Phase Variation which not only affects the amplitude of the received signal but can also introduce phase variations. Phase noise is the random fluctuation of the phase of a signal, and it can arise from various sources, including imperfections in the transmitter and receiver components. In a more complex scenario, where phase noise is explicitly considered, the received signal with phase noise can be represented as in Eq. (14):

$$(t) = s(t) \cdot \exp\left(j\theta s(t)\right) + n(t) \quad (14)$$

Here, $\theta(t)$ represents the phase noise term that varies over time. The noise term (t) can also have phase variations, but for simplicity, let's focus on the phase variation introduced by phase noise. Phase noise can degrade the quality of the received signals, impacting various aspects of the CR system. For instance, in spectrum sensing, accurate detection of primary users' signals can be compromised if the phase noise masks the actual signal characteristics.

3.3.2 Designing of RL Agent

Consider a simplified cognitive radio scenario where the CR agent aims to select the optimal channel from a set of available channels. The goal is to maximize the cumulative reward over time while avoiding interference with primary users of the spectrum. Let's represent the state as a vector that includes information about the available channels, signal strengths, and interference levels. For simplicity, let's denote the state vector as "S". The action space consists of selecting a channel from the available channels. Let's denote the action as "a".

The reward function evaluates the quality of the chosen action in a given state. It encourages selecting channels with low interference and successful data transmission. The reward function can be defined as follows in Eq. (15).

$$R(s, a) = \\ = Successuful_{TransRwed}(s, a) - Intrfrnce_{Penalty}(s, a) \quad (15)$$

Here, $Successful_{TransRwrd}$ (s, a) represents the reward for successfully transmitting data on channel "a" in state "s", and $Intrfrnce_{Penalty}$ (s, a) represents the penalty for causing interference to primary users on channel "a" in state "s".

3.3.3 Training and Exploitation

The training phase of Deep Q-Learning (DQL) for Cognitive Radio (CR) applications involves the iterative process of updating the Q-values of actions in different states, enabling the agent to learn an optimal policy. The training loop consists of episodes, where each episode involves interactions with the environment. Within each episode.

Algorithm 1: Training of CR environment.

Step 1. Initialize the Q-network with random weights or pre-trained weights from another task (transfer learning).

Step 2. Set hyper parameters such as learning rate (α), discount factor (γ), exploration rate (ε), and target network update frequency.

Step 3. Then initialize the state: S.

Step 4. for each time step t within the episode:

- Choose an action using a ε-greedy or other exploration strategy: a_t = ε-greedy (s_t, Q), where s_t is the current state and Q is the Q-network $(s, a) = r + \gamma * \max_{a+1} Q(s_{+1}, a_{+1})$.
- Execute the action: a_t.
- Receive a reward: r_t and transition to the next state: s_{t+1}.
- Store the transition (s_t, a_t, r_t, s_{t+1}) in the experience replay buffer.

Step 5. Sample a batch of transitions from the experience replay buffer (s, a, r, s_{+1}) to update the Q-network. This helps stabilize learning by breaking correlations between consecutive samples.

Step 6. For each transition (s, a, r, s_{+1}), calculate the target Q-value using the Bellman equation:

$target\ (s, a) = r + \gamma \max_{a+1} Q(s_{+1}, a_{+1})$ .Update the Q-value of the chosen action in the current state using the target Q-value as $Q(s, a) = (1 - \alpha) * Q(s, a) + \alpha * targetQ(s, a)$.

Step 7. Periodically update the target network to stabilize training. This involves copying the weights of the Q-network to the target network.

Step 8. Decrease the exploration rate ε over time to shift from exploration to exploitation. End the episode after a predefined number of time steps or when a termination condition is met. Periodically evaluate the learned policy in an evaluation environment to track progress.

This training process iterates over episodes and gradually improves the Q-values, allowing the DQL agent to learn a policy that maximizes the expected cumulative reward. In the context of CR, the state, action, and reward definitions would be specific to the CR problem you are addressing, such as dynamic spectrum access, power control, or channel selection.

3.3.4 Continual Learning and Adaptation Phase

The continual learning and adaptation of Deep Q-learning in Cognitive Radio (CR) necessitates training the RL agent to navigate dynamic environments, ensuring it can adeptly adjust to evolving conditions. This is crucial for CR scenarios characterized by fluctuating wireless environments. The Q-learning update equation, described in Eq. (16), refines Q-values, encapsulating cumulative rewards for state-action pairs, serving as the core of this adaptive process.

$$(s,a) \leftarrow Q(s,a)+\alpha\cdot\left(r+\gamma\cdot \max a' Q(s',a')-Q(s,a)\right) \quad (16)$$

Here, s represents the current state, a is the action taken, r is the immediate reward, s' is the next state, a' is the next action, α is the learning rate, and γ is the discount factor. This equation signifies the continual update of Q-values based on new experiences. To adapt the DQN, strategies such as Experience Replay and Target Network Update are employed. Experience Replay stores past transitions (state, action, reward, next state) in a buffer, facilitating learning from diverse experiences. Target Network Update introduces stability by maintaining a separate network with delayed updates.

4 RESULTS AND DISCUSSION

We specify the number of users or channels, a range of SNR values in dB, and other system parameters. Inside the simulation loop, for each SNR point, we calculate the optimal power allocation to maximize energy efficiency for each user or channel. We assume equal power allocation initially and then adjust the power allocation for the user under consideration. We calculate the energy efficiency for each user and store the results in the matrix. Finally, we plot the Power Allocation vs. Energy Efficiency for each user, showing how energy efficiency changes with different power allocation strategies shown in Fig. 4

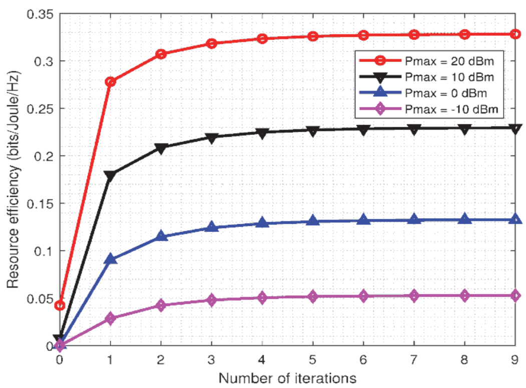

Figure 4 Energy efficient and power allocation

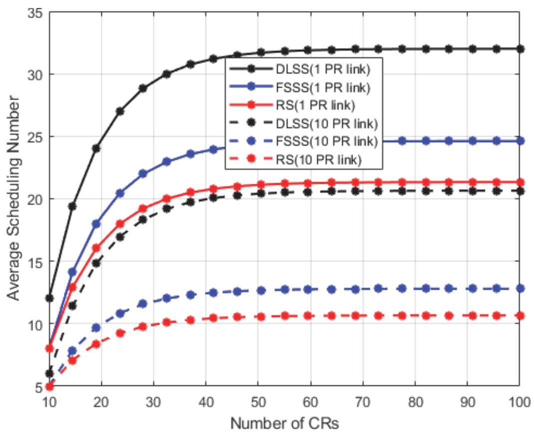

Figure 5 Average scheduling and number of CRs

Scheduled user of CR in the proposed Deep Learning based Spectrum Sharing (DLSS) is 33 which is 32% higher than the earlier work of Full Space Spectrum Sharing (FSSS) and 57% higher than Random Scheduling (RS) method respectively as shown in Fig. 5.

We specify a range of SNR values in dB, representing the SNR points you want to evaluate. For each SNR point, we calculate the energy efficiency, which is defined as the bit rate divided by the total power consumption (transmit power + received power). We plot SNR vs. Energy Efficiency to visualize how energy efficiency changes with varying SNR levels shown in Fig. 6.

The way an antenna transmits electricity into the surrounding environment is described in three dimensions by its radiation pattern. The far-field, or the distance away from the antennas, is where this type of pattern is often measured. It may be described simply as the power emitted in a certain direction by a short dipole antenna. Fig. 7 illustrates the radiation pattern of massive MIMO CBS and Tab. 1 shows the performance analysis.

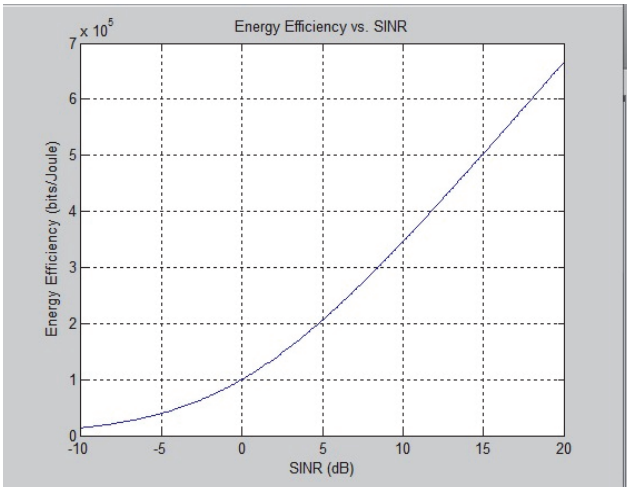

Figure 6 Energy efficiency and SNR

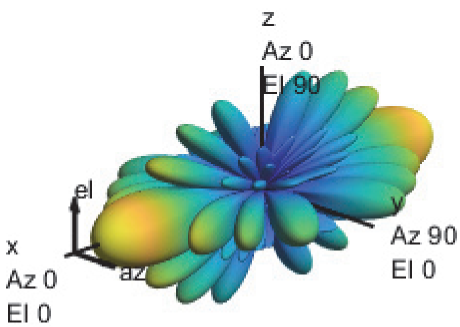

Figure 7 3Dradiation pattern of CBSURA antenna

Table 1 Performance analysis

SNR / Db	Transmit Power / w	Received power / w	Energy Efficiency / bps
–10	1.0	0.1	0.05
–8	1.2	0.12	0.06
–6	1.5	0.15	0.07
–4	2.0	0.2	0.08
–2	2.5	0.25	0.09
0	3.0	0.3	0.1
2	3.5	0.35	0.11
4	4.0	0.4	0.12
6	4.5	0.45	0.13
8	5.0	0.5	0.14
10	5.5	0.55	0.15

5 CONCLUSION

In conclusion, our study introduces an innovative strategy to address complex challenges in spectrum sharing within cognitive radio networks (CRNs). We leverage massive compound inputs and outputs (mMIMO) technology alongside cognitive radio principles, presenting a novel 3D spatial data acquisition approach enhanced by Deep Learning for informed decision-making. By integrating mMIMO structures into cognitive radio base locations (CRBS), we extract detailed angular insights from user equipment (UE), improving spatial resolution. We efficiently estimate Direction of Arrival (DoA) using the Iterative Hard Thresholding (IHT) technique and employ advanced Deep Learning for spectrum sensing, surpassing conventional reinforcement learning (RL) methods and enhancing CRN performance. Our spectrum scheduling strategy aims to maximize CR coverage and optimize average transmission rates in CRN mMIMO scenarios. Through careful tuning of deep learning parameters like learning rate and batch size, our model significantly improves Average Area Spectral Efficiency (AASR), surpassing established CRN spectrum sharing methods. Importantly, this progress is achieved without relying on reinforcement learning for spectrum sensing. Our research demonstrates the effectiveness of integrating Deep Learning into cognitive radio networks, paving the way for enhanced spectrum utilization and network performance. We also consider energy efficiency, aligning with sustainability goals. Rigorous experimental evaluation focuses on key metrics such as channel occupancy and energy efficiency. This study highlights the transformative potential of our approach, ushering in a new era of capabilities for CRNs and advancing spectrum allocation and network efficiency in next-generation wireless communication systems.

6 REFERENCES

[1] Vijay Anand, J., Manoharan, P. S., Jeyadheep Vignesh, J., Varatharajan, M., & Rubina Sherin, M. (2022). Spider Search Algorithms for MIMO System and Assessment Using Simatic PCS7. *Technical Gazette, 28*(4), 1118-1126. https://doi.org/10.17559/TV-20200513113443

[2] Akbar, N. S., Tufail, M., & Vasilakos, A. V. (2019). A survey of spectrum sensing algorithms for cognitive radio applications. *IEEE Access, 7*, 132805-132824.

[3] Nasir, A. A., Liu, L., & Duong, T. Q. (2017). Energy-efficient resource allocation in OFDM-based cognitive radio networks: A survey. *IEEE Communications Surveys & Tutorials, 19*(1), 558-577.

[4] Vasudevan, K. & Garg, V. K. (2018). Deep learning for cognitive radio networks: A comprehensive survey. *IEEE Communications Surveys & Tutorials, 20*(4), 2595-2620. https://doi.org/10.3390/s19010126

[5] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., & Petersen, S. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529-533. https://doi.org/10.1038/nature14236

[6] Xu, Y. & Zhang, R. (2018). A survey on multi-antenna techniques for millimeter-wave communications. *IEEE Transactions on Vehicular Technology, 67*(10), 9774-9789. https://doi.org/10.1109/OJCOMS.2020.3015394

[7] Chang, Z.,Wang, Z., Guo, X., & Han, Z. (2017). Energy-Efficient Resource Allocation for Wireless Powered Massive MIMO System with Imperfect CSI. *IEEE Transactions on Communications, 67*(9), 6327-6341. https://doi.org/10.1109/TGCN.2017.2696161

[8] Yaqiu, P. & Mingqi, L. (2020). A Novel PTS Scheme forPAPR Reduction of Filtered-OFDM Signals without Side Information. *Technical Gazette, 27*(4), 1305-1310. https://doi.org/10.1109/TBC.2017.2711141

[9] Nguyen, H. H., Nguyen, H. H., Tran, V. H., & Le-Ngoc, T. (2017). Joint spectrum sensing and access in cognitive Radio networks: A survey. *IEEE Transactions on Vehicular Technology, 66*(11), 10429-10442. https://doi.org/10.3390/s21072408

[10] Zhang, Y., Hu, R. Q., & Qian, Y. (2018). Cognitive radio networks with massive MIMO: A comprehensive survey. *IEEE Communications Surveys& Tutorials, 20*(3), 1740-1765. https://doi.org/10.3390/s19010126

[11] Rathika, P D. & Sivaraj, D. (2023). Spectrum Estimation and Optimal Secondary User Selection in Cognitive Radio Networks. *Technical Gazette, 30*(6), 1744-1752. https://doi.org/10.17559/TV-20230223000377

[12] Yang, K., Yuan, Y., & Guizani, M. (2019). Reinforcement learning for cognitive radio networks: Acomprehensive survey. *IEEE Access, 7*, 114403-114420. https://doi.org/10.1109/JIOT.2020.3040957

[13] Al-Rubaye, S. & Liang, Y. C. (2020). Deepreinforcement learning for spectrum sharing in cognitive radio networks: A survey. *IEEE CommunicationsSurveys & Tutorials, 22*(3), 1862-1885. https://doi.org/10.3390/s23052622

[14] Zhang, K., Wang, F., & Wang, C. X. (2021). Reinforcement learning for cognitive radio networks: An overview and recent advances. *IEEE Transactionson Cognitive Communications and Networking, 7*(1), 3-20.

[15] Jayamathi, A., Jayasankar, T., & Vinoth Kumar, K. (2022). Novel Selective Mapping with Oppositional Hosted Cuckoo Optimization Algorithm for PAPR Reduction in 5G UFMC Systems. *Technical Gazette, 29*(2), 464-471. https://doi.org/10.17559/TV20210524085655

Contact information:

Madhan KRISHNAMURTHI, Mr.
Department of Electronics and Communication Engineering,
Anna University Regional Campus-Madurai
E-mail: madhan287@gmail.com

Vinoth Kumar KALIMUTHU, Professor
(Corresponding Author)
Department of Electronics and Communication Engineering,
Vivekanandha College of Engineering for Women (Autonomous), Tiruchengode
E-mail: vinodkumaran87@gmail.com