(Received 19 November 2023, Accepted 27 December 2023, Scheduled 8 January 2024)

Q-Learning Empowered Cavity Filter Tuning with Epsilon Decay Strategy

Amina Aghanim*, Hamid Chekenbah, Otman Oulhaj, and Rafik Lasri

TED: AEEP, FPL, Abdelmalek Essaâdi University, Tetouan, Morocco

ABSTRACT: In the ever-evolving landscape of engineering and technology, the optimization of complex systems is a perennial challenge. Cavity filters, pivotal in Radio Frequency (RF) systems, demand precise tuning for optimal performance. This article introduces an innovative approach to automate cavity filter tuning using Q-learning, enhanced with epsilon decay. While reinforcement learning algorithms like Q-learning have shown effectiveness in complex decision-making, the exploration-exploitation trade-off remains a crucial challenge. The study conducts a thorough investigation into the application of epsilon decay in conjunction with Q-learning, employing the well-established epsilon-greedy strategy. This research focuses on systematically decaying the exploration rate ε over time, aiming to strike a balance between exploring new actions and exploiting acquired knowledge. This strategic shift serves to not only refine the convergence of the Q-learning model but also remarkably elevate the overall tuning performances. Impressively, this optimization is achieved with a notable reduction in the number of tuning steps, demonstrating an efficiency boost of up to 45 steps.

1. INTRODUCTION

In the field of microwave engineering, cavity filters are broadly used in distinct application fields owing to their exceptional performance and distinctive characteristics [1]. The key application domains include communication systems, such as satellite communications [2, 3], wireless communications [4], cellular networks, and base stations [5], where the filters are positioned at the front end connected to the antenna inside the transceiver [6]. Additionally, the domain of wireless technologies includes: Wi-Fi, Bluetooth, and RFID operating in the S-band. Furthermore cavity filters, renowned for their reliability, play a pivotal role in the defense and aerospace sector, guiding missile protocols, aiding electronic warfare, and enhancing radar systems [7]. Their contribution extends beyond accurate target detection to ensuring secure communications in these critical applications [8]. In television broadcasting, cavity filters are instrumental for precise frequency channel selection, optimizing transmission in television systems [9]. Moreover, their impact reaches the medical imaging domain, with applications in Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) [10]. Across these diverse domains, cavity filters prove essential, excelling in accurate signal processing, secure communication, and frequency management.

In the literature, numerous studies have been conducted on coaxial cavity filters, where researchers have focused their interest especially on the topology to enhance the rejection by minimizing the number of resonators [11]. Combline and interdigital filters are the typical variety of cavity filters because of their high-performance features, such as the high power handling capabilities [5, 12] since they are usually made of bulky metal, the low loss characteristics because of the high-quality factor Q ranging from [2000–5000] which enables them to be used in the front end of the UHF, both S and L bands, and receivers. Furthermore, combline and cavity filters are famous for their excellent stopband and high selectivity [13], where they can be used up to 10 GHz. The post-fabrication tuning capabilities represent highly appealing characteristics of these filter types, as they provide the opportunity to fine tune the mismatched frequency response of the filter by modifying the depth of screw penetrations [14, 15].

Unfortunately, in most cases of the manufactured microwave and Radio Frequency (RF) structures, the frequency response does not meet the required specifications due to design uncertainties, manufacturing tolerances, the adopted materials at the fabrication phase, and the insufficiency of exceedingly precise design models [16]. To address this issue, a post-fabrication tuning procedure has emerged as a decisive concern, wherein this process can be carried out manually or automatically [17], depending on the complexity of the circuit and the available resources.

The reason behind the necessity of this automated process is the complexity of cavity filters as they are treated as complex nonlinear systems, since there is no linear link between the depth of penetration of the screws inside the cavity filter and S-parameters. Moreover, manual tuning has always been considered a tedious and time-consuming task. Besides, tuning cavity filters require a vast amount of knowledge. This is why this task has always been delegated to technologists and experts.

Generally, when it comes to the tuning method of cavity filters, it is either based on the features, elements, and physical models of the filter such as the admittance matrix Y, S matrix, and coupling matrix M, or the data-driven modeling techniques which could be characterized by a mysterious internal

* Corresponding author: Amina Aghanim (amina.aghanim@etu.uae.ac.ma).

mechanism linking the tuned system inputs and outputs. The first approach could be adopted in the form of poles and zeros of the input reflection coefficient [18], circuit model parameter extraction [19], otherwise the method of time domain tuning [20]. The second approach could be achieved either by using Support Vector Regression (SVR) [17], Artificial Neural Network (ANN) [21], Fuzzy Logic Controllers (FLC) [22], or Neural-Fuzzy approach.

This paper presents an original approach based on using Epsilon decay strategy in the context of cavity filter tuning with the Q-learning. Consequently, the key contributions of this study encompass the following:

- Fine-tune filter parameters with precision, ultimately leading to superior filter characteristics.
- Find the right balance between exploration and exploitation during the tuning process. In the early stages of learning, a higher epsilon value encourages the algorithm to explore a wide range of tuning parameters.
- Progressively reduce the exploration rate over time as it accumulates experience and gains a better understanding of the plant behavior.
- Promote efficient convergence to the optimal filter settings and prevent spending excessive time exploring.

The adoption of the Q-learning algorithm with an epsilon decay strategy for tuning cavity filters stems from a careful evaluation of multiple factors that highlight its advantageous position over alternative techniques. The primary factor guiding this choice is the algorithm's inherent adaptability, allowing it to dynamically strike a balance between exploration and exploitation. The epsilon decay strategy systematically fine-tunes this balance over time, ensuring a nuanced approach to the exploration-exploitation trade-off. Unlike rule-based systems, such as FLCs [23], the simplicity and adaptability of Q-learning reduce the need for intricate rule bases, providing a more streamlined and efficient tuning process. Moreover, the model-free nature of Q-learning contributes to its computational efficiency, a crucial factor in real-time tuning scenarios. In contrast to more complex models like SVR [17], ANN, and Neural-Fuzzy approaches [21], which may demand substantial computational resources during training, Q-learning offers an elegant and effective solution with reduced computational complexity. The distinct advantage of Q-learning becomes even more pronounced when data requirements are considered. The algorithm's ability to operate with fewer labeled data points for training positions it favorably against supervised learning methods like SVR and ANN. This characteristic is particularly advantageous in scenarios where obtaining extensive datasets for training purposes may be challenging. Furthermore, Q-learning exhibits commendable generalization capabilities, allowing it to adapt efficiently to diverse filter characteristics. This stands in contrast to FLCs, which may struggle with generalization if the rule base is not comprehensive, and other machine learning models that may face challenges in handling unforeseen variations in filter characteristics.

In conclusion, the adopted method is rooted in its unique blend of adaptability, computational efficiency, reduced data requirements, and strong generalization capabilities. These qualities position Q-learning as a compelling and advantageous choice.

The subsequent structure of this paper is organized into distinct sections to offer an exhaustive insight of the application of Q-learning in cavity filters tuning. Section 2 examines the methodology where it elaborates the fundamentals of cavity filters, Reinforcement Learning (RL) and details the adaptation of Q-learning for cavity filter tuning. The article then proceeds to present experimental results and engage in in-depth discussions in Section 3. Finally, the conclusion section encapsulates the main findings and contributions of the study.

2. METHODOLOGY

2.1. Operating Mode of Cavity Filters

While insertion loss stands as a pivotal performance metric, quantifying the reduction in signal strength as it traverses the filter, return loss emerges as another crucial parameter. This metric specifically addresses the magnitude of power reflected back towards the source, a phenomenon primarily attributed to impedance mismatches within the filter. A higher return loss signifies a lower level of reflected power, indicating a superior impedance match and, consequently, more efficient power transmission through the filter.

In an optimally designed cavity filter, the aim extends beyond merely minimizing insertion loss within the passband [24]. It encompasses the strategic management of return loss to enhance the overall network performance. By ensuring substantial attenuation of stopband frequencies and minimizing reflected power, the filter upholds its performance integrity. Thus, return loss becomes an integral aspect of filter design, directly impacting the filter's efficiency and the reliability of the entire communication system [17].

2.1.1. Cavity Filters and S-Parameters

Scattering parameters, or S-parameters, play an essential role in assessing the performance of cavity filters. Specifically, S_{11} and S_{21} are vital metrics that provide deep insights into the behavior of the filter. The reflection coefficient at the input port, S_{11}, indicates the proportion of the signal that is reflected back from the filter. A low S_{11} value, particularly within the passband frequencies, is desirable as it signifies minimal signal reflection, ensuring that most of the incident power is transmitted through the filter. This is a critical parameter in assessing how well-matched the filter is to the source impedance, with lower values of S_{11} indicating better matching and, consequently, more efficient filter performance. On the other hand, S_{21} represents the transmission coefficient, measuring the portion of the signal that successfully passes through the filter from the input to the output port [24]. Elevated values of S_{21} within the passband typically correspond to low insertion loss. It is important to clarify that, particularly for passive lossless devices, the maximum magnitude of S_{21} is 1 or 0 dB. In this context, a value approaching 1 indicates minimal attenuation, highlighting the efficient transmission of desired frequencies through the filter.

In this purpose, careful optimization of these parameters ensures that the filter is well-matched and exhibits good insertion loss, which are key factors in achieving efficient and reliable signal transmission.

2.1.2. Tuning the Response of Cavity Filters with Screws

The mechanical adjustment of cavity filters, commonly achieved through the manipulation of screws or other variable components, stands as a prevalent technique for the refinement of filter performance. This practice primarily involves alterations to the physical dimensions of the filter cavities. By varying the positioning of the tuning screws, the effective volumetric attributes of the resonant cavity are modified, subsequently influencing the resonant frequency [25]. This adjustment enables the achievement of specific filter characteristics, ensuring alignment with desired performance criteria. Moreover, these adjustments extend beyond resonant frequency, impacting the filter's bandwidth and selectivity. It necessitates a strategic balance to optimize the filter's operation. However, it is imperative to acknowledge the challenges and required precision associated with mechanical tuning.

2.2. Reinforcement Learning

RL stands as a pivotal paradigm in machine learning, where an autonomous agent learns to make decisions by interacting with its environment. The unique nature of RL stems from its learning process, wherein the agent, through a series of actions and received feedback, iteratively refines its decision-making policy. The principles of RL are grounded in the agent's quest to optimize cumulative rewards over time, making it particularly well-suited for a myriad of optimization problems across diverse domains [26].

2.2.1. Basics and Structure

Diving deeper into the basics, RL operates on the premise of trial and error, where the agent explores the action space and learns from the consequences of its actions. The states provide a snapshot of the environment, encapsulating all the necessary information for decision-making. Actions, emanating from the agent, instigate changes in the environment, leading to new states and associated rewards [27]. The rewards serve as the immediate feedback, signaling the efficacy of the agent's actions. A positive reward reinforces the action taken in the particular state, while a negative reward discourages it. The policy, a critical component of the RL framework, dictates the agent's behavior, determining the actions taken in various states [28]. The policy can be deterministic, mapping states to specific actions, or stochastic, providing probabilities for each action in a state. The value function, often denoted as $V(s)$ for states or $Q(s, a)$ for state-action pairs, encapsulates the expected returns from states or actions, providing a measure of their long-term benefit. The learning algorithm, another cornerstone of RL, leverages the observed rewards and transitions to iteratively update the policy and value function, steering the agent toward optimal behavior.

2.2.2. Q-Learning

Q-learning represents a model-free RL algorithm that seeks to find the optimal policy, denoting the best action to take in each state to maximize cumulative rewards. This algorithm is integral in problems where the environment is uncertain, and the agent must learn from its interactions. The Q-learning algorithm functions by estimating the values associated with state-action pairs, represented as $Q(s_k, a_k)$, which indicates the expected cumulative rewards of taking action a_k in state s_k and subsequently following the optimal policy [29]. The core of the Q-learning algorithm involves iteratively updating these Q-values from the Bellman equation [28], which provides a recursive definition for the optimal policy. The update rule for Q-values in Q-learning is defined as in Eq. (1) [30]:

$$Q(s_k, a_k) \leftarrow Q(s_k, a_k) + \alpha[r(s_k, a_k) + \gamma \max(Q(s_{k+1}, a_{k+1})) - Q(s_k, a_k)] \quad (1)$$

where:

- s_k: The immediate state.
- a_k: The action taken.
- r: Immediate reward received after taking action a_k in state s_k.
- s_{k+1}: New state after taking action a_k.
- α: Learning rate, determining the weight of new experiences.
- γ: Discount factor, balancing the significance of instant and delayed rewards.

The term $\max(Q(s_{k+1}, a_{k+1}))$ represents the estimation of the optimal future value from the new state s_{k+1}, and the entire update rule makes the Q-values converge towards the optimal Q-values over time. Unlike model-based approaches, Q-learning does not mandate a model of the environment and can learn the optimal policy directly from interactions, making it particularly valuable for problems where the environment is complex or not fully understood. The convergence of the Q-learning algorithm to the optimal policy is guaranteed under certain conditions, such as all state-action pairs being visited infinitely often and a proper choice of learning rate and discount factor. This ensures that, given enough time and exploration, the algorithm will determine the most effective strategy that maximizes the total rewards, thus solving the optimization problem at hand. The illustration in Fig. 1 summarizes this process.

2.2.3. Exploration/Exploitation (Epsilon Greedy Strategy)

The fundamental concepts of exploration and exploitation play a pivotal role in guiding the decision-making process of our autonomous agent. Exploration involves the deliberate pursuit of novel actions or states within our research environment, aimed at uncovering previously unknown strategies and refining the agent's understanding over time. Conversely, exploitation centers on the selection of actions known to yield the highest expected rewards based on the agent's current knowledge or learned policy, with the objective of maximizing immediate gains. Striking an optimal balance between these two strategies is essential to ensure the robustness and efficiency of the

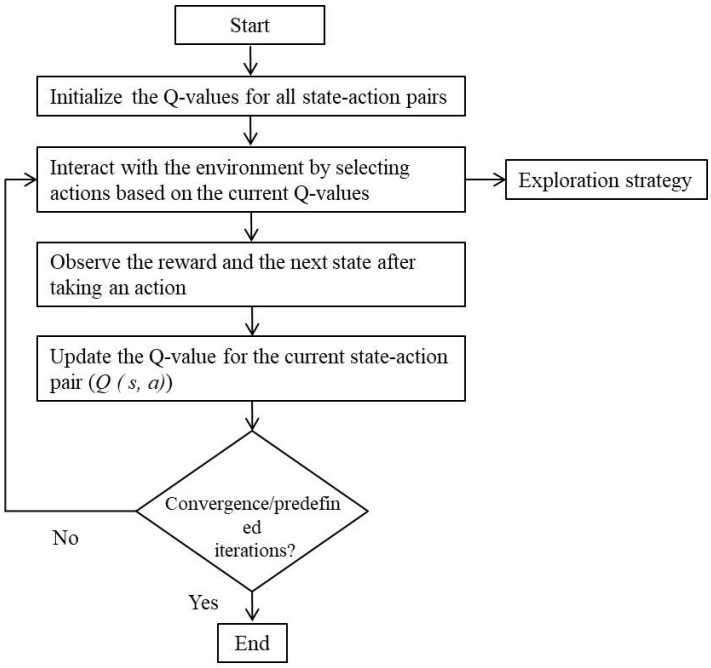

FIGURE 1. General Q-Leaning process algorithm.

model, as excessive exploration may delay rewards, while over-reliance on exploitation may hinder the discovery of superior strategies.

The epsilon greedy strategy provides an efficient method to balance exploration and exploitation, crucial elements in the RL paradigm. It operates under the governance of a parameter ϵ, which ranges between 0 and 1 and dictates the agent's propensity to explore or exploit. The mathematical framework for this strategy can be delineated as follows [31]:

- Action Selection: The agent picks the action a that maximizes the estimated Q-value for the immediate state s, with a probability of $1-\varepsilon$, which could be described as: $a = \max Q(s, a')$. Alternatively, with a probability of ε, the agent elects an action a at random from the available action space $A(s)$, fostering exploration.
- Q-value Update: Subsequent to action execution, the Q-value for the state-action pair (s, a) is updated. This update hinges on the received reward r as in Eq. (1). This mathematical model underpins the epsilon greedy strategy, ensuring a balanced approach to learning by intertwining periods of exploration with phases of exploitation, thereby enhancing the agent's performance and learning efficiency.

2.3. Tuning Cavity Filters with Q-Learning

In this section, we delve into the application of Q-Learning, to the specific task of tuning cavity filters. The goal is to optimize the filter's performance through intelligent adjustments based on the feedback received from the environment as in Fig. 2.

2.3.1. Specifications

The goal of tuning our cavity filters is to enhance the filter's attributes in alignment with specific performance criteria, with a particular focus on maximizing the return loss to a targeted 21 dB. This tuning process is carried out on a 6th-order combline cavity filter, where the outer dimensions are determined by a housing and corresponding cover precisely crafted from 6061 aluminum alloy and LY12 aluminum alloy, respectively. The filter measures 164 mm in length, 52 mm in width, and 34 mm in height. To achieve precise adjustments, the filter incorporates copper tuning screws, each with a diameter of 10 mm and a height of 43 mm. The combline cavity filter is uniquely configured with a central frequency of 941 MHz. In the tuning procedure, four tuning screws are available, although the active utilization is limited to two screws with a full rotation angles of 360° in both directions. This scientific approach ensures a systematic and thorough optimization of the cavity filter's characteristics to meet the specified performance requirements. During the evaluation, the primary emphasis is placed on the S_{11} coefficient within the scattering parameters. This evaluation is performed with the Vector Network Analyzer measurement setup using inputs and output ports. These parameters, along with the current positions of the tuning screws, serve as the sole inputs for making adjustments, aiming to emulate the precision of human tuning. The performance of the filter is continuously assessed against the return loss specification, with the tuning process striving to find the optimal configuration that meets this requirement. In the context of our Q-Learning application, meeting or exceeding the return loss target results in positive rewards, while any deviation incurs penalties.

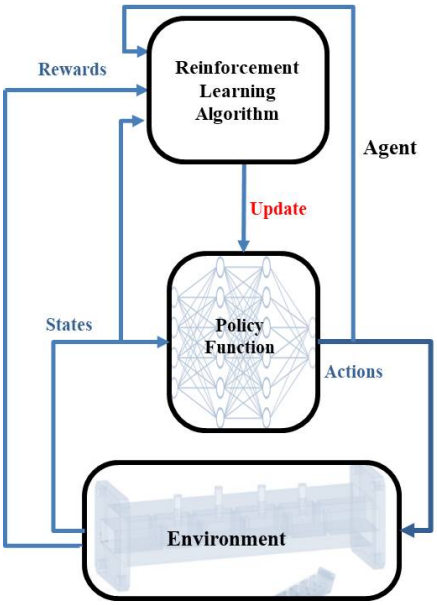

FIGURE 2. Q-leaning in tuning cavity filter.

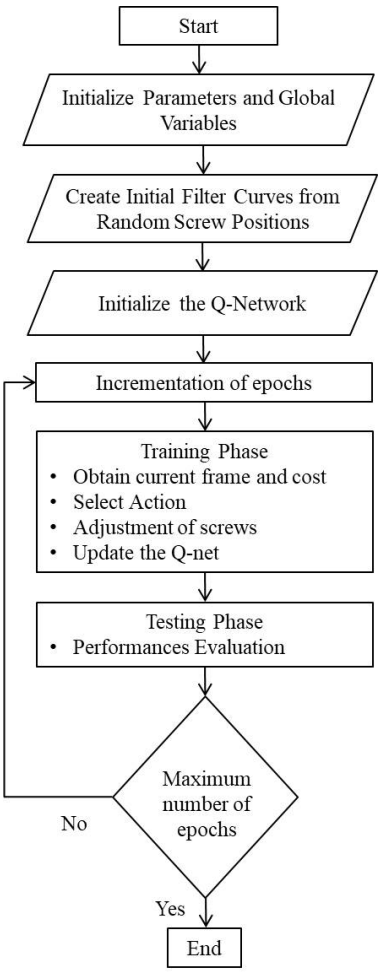

FIGURE 3. The adopted Q-Leaning algorithm.

This reinforcement mechanism guides the learning process of the agent, allowing it to incrementally refine its strategy and enhance the filter's performance. It is imperative to note that the passband fluctuations and stopband ripples, represented by the S_{21} coefficient, are not considered in this tuning process; instead, the focus is solely on the curve of return loss, S_{11}, as the state. The filter is considered optimally tuned only when it achieved the predefined criteria based on the magnitude of the scattering parameters. In other words, the tuning success is contingent on the curve of S_{11} being below the fixed return loss line of 21 dB. This approach ensures a systematic and data-driven process for filter tuning, aligning with the specific criteria associated with scattering parameter magnitudes.

2.3.2. Architecture of the Proposed Q-Learning Model

The Q-Learning model proposed for this task consists of several key components. The state space represents the possible configurations of the cavity filter, with each state corresponding to a specific combination of tuning parameters. The action space includes the possible adjustments that can be made to the filter, such as turning the tuning screws by a certain amount.

The Q-table holds the values associated with each state-action pair, representing the expected cumulative reward of selecting a specific action from a particular state. The Q-values undergo iterative updates determined by the rewards received and the predicted future rewards, adhering to the Q-Learning update rule. The agent interacts with the environment (the cavity filter) by taking actions (adjusting the tuning parameters) and receiving feedback in the form of rewards or penalties. This engagement allows the agent to acquire the most effective tuning approach as time progresses

Fig. 3 illustrates the flowchart of the adopted Q-learning algorithm through a Deep Q Networks (DQN) RL framework. Initially, we establish a solid foundation by setting up essential parameters and global variables, along with defining the initial filter curves. This initial phase is imperative as it prepares the system for the subsequent optimization tasks. Following this foundational setup, we introduce a Q Network (Qnet), a predictive model designed to identify the optimal actions based on the current state of the filter parameters.

The iterative optimization process unfolds through distinct epochs, each comprising a dedicated training and testing phase. Within the training phase, the algorithm engages in a sophisticated tuning routine. This sequence begins with the extraction of the current filter curve and the computation of its associated cost, offering a quantitative evaluation of the filter's performance. The choice of action is determined through an

epsilon-greedy strategy, delicately navigating the balance between exploring novel parameter configurations and exploiting known effective ones. This strategic choice results in adjusted filter parameters and an updated Qnet, implemented through a mini-batch approach.

$$Q(s', a') = R(s_k, a_k) + \gamma^* \max(Q(s_{k+1}, a_{k+1})) \quad (2)$$

where:

- $R(s_k, a_k)$ is the immediate reward.
- $\gamma^* \max(Q(s_{k+1}, a_{k+1}))$ the devaluated/upcoming reward.

Parallelly, Eq. (2) as the learning mechanism or the Q-function, iteratively refines the policy and propels the algorithm toward optimal cumulative rewards. Consequently, maintaining an optimal discounted upcoming reward decisively determines the action. $Q(s', a')$ represents the summation of the immediate and upcoming rewards.

In cases where the S_{11} parameters are already situated below the specified target line, the designated distance $x(i)$ is deemed zero. Conversely, if the S_{11} parameters fall short of the target line, the distance $x(i)$ is computed as the absolute difference between the two curves as described in Eq. (3). A reward mechanism is then applied, assigning a reward of 1 for reduced distance, indicating an improved alignment with the target. Conversely, a reward of 0 is assigned if the distance does not decrease. Each encountered state prompts the system to take a corresponding action, receiving the associated reward and transitioning to the subsequent s_{k+1}, with the entire sequence being meticulously recorded. Following this, a mini-batch is extracted from the stored sequences to facilitate the training of the designated network. Additionally, the future action a_{k+1} is determined strategically to maximize the Q-value of the future state s_{k+1} in the forthcoming cycle.

$$x(i) = \begin{cases} |S_{11}(i) - TRL|, & S_{11} > TRL \\ 0, & \text{if } TRL \geq S_{11}(i) \end{cases} \quad (3)$$

On the other hand, Eq. (4) defines the cost function in terms of the transformed S_{11} value, penalizing larger deviations and guiding the algorithm toward better filter performance [30]. On the other hand, the cost is directly related to Temporal Difference Error δ as depicted in Eq. (4).

$$\text{Cost}(s_k) = |x(i)|^2 = \frac{1}{2}\delta^2 \quad (4)$$

The quadratic form in Eq. (3) transforms the scattering parameter S_{11} into a reward signal, guiding the algorithm towards configurations that minimize S_{11} and satisfy the threshold Target Return Loss (TRL).

Lastly, Eq. (5) calculates δ, a critical element for the stability and convergence of the Q-Learning algorithm. Upon concluding each epoch, we arrive at a critical juncture. The decision here is straightforward: we will invariably proceed to the subsequent epoch, continuing this progression until we reach the predefined maximum number of epochs.

$$\min(r_k + \gamma_{a_{k+1}} \max Q'(s_{k+1}, a_{k+1}; \mu; \beta) - Q(s_k, a_k; \mu; \beta))^2 \quad (5)$$

Only then will we assess if the optimization process has satisfactorily aligned with our performance targets. This method ensures that the entirety of the optimization process is exhaustively explored, leveraging all available epochs to refine and enhance filter performance.

2.3.3. Optimizing the Exploration-Exploitation Strategy (By Adopting the Epsilon Decay Strategy)

The epsilon-greedy strategy relies on a fixed epsilon value to determine the trade-off between exploration and exploitation. While this approach provides a straightforward mechanism to balance these aspects, it may not be optimal throughout the whole learning process. To address this, we transition to the Epsilon Decay Strategy, where epsilon is no longer static but dynamically adjusted. A mathematical model defines how epsilon decreases over time. We have considered the exponential decay, which is expressed as in Eq. (6):

$$\varepsilon(t) = \varepsilon_0 \cdot e^{-\beta t} \quad (6)$$

where:

- $\varepsilon(t)$: The epsilon value at time t.
- ε_0: The initial value of epsilon.
- β: The decay rate.
- t: The number of episodes.

The integration of this strategy optimizes our algorithm in several ways: firstly, in the initial stages, when the Q-values are less accurate, the high epsilon value ensures extensive exploration. As the Q-values stabilize, the decreasing epsilon fosters exploitation of the acquired knowledge. Secondly, in terms of efficient convergence: by dynamically adjusting epsilon, the strategy helps the algorithm converge more quickly to the optimal set of filter parameters, reducing the time and computational resources required. Also, as epsilon decreases, the algorithm focuses more on exploitation, allowing for fine-tuning of the filter parameters and optimization of the filter performance.

3. EXPERIMENTAL TASK AND RESULTS

In our research, the Q-learning model was trained and tested over consistent settings: 100 epochs for training with up to 1000 tuning steps each and 100 epochs for testing with up to 200 tuning steps each. The training dynamics using the Epsilon Greedy strategy is showcased in Fig. 4.

During the initial phase, the tuning process displayed a dynamic pattern, varying between 1000 and 50 tuning steps. After the first 45 epochs, a consistent pattern began to manifest, hinting at convergence. However, a noticeable alteration appeared around the 80th epoch, shifting the tuning dynamics. The variations observed warrant deeper examination into the Q-learning model's behavior. Potential factors might include model sensitivity to specific epochs or complexities inherent to the learning environment. Nevertheless, the desired state remained elusive for the Q-network model.

The testing phase dynamics, as detailed in Fig. 5, also varied, fluctuating between 200 and 50 tuning steps. The out-

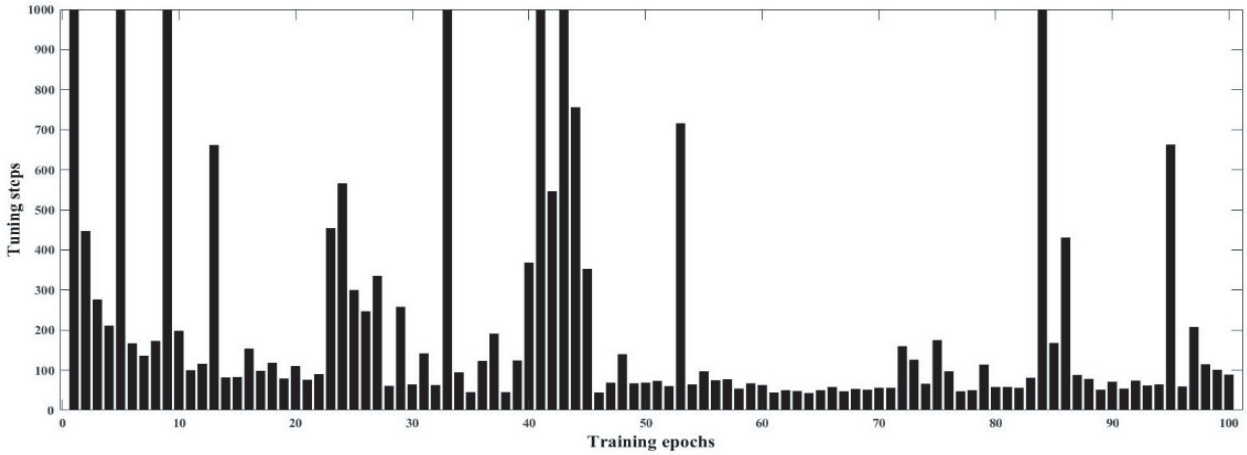

FIGURE 4. Training phase of Epsilon greedy Strategy.

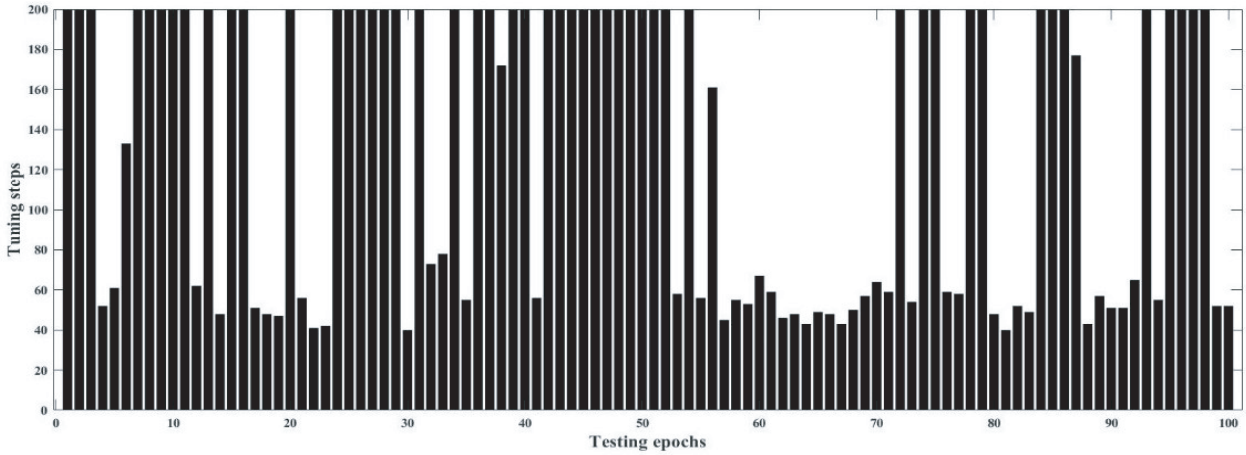

FIGURE 5. Testing phase of Epsilon greedy Strategy.

comes, however, indicated challenges in achieving optimal tuning. This calls for a deeper assessment of the model's adaptability to testing dynamics and a comprehensive understanding of the factors affecting the tuning process.

Figure 6 illustrates the training dynamics for a distinct experiment that combined both the Epsilon greedy and decay strategies. Initial phases saw the model grappling with achieving desired outcomes. Nevertheless, a significant shift in learning behavior was evident post the 20th epoch, marking a crucial turning point. Following this, the model displayed enhanced learning and began converging to improved outcomes.

The model's performance, after training with the combined strategies, is depicted in Fig. 7. This phase witnessed a consistent success in the tuning process, often achieving the target within impressive 40 steps. This robust performance indicates the potential of the combined strategy approach, emphasizing its efficacy and adaptability. The combined approach accelerates convergence while minimizing tuning iterations needed for optimal performance. While the Epsilon-Greedy strategy starts with extensive exploration due to a high initial epsilon value, leading to prolonged discovery of optimal actions, the Decay Epsilon approach, in contrast, decreases epsilon more rapidly, enabling a faster transition from exploration to exploitation. This results in quicker convergence by capitalizing on accumulated knowledge.

Fig. 8 illustrates the variation in exploration levels of two strategies employed in the Q-learning algorithm: the epsilon greedy strategy and epsilon decay strategy. At a first glance, both strategies begin with a high exploration rate. However, as epochs progress, distinct behaviors between the two become evident:

- Epsilon Greedy Strategy: This strategy appears to follow a consistent decline, reaching a somewhat stable exploration rate. While this ensures that there is always some level of exploration, it may also imply that the model continues to second-guess its decisions even after many learning iterations. The consequence of this can be a slower learning process and potentially sub-optimal results, especially if the algorithm continually explores options it has already determined to be non-optimal.
- Epsilon Decay Strategy: The epsilon decay strategy demonstrates a more aggressive decline in exploration. Early on, the system is open to trying out various possi-

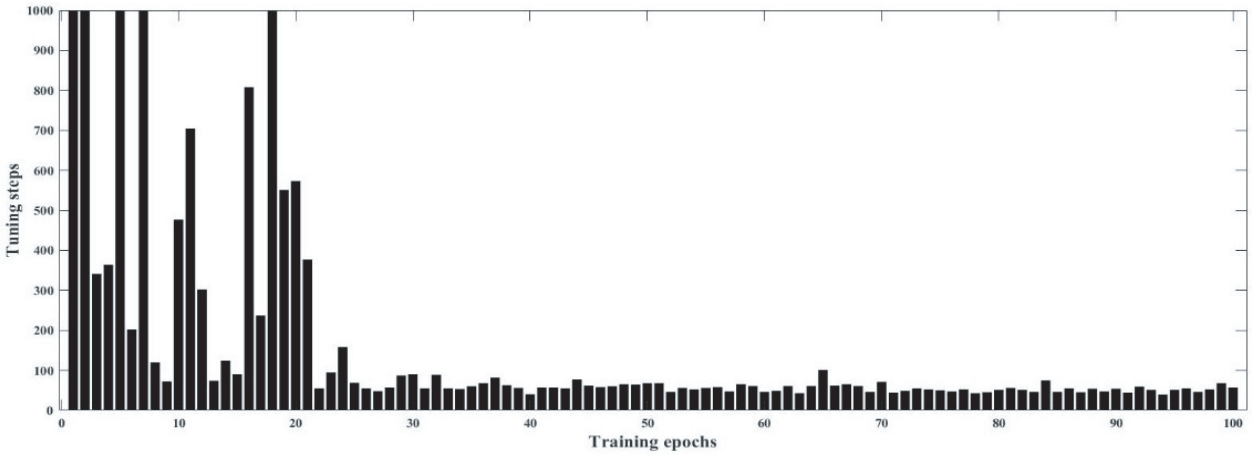

FIGURE 6. Training phase of Epsilon decay Strategy.

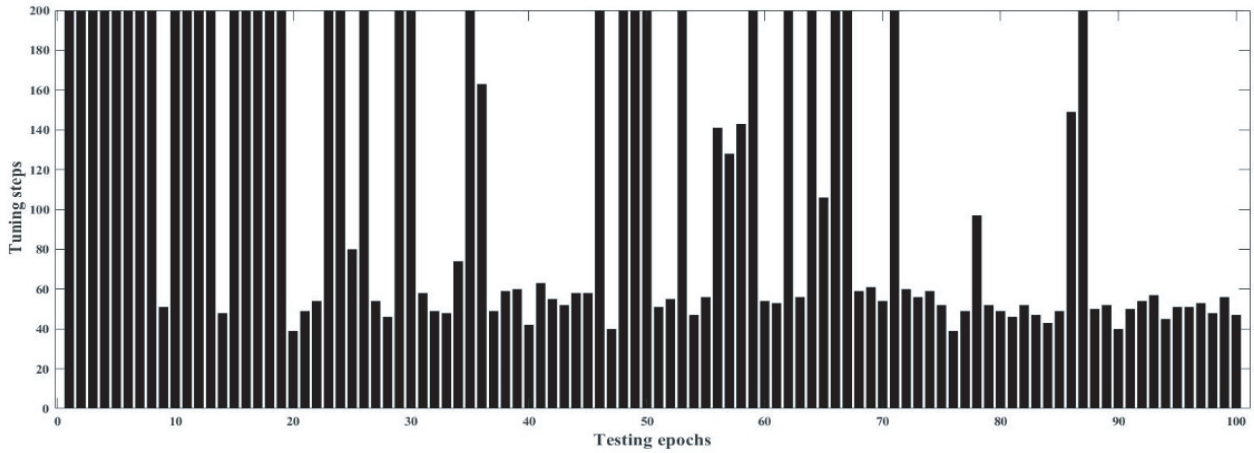

FIGURE 7. Testing phase of Epsilon decay Strategy.

bilities, ensuring a thorough search of the solution space. As learning advances, the rapid reduction in exploration implies the system becomes increasingly confident in its decisions. By allowing exploration to decrease more significantly, the algorithm transitions from a broad search to fine-tuning its choices, focusing more on exploiting the best-known actions. This potentially results in faster convergence to optimal or near-optimal solutions.

In terms of efficiency, the epsilon decay strategy appears advantageous. By tailoring the exploration-exploitation balance over time, it is ensured that the system learns efficiently. Initially, when the knowledge about the environment is limited, a higher exploration rate aids in understanding the landscape. As familiarity grows, the decreased reliance on exploration means that the system can capitalize on its accumulated knowledge, optimizing its actions based on prior learning.

Fig. 9 vividly demonstrates the progressive tuning results achieved through the novel technique at various frequency intervals. Each curve represents a distinct stage in the regulation process, with the S_{11} coefficient charted against frequency. Examining the curve of the initial step in yellow, where the tuning screws have zero penetration depth, i.e., the initial state before regulation, the return loss remains below -7 dB, which is considered suboptimal. Conversely, the brown curve corresponding to step 30 exhibits a return loss ranging from -10 to -17 dB across the entire passband. On the other hand, the blue curve representing step 41 demonstrates a remarkable return loss of up to -30 dB. Unfortunately, this performance is not sustained throughout the entire passband, as there is a noticeable deviation below the specified line of -21 dB between 902 and 910 MHz.

On the other hand, in the curve that represents the outcomes at the 45th tuning step, it is evident that a notable achievement has been reached: The S_{11} values for all sampled frequencies comfortably lie within the desired passband criterion of -21 dB. This not only validates the effectiveness of the Q-learning algorithm but also underscores its ability to swiftly align system performance with the targeted threshold. Furthermore, the achievement of this benchmark within just 45 tuning iterations stands as a testament to the efficiency and speed of the Q-learning optimization. It is noteworthy how the system behavior was seamlessly calibrated to meet the stringent per-

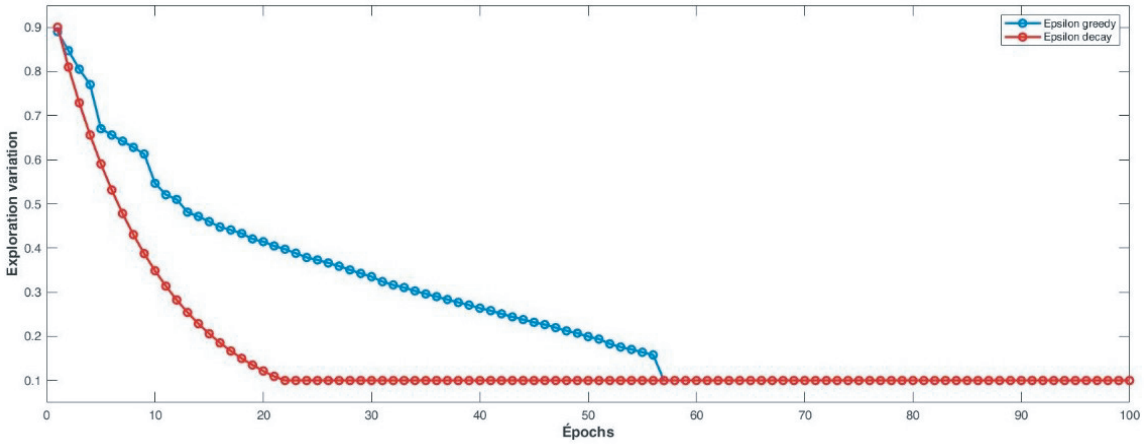

FIGURE 8. Exploration variation of both Epsilon greedy and decay strategies.

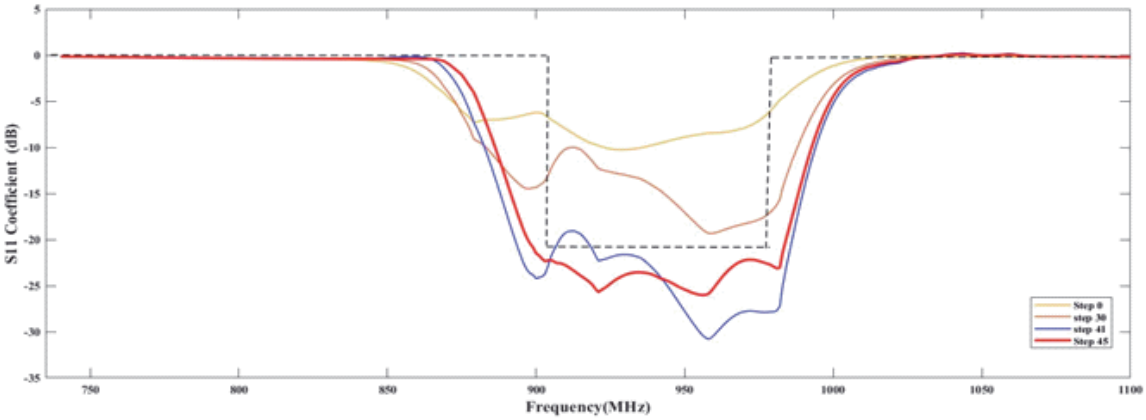

FIGURE 9. Test experimental results of the trained model.

formance objective of -21 dB, demonstrating the robustness of the adopted technique.

4. CONCLUSION

This research presented a rigorous application of the Q-learning algorithm to the intricate process of cavity filter tuning. Initially, the epsilon greedy strategy was employed, a method commonly favored for its simplistic balance between exploration and exploitation. While insightful, this strategy presented certain limitations in achieving the desired filter tuning outcomes. To address these challenges, we transitioned to the epsilon decay strategy. This adaptive methodology, which strategically diminishes the exploration rate over time, exhibited more efficient system behavior. The adoption of epsilon decay resulted in a more streamlined and effective optimization process, circumventing some of the obstacles associated with the epsilon greedy approach.

A notable milestone in our work was the successful fine-tuning of the S_{11} coefficient across diverse frequency steps. This achievement, realized through the novel application of the Q-learning algorithm with the epsilon decay strategy, highlighted the method's capability to meet stringent filter tuning standards. The rapid convergence observed affirmed the efficacy of the Q-learning algorithm when being combined with an optimal exploration strategy. Moreover, a comparative analysis between epsilon greedy and epsilon decay strategies bolstered our findings. This analysis, though non-graphical in nature, served to emphasize the advantages of the epsilon decay approach for this specific application. In conclusion, our research illuminates the potential of the Q-learning algorithm, particularly when it is paired with the epsilon decay strategy, in revolutionizing cavity filter tuning. This innovative methodology promises enhanced efficiency, speed, and precision in tuning procedures, setting a new standard in the field.

REFERENCES

[1] Xie, Y., F.-C. Chen, and Q.-X. Chu, "Tunable cavity filter and diplexer using in-line dual-post resonators," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 70, No. 6, 3188–3199, Jun. 2022.

[2] Rehman, A. and C. Tomassoni, "Spurious self-suppression method: Application to TM cavity filters," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 71, No. 3, 1201–1215, Mar. 2023.

[3] Laplanche, E., N. Delhote, A. Perigaud, O. Tantot, S. Verdeyme, S. Bila, D. Pacaud, and L. Carpentier, "Tunable filtering devices in satellite payloads: A review of recent advanced fabrication technologies and designs of tunable cavity filters and multiplexers using mechanical actuation," *IEEE Microwave Magazine*, Vol. 21, No. 3, 69–83, Mar. 2020.

[4] Mansour, R. R., "RF filters and diplexers for wireless system applications: State-of-the-art and trends," in *Radio and Wireless Conference, 2003. RAWCON'03. Proceedings*, 373–376, IEEE, Boston, Massachusetts, USA, Aug. 2003.

[5] Mansour, R. R., "Filter technologies for wireless base stations," *IEEE Microwave Magazine*, Vol. 5, No. 1, 68–74, Mar. 2004.

[6] Zhao, B. and F. Yang, "Compatibility evaluation and technical analysis of C-band broadcasting satellite receiving stations against 5G base station," in *Second International Conference on Digital Signal and Computer Communications (DSCC 2022)*, Vol. 12306, 162–168, SPIE, Changchun, China, Aug. 2022.

[7] Zhao, J., C. Ma, J. Zhou, J. Li, J. Liu, and Y. Luo, "Design of wide stopband and high suppression cavity filter," in *2023 24th International Vacuum Electronics Conference (IVEC)*, 1–2, IEEE, Chengdu, China, 2023.

[8] Varikuntla, K. K. and R. Singaravelu, "Review on design of frequency selective surfaces based on substrate integrated waveguide technology," *Advanced Electromagnetics*, Vol. 7, No. 5, 101–110, Nov. 2018.

[9] Li, M., Y. Yang, F. Iacopi, M. Yamada, and J. Nulman, "Compact multilayer bandpass filter using low-temperature additively manufacturing solution," *IEEE Transactions on Electron Devices*, Vol. 68, No. 7, 3163–3169, Jul. 2021.

[10] Rehman, H. Z. U., H. Hwang, and S. Lee, "Conventional and deep learning methods for skull stripping in brain MRI," *Applied Sciences*, Vol. 10, No. 5, 1773, Mar. 2020.

[11] Suryanarayana, G., K. Chandran, O. I. Khalaf, Y. Alotaibi, A. Alsufyani, and S. A. Alghamdi, "Accurate magnetic resonance image super-resolution using deep networks and gaussian filtering in the stationary wavelet domain," *IEEE Access*, Vol. 9, 71 406–71 417, 2021.

[12] Palanisamy, S., B. Thangaraju, O. I. Khalaf, Y. Alotaibi, and S. Alghamdi, "Design and synthesis of multi-mode bandpass filter for wireless applications," *Electronics*, Vol. 10, No. 22, 2853, Nov. 2021.

[13] Song, J., B. Deng, Y. Pang, and L. Sun, "A compact and high selective combline bandpass filter using GaAs IPD technology," in *2019 IEEE 5th International Conference on Computer and Communications (ICCC)*, 309–312, IEEE, Chengdu, China, 2019.

[14] Sekhri, E., R. Kapoor, and M. Tamre, "Double deep Q-learning approach for tuning microwave cavity filters using locally linear embedding technique," in *2020 International Conference Mechatronic Systems and Materials (MSM)*, 1-6, IEEE, Bialystok, Poland, Jul. 2020.

[15] Aghanim, A., R. Lasri, and O. Oulhaj, "Implementation of a fuzzy controller to tune the response of a waveguide cavity filter," *E-Prime — Adv. Electr. Eng. Electron. Energy*, Vol. 2, 100078, 2022.

[16] Yigit, Y. and O. Suvak, "Control architecture for autonomous RF cavity filter and multiplexer tuning," in *2022 IEEE Autotestcon*, 1-5, IEEE, National Harbor, MD, USA, Aug. 2022.

[17] Lindstahl, S. and X. Lan, "Reinforcement learning with imitation for cavity filter tuning," in *2020 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, 1335–1340, IEEE, Boston, MA, USA, Jul. 2020.

[18] Amari, S. and G. Macchiarella, "Synthesis of inline filters with arbitrarily placed attenuation poles by using nonresonating nodes," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 53, No. 10, 3075–3081, Oct. 2005.

[19] Thal, H. L., "Computer-aided filter alignment and diagnosis," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 26, No. 12, 958–963, 1978.

[20] Alvarez, J., L. D. Angulo, A. R. Bretones, and S. G. Garcia, "A spurious-free discontinuous galerkin time-domain method for the accurate modeling of microwave filters," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 60, No. 8, 2359–2369, Aug. 2012.

[21] Sadrossadat, S. A., Y. Cao, and Q.-J. Zhang, "Parametric modeling of microwave passive components using sensitivity-analysis-based adjoint neural-network technique," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 61, No. 5, 1733–1747, May 2013.

[22] Bi, L., W. Cao, W. Hu, and M. Wu, "A dynamic-attention-based heuristic fuzzy expert system for the tuning of microwave cavity filters," *IEEE Transactions on Fuzzy Systems*, Vol. 30, No. 9, 3695–3707, Sep. 2022.

[23] Peng, S., W. Cao, L. Bi, Y. Yuan, and M. Wu, "A tuning strategy for microwave filter using variable universe adaptive fuzzy logic system," in *2021 China Automation Congress (CAC)*, 6061–6066, Oct. 2021.

[24] Yao, S., X.-C. Wei, and L. Ding, "A deembedding method for the S-parameter extraction of surface-mounted devices with asymmetric fixtures," *IEEE Microwave and Wireless Components Letters*, Vol. 31, No. 2, 211–214, Feb. 2021.

[25] Yigit, Y. and E. Afacan, "Autonomous RF cavity filter tuning," *IEEE Instrumentation & Measurement Magazine*, Vol. 26, No. 5, 39–44, Aug. 2023.

[26] Nian, R., J. Liu, and B. Huang, "A review on reinforcement learning: Introduction and applications in industrial process control," *Computers & Chemical Engineering*, Vol. 139, 106886, Aug. 2020.

[27] Mlika, Z. and S. Cherkaoui, "Network slicing with MEC and deep reinforcement learning for the internet of vehicles," *IEEE Network*, Vol. 35, No. 3, 132–138, May 2021.

[28] Akalin, N. and A. Loutfi, "Reinforcement learning approaches in social robotics," *Sensors*, Vol. 21, No. 4, Feb. 2021.

[29] Clifton, J. and E. Laber, "Q-Learning: Theory and applications," *Annu. Rev. Stat. Its Appl.*, Vol. 7, No. 1, 279-301, 2020.

[30] Ding, Z., Y. Huang, H. Yuan, and H. Dong, "Introduction to reinforcement learning," *Deep Reinforcement Learning: Fundamentals, Research and Applications*, 47–123, 2020.

[31] Wilson, R. C., E. Bonawitz, V. D. Costa, and R. B. Ebitz, "Balancing exploration and exploitation with information and randomization," *Current Opinion in Behavioral Sciences*, Vol. 38, No. SI, 49–56, Apr. 2021.