

姓名: 陈弘毅

中学: 上海外国语大学附属外国语学校

省份: 上海

国家/地区: 中国

指导老师: 朱皓斌

指导老师单位: 上海外国语大学附属外国语学校

论文题目: *Expert-Informed Dueling Deep Q-Learning for mBVD Equivalent-Circuit Optimization of BAW Filters*

Expert-Informed Dueling Deep Q-Learning for mBVD Equivalent-Circuit Optimization of BAW Filters

Hongyi Chen

Abstract

The automated design of fixed-topology bulk acoustic wave (BAW) filters is challenging because the design space is high-dimensional, the resonator parameters are strongly coupled, and repeated circuit simulations are costly. This paper presents an automatic parameter-optimization framework that combines a modified Butterworth–Van Dyke (mBVD) equivalent circuit with a dueling deep Q-network (DQN). The mBVD model describes the frequency-dependent impedance of each resonator, while cascaded two-port ABCD matrices provide efficient evaluation of the filter transmission response, $S_{21,\text{dB}}$. Parameter tuning under a fixed ladder topology is formulated as a Markov decision process. The state jointly represents the sampled frequency response, normalized resonator parameters, and a global performance summary. Instead of exploring a 28-dimensional continuous action space directly, the agent selects from an expert-informed discrete action set comprising physically interpretable global adjustments and local completion actions. A deep value network with multiscale residual convolutions, dilated convolutions, attention pooling, and dueling value decomposition extracts both local and global response patterns. Experiments on a ladder filter with four series and four shunt resonators show that the proposed method improves the composite performance score from approximately -33.9 to -3.99 and meets the prescribed passband and stopband limits. Under the evaluated settings, it also outperforms a proximal policy optimization (PPO) baseline that directly controls 28 continuous parameter increments. An ablation study further indicates that explicit global performance statistics improve action-value estimation. These results demonstrate the feasibility of combining fast equivalent-circuit simulation with deep reinforcement learning for BAW-filter parameter optimization.

Keywords—bulk acoustic wave filter, mBVD model, automated circuit design, deep reinforcement learning, dueling DQN, parameter optimization

Contents

Abstract	i
1 Introduction	1
2 Background and Related Work	3
2.1 Reinforcement Learning, DQN, and the Dueling Architecture	3
2.2 RL-Driven Automated Circuit Design	3
2.3 Recent Automated Optimization of RF and Microwave Filters	4
2.4 Equivalent-Circuit Simulation of BAW Filters	5
3 Proposed Method	6
3.1 Problem Formulation	6
3.2 State Representation	6
3.3 Expert-Informed Discrete Action Space	6
3.4 Transition Dynamics and Response Score	7
3.5 Topology-Aware Initialization	7
3.6 Multiscale Dueling DQN	7
4 Experiments and Results	8
4.1 Task and Experimental Setup	8
4.2 Initial Parameter Setting	8
4.3 DQN Optimization Results	9
4.4 Additional Design Specifications and Resonator Counts	9
4.5 Comparison and Ablation Studies	9
5 Conclusion	10
5.1 Main Findings	10
5.2 Limitations and Future Work	11
A Expert-Informed Actions and Their Response Effects	12
A.1 Action-Space Design Principles	12
A.2 Parameter-Update Conventions	12
A.3 Expert Actions	12
A.4 Local Completion Actions	12
A.5 Action-Space Summary	13
B Filter Specifications and Initial Design	13
B.1 Filter Specifications	13
B.2 Default Initial Equivalent-Circuit Parameters	14
B.3 Initial S_{21} Response and Score	14
References	16

1 Introduction

The rapid development of wireless communications, satellite navigation, radar sensing, and the Internet of Things has placed increasingly stringent requirements on radio-frequency (RF) front ends. Within limited spectrum resources, an RF front end must provide precise channel selection and effective interference rejection. Filters are therefore critical components: their frequency responses directly determine insertion loss, out-of-band rejection, frequency selectivity, and ultimately communication quality. Bulk acoustic wave (BAW) filters offer high operating frequencies, high quality factors, small footprints, and favorable integration characteristics. They have consequently become an important filtering technology for mobile communications and high-frequency RF systems. In a ladder topology, staggered resonance characteristics of cascaded series and shunt piezoelectric resonators can be exploited to form the desired passband while maintaining strong out-of-band rejection [1, 2].

Designing a BAW filter is intrinsically a high-dimensional, nonlinear, and strongly coupled optimization problem. For a prescribed topology, the designer must simultaneously determine the resonance frequency, effective electromechanical coupling coefficient, reference impedance, and, where applicable, additional inductive or capacitive compensation for multiple series and shunt resonators. The resulting $S_{21,\text{dB}}$ response must satisfy specifications on passband insertion loss, passband ripple, stop-band attenuation, and transition-band roll-off. Adjusting a single parameter may alter the center frequency, bandwidth, passband flatness, and transmission-zero locations at the same time. The resulting objective landscape is generally nonconvex, multimodal, and locally sensitive, which limits the effectiveness of analytical or one-variable-at-a-time tuning.

Conventional filter-design workflows rely heavily on empirical formulas, manual tuning, or population-based heuristics such as genetic algorithms and particle swarm optimization. Although such methods can handle nonconvex objectives, they often require many repeated simulations and do not readily reuse tuning experience accumulated across optimization trajectories. Recent surrogate-assisted evolutionary and Bayesian methods reduce this

cost by learning an inexpensive approximation to the electromagnetic response, but their performance still depends on sampling quality and surrogate fidelity [3–5]. Their overall cost can therefore rise rapidly as the number of resonators and design variables increases.

Deep reinforcement learning (RL) has recently emerged as a promising approach to automated analog- and RF-circuit design. Unlike one-shot black-box optimization, RL treats parameter tuning as sequential decision making: an agent observes the current circuit, selects a parameter-update action, and continues to refine the design according to the measured performance change. Through repeated interaction with a simulator, the agent can learn a mapping from circuit states to effective tuning directions. AutoCkt, L2DC, and subsequent approaches that incorporate graph neural networks or domain knowledge have demonstrated the potential of RL for analog-circuit sizing [6–10]. More recent studies have extended this idea to microwave-filter geometry, coupling-matrix synthesis, and distributed filtering circuits [11–14].

Current frontier studies frequently formulate filter design in terms of physical geometry, full-wave electromagnetic surrogate models, or coupling matrices [3, 4, 12–14]. These representations are closer to fabrication-oriented practice because they either optimize realizable dimensions, preserve high-fidelity electromagnetic behavior, or connect directly to established filter-synthesis procedures. Their practical advantage, however, comes with a substantial entry barrier. Geometry-based surrogate learning commonly requires a commercial solver to run for long periods and generate hundreds to many thousands of labeled designs before optimization or RL training can begin. In addition to workstation or cluster time, the required licenses, solver modules, and parallel-computing options can be expensive. Publicly available academic and commercial pricing examples indicate that the resulting software budget may range from tens of thousands to several hundred thousand Chinese yuan, depending strongly on the license type, module combination, institutional discount, and computing configuration. Such resources are difficult to obtain for an individual student-led project.

For this reason, the present study adopts the mBVD equivalent circuit together with cascaded ABCD matrices as a deliberately low-cost research abstraction. The

mBVD model retains the dominant resonance, antiresonance, electromechanical-coupling, and loss characteristics of a BAW resonator, while the ABCD formulation provides a transparent and computationally inexpensive way to cascade multiple series and shunt elements and obtain the overall two-port response [1, 15]. This representation is sufficiently expressive to study the central algorithmic question—whether an RL agent can learn effective parameter-tuning strategies for a strongly coupled multi-resonator filter—without first constructing a large full-wave dataset. Its equations also have clear physical and circuit interpretations that can be understood and implemented within a secondary-school research project after focused study. In particular, the relationships among resonance frequency, coupling, impedance, and ladder-network transmission make it possible to reason about parameter effects and design physically interpretable expert actions. The mBVD-ABCD model is therefore used as an appropriate first-stage scientific model rather than as a replacement for subsequent multiphysics, fabrication, and measurement validation.

Directly transferring existing RL methods to BAW-filter design nevertheless raises three difficulties. First, multiple resonators create a high-dimensional continuous action space, and the relationship between parameter increments and changes in the S_{21} response is highly nonlinear. Unguided continuous exploration therefore produces few useful actions. Second, a filter response contains local features, such as resonant peaks, transmission zeros, and ripple, as well as global attributes, such as center frequency, passband width, and stopband attenuation. A single-scale representation may fail to capture both. Third, RL requires a large number of simulator calls; an expensive performance model can make training impractical.

To address these issues, this work develops an equivalent-circuit-driven RL framework for automatic BAW-filter design. The mBVD equivalent circuit and cascaded ABCD matrices form an efficient simulator that maps resonator parameters to $S_{21,\text{dB}}$. The original high-dimensional continuous update space is compressed into a set of physically interpretable compound actions, including center-frequency shifts, bandwidth adjustments, passband-transmission enhancement, and stopband-rejection enhancement. Additional local completion actions cover pa-

rameter directions omitted by the global expert actions. This construction retains essential tuning flexibility while suppressing physically uninformative exploration.

The state combines the sampled frequency response, normalized resonator parameters, and explicit global performance statistics. The response tensor contains normalized frequency, normalized $S_{21,\text{dB}}$, passband margin, stopband margin, and band-importance weight. The global summary reports the current and best scores, specification violations, passband ripple, center-frequency error, and episode progress. A multiscale one-dimensional residual encoder with dilated convolutions and attention pooling extracts the response features. Separate context branches encode the physical parameters and summary statistics, after which a dueling DQN estimates the state value and action advantages [16, 17].

The main contributions are summarized as follows.

1. An equivalent-circuit-driven RL framework integrates the mBVD resonator model, cascaded ABCD-network analysis, and DQN-based parameter tuning into a closed-loop BAW-filter design process.
2. An expert-informed discrete action space converts high-dimensional continuous optimization into a finite set of physically meaningful filter-tuning operations. Local completion actions preserve basic reachability in the parameter space and reduce the exploration burden.
3. A multisource dueling DQN jointly processes multiscale frequency-response features, normalized resonator parameters, and explicit global performance statistics to improve action-value representation.
4. A composite score accounts for mean and worst-case specification violations, the passband-to-stopband power ratio, passband ripple, and center-frequency error. The score increment between consecutive states is used as the immediate reward. In the principal experiment, the score improves from about -33.9 to -3.99 in 300 episodes (approximately 30 min). By comparison, the best score of the 28-dimensional continuous-action PPO baseline remains below approximately -21.5 after 500 episodes, and removing the global performance summary reduces the best DQN score to approximately -5.2 .

The remainder of this paper is organized as follows. Section 2 reviews RL, DQN, the dueling architecture, re-

cent RF- and microwave-filter optimization methods, and equivalent-circuit simulation of BAW filters. Section 3 formulates the optimization problem and describes the state, action, transition, score, initialization, and value-network designs. Section 4 presents the experiments, comparisons, and ablations. Section 5 concludes the paper and discusses future work. The appendices document the complete action set, filter specifications, and initial circuit parameters.

2 Background and Related Work

RL is increasingly used in electronic design automation (EDA) because it does not require an explicit analytical gradient of circuit performance with respect to design variables. Instead, it learns a sequential tuning policy from repeated interactions with a circuit simulator. This property is particularly attractive for the high-dimensional, strongly coupled, nonconvex, and simulation-driven black-box problems encountered in analog and RF design [6–8, 10].

2.1 Reinforcement Learning, DQN, and the Dueling Architecture

The interaction between an agent and its environment is modeled as a Markov decision process (MDP)

$$\mathcal{M} = (\mathcal{S}, \mathcal{A}, P, R, \gamma), \quad (1)$$

where $\mathcal{S}$ and $\mathcal{A}$ denote the state and action spaces, P is the transition kernel, R is the reward function, and $\gamma \in [0, 1]$ is the discount factor. At step t , the agent selects action a_t in state s_t , and the environment returns reward r_t and the next state s_{t+1} . The objective is to learn a policy $\pi(a | s)$ that maximizes the expected discounted return

$$G_t = \sum_{k=0}^{T-t-1} \gamma^k r_{t+k}. \quad (2)$$

The action-value function $Q^\pi(s, a)$ is the expected return obtained by taking action a in state s and subsequently following policy π . The optimal action-value function satisfies the Bellman optimality equation

$$Q^*(s_t, a_t) = \mathbb{E} \left[r_t + \gamma \max_{a'} Q^*(s_{t+1}, a') \right]. \quad (3)$$

The corresponding greedy action is

$$a_t = \arg \max_{a \in \mathcal{A}} Q^*(s_t, a). \quad (4)$$

A deep Q-network (DQN) approximates Q^* by a neural network $Q(s, a; \theta)$ and is well suited to high-dimensional observations with discrete actions [16]. For a transition (s_t, a_t, r_t, s_{t+1}) , the temporal-difference target is

$$y_t = r_t + \gamma \max_{a'} Q(s_{t+1}, a'; \theta^-), \quad (5)$$

where θ^- denotes the target-network parameters. The on-line network is trained by minimizing

$$\mathcal{L}(\theta) = \mathbb{E} \left[(y_t - Q(s_t, a_t; \theta))^2 \right]. \quad (6)$$

Experience replay reduces correlations among training samples, while delayed target-network updates improve stability. An ϵ -greedy behavior policy balances random exploration and greedy action selection.

The dueling architecture decomposes the action value into a state-value term $V(s)$ and an action-advantage term $A(s, a)$ [17]:

$$Q(s, a) = V(s) + A(s, a) - \frac{1}{|\mathcal{A}|} \sum_{a' \in \mathcal{A}} A(s, a'). \quad (7)$$

Mean-centering the advantages removes the identifiability ambiguity between V and A . The network can therefore learn the overall quality of a state separately from the relative merit of each action, which is beneficial when several small tuning actions have similar values.

2.2 RL-Driven Automated Circuit Design

A typical RL-based circuit-design loop begins with a target specification spec and an initial parameter vector θ_0 obtained from expert knowledge, an existing design, or random sampling. A simulator returns

$$y_0 = \text{sim}(\theta_0), \quad (8)$$

from which the initial state is constructed as

$$s_0 = (\theta_0, y_0, \text{spec}). \quad (9)$$

At step t , the policy proposes an update action

$$a_t \sim \pi_\phi(\cdot | s_t), \quad (10)$$

and an update operator $\mathcal{U}$ modifies the circuit:

$$\theta_{t+1} = \mathcal{U}(\theta_t, a_t). \quad (11)$$

The simulator then evaluates

$$y_{t+1} = \text{sim}(\theta_{t+1}), \quad (12)$$

and the environment returns

$$r_t = R(s_t, a_t, s_{t+1}). \quad (13)$$

An episode terminates when the specification is met, simulation fails, or the maximum number of design steps is reached. The resulting trajectory is

$$\tau = \{(s_t, a_t, r_t, d_t, s_{t+1})\}_{t=0}^{T-1}, \quad (14)$$

where d_t indicates termination. DQN or a discrete actor-critic method can be used for discrete tuning actions, whereas DDPG, TD3, SAC, or continuous-action PPO is more natural for continuous parameter increments.

This formulation turns static parameter search into a sequential process that can accumulate tuning experience. When the target specification is included in the state and training spans a distribution of specifications, the policy may further become goal-conditioned and transfer knowledge across related tasks. AutoCkt, GCN-RL, and later goal-conditioned methods illustrate this direction [7, 9, 18]. The present work focuses on fixed-topology optimization for one prescribed specification and does not claim cross-specification generalization.

2.3 Recent Automated Optimization of RF and Microwave Filters

Recent filter-design studies can be broadly divided into surrogate-assisted black-box optimization, neural inverse or forward modeling, and RL-based sequential optimization. In the first category, Xue et al. proposed H-SMEAFO, which combines response-feature objectives, hybrid surrogate models, and evolutionary search to perform unsupervised three-dimensional optimization of high-order waveguide filters [3]. Belen and Belen compared several regression models for an edge-coupled sub-5G bandpass filter and used the resulting data-driven surrogate within a Bayesian-optimization loop [4]. More recently, Lindemans et al. introduced a physics-informed Gaussian process with a complex-valued Szegő kernel. By modeling the spectral S-parameters rather than only a scalar score, their Bayesian optimizer can evaluate different response-based objectives from the same surrogate [5]. These methods illustrate how data-efficient global search can reduce the number of expensive full-wave evaluations.

RL-based approaches instead learn how design parameters should be changed as a function of the current response. Dai et al. proposed a relational induction neural network that discretizes geometric-parameter actions and applies deep RL to second- and sixth-order microwave filters [11]. Ye et al. trained an artificial-neural-network surrogate for a hairpin filter and then used continuous-action PPO to optimize its geometric dimensions; the trained agent typically reached the prescribed goal within 5–15 design steps in the learned environment [12]. Fan et al. applied an Actor-Critic PPO formulation to microwave-filter coupling-matrix synthesis and designed task-specific action and reward functions for the resulting combinatorial problem [13]. Gao et al. formulated distributed-filter layout design as sequential geometric modification and incorporated initialization, state evaluation, and invalid-action penalties to prevent overlapping resonators [14].

Neural models can also accelerate a conventional optimizer without replacing it with a learned policy. Zhang et al. combined an LSTM-feedforward network with polynomial transfer-function approximation and a trust-region method, avoiding repeated retraining of the neuro-transfer-function model during low-pass-filter optimization [19]. Luo et al. used a multilayer-perceptron predictor to initialize particle-swarm optimization and retained CST simulations for final correction of a hairpin bandpass filter [20]. Such hybrid methods preserve a direct connection to high-fidelity simulation, whereas RL can reuse sequential tuning experience after training.

Most of these studies address microstrip, waveguide, cavity, or distributed filters and optimize geometric variables or coupling matrices. In contrast, the present work targets a ladder BAW filter and optimizes resonator-level mBVD parameters under a fixed topology. Its expert-informed discrete actions occupy a middle ground between unrestricted continuous increments and a fully prescribed design rule: they encode common coordinated adjustments while leaving the action sequence to the dueling DQN. The fast mBVD-ABCD simulator also permits direct interaction during training, avoiding the additional approximation error introduced by a learned electromagnetic surrogate.

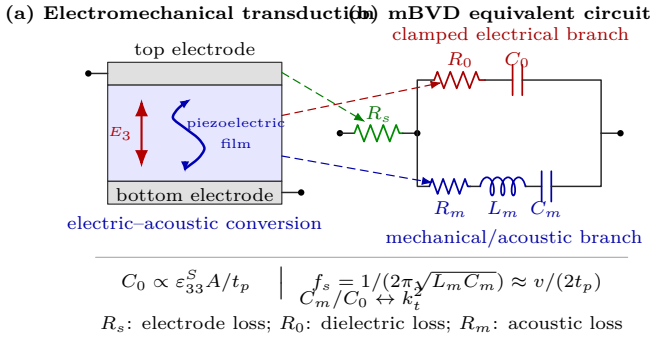


Figure 1: Simplified physical-to-circuit interpretation of a BAW resonator. The electrode–piezoelectric stack performs electromechanical transduction and supports longitudinal thickness oscillation. In the mBVD model, (C_0, R_0) form the static electrical branch, (L_m, C_m, R_m) form the motional branch, and R_s represents electrode and lead loss.

2.4 Equivalent-Circuit Simulation of BAW Filters

2.4.1 The mBVD Model

A BAW resonator exploits electromechanical coupling in a piezoelectric material to convert electrical energy into acoustic energy and vice versa. Near resonance, strong thickness-mode vibration produces characteristic resonance and antiresonance behavior. In a ladder network, appropriately staggered series and shunt resonators provide low-loss in-band transmission and attenuate out-of-band signals through the combined effects of series impedance and shunt bypass paths [1, 2].

Figure 1 highlights the two relations needed for the circuit-level model. First, the electrode–piezoelectric–electrode stack converts the applied electric field into longitudinal thickness vibration and converts the resulting acoustic motion back into an electrical response. Second, this physical behavior maps to two parallel mBVD branches: C_0 and R_0 describe the clamped capacitance and dielectric loss, whereas L_m , C_m , and R_m represent effective acoustic mass, compliance, and damping. The series resistance R_s accounts primarily for electrode and lead loss. In approximate physical terms, $C_0 \propto \epsilon A / t_p$, $f_s = 1 / (2\pi \sqrt{L_m C_m}) \approx v / (2t_p)$, and the ratio C_m / C_0 is governed by the electromechanical coupling coefficient k_t^2 .

Finite-element models can accurately resolve internal acoustic fields but are expensive for iterative circuit-

level optimization. The modified Butterworth–Van Dyke (mBVD) model is therefore commonly used as an efficient equivalent circuit [15]. At angular frequency ω , the impedance used by the simulator is

$$Z_{\text{mBVD}}(\omega) = R_s + \frac{Z_{\text{static}}(\omega) Z_{\text{motional}}(\omega)}{Z_{\text{static}}(\omega) + Z_{\text{motional}}(\omega)}, \quad (15)$$

where

$$Z_{\text{static}} = R_0 + \frac{1}{j\omega C_0},$$

$$Z_{\text{motional}} = R_m + j\omega L_m + \frac{1}{j\omega C_m}. \quad (16)$$

This representation separates the ordinary electrical capacitance from the electromechanically induced resonance while retaining the principal loss mechanisms.

2.4.2 Two-Port Networks and ABCD Matrices

A ladder filter can be decomposed into cascaded two-port elements, avoiding repeated solution of every internal node voltage and current. This work represents each two-port network by an ABCD (transmission) matrix. With the output-port current defined as positive when leaving the network,

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} V_2 \\ -I_2 \end{bmatrix}, \quad (17)$$

where A , B , C , and D are frequency-dependent transmission parameters.

For a series element with impedance $Z(\omega)$,

$$\mathbf{M}_{\text{series}}(\omega) = \begin{bmatrix} 1 & Z(\omega) \\ 0 & 1 \end{bmatrix}. \quad (18)$$

For a shunt element with admittance $Y(\omega)$,

$$\mathbf{M}_{\text{shunt}}(\omega) = \begin{bmatrix} 1 & 0 \\ Y(\omega) & 1 \end{bmatrix}, \quad Y(\omega) = \frac{1}{Z(\omega)}. \quad (19)$$

Substituting $Z_{\text{mBVD}}(\omega)$ into the appropriate matrix yields the two-port representation of each series or shunt resonator.

If N two-port elements appear in propagation order as $\mathbf{M}_1, \dots, \mathbf{M}_N$, the overall transmission matrix is

$$\mathbf{M}_{\text{total}}(\omega) = \prod_{i=1}^N \mathbf{M}_i(\omega) = \begin{bmatrix} A(\omega) & B(\omega) \\ C(\omega) & D(\omega) \end{bmatrix}. \quad (20)$$

Because matrix multiplication is noncommutative, the multiplication order must match the physical cascade.

With equal real reference impedances Z_0 at both ports and a matched output port, the forward transmission coefficient is

$$S_{21}(\omega) = \frac{2}{A(\omega) + B(\omega)/Z_0 + C(\omega)Z_0 + D(\omega)}. \quad (21)$$

For the conventional RF reference impedance $Z_0 = 50 \Omega$,

$$S_{21,\text{dB}}(\omega) = 20 \log_{10} |S_{21}(\omega)|. \quad (22)$$

This formulation provides a uniform and scalable simulator for the repeated frequency-response evaluations required during parameter optimization.

3 Proposed Method

3.1 Problem Formulation

Consider a ladder BAW filter containing N_{se} series resonators and N_{sh} shunt resonators. Each series resonator is parameterized by its series-resonance frequency, effective electromechanical coupling coefficient, and impedance-scaling parameter,

$$\theta_{\text{se}}^{(i)} = (f_{s,i}, k_{t,i}^2, Z_i), \quad (23)$$

whereas each shunt resonator has an additional compensation parameter h_j ,

$$\theta_{\text{sh}}^{(j)} = (f_{s,j}, k_{t,j}^2, Z_j, h_j). \quad (24)$$

The complete design vector is therefore

$$\theta = [\theta_{\text{se}}^{(1)}, \dots, \theta_{\text{se}}^{(N_{\text{se}})}, \theta_{\text{sh}}^{(1)}, \dots, \theta_{\text{sh}}^{(N_{\text{sh}})}], \quad (25)$$

with dimensionality

$$d_\theta = 3N_{\text{se}} + 4N_{\text{sh}}. \quad (26)$$

For example, a 4s4p topology has $d_\theta = 28$ design variables.

Given passband and stopband specifications, the objective is to find a feasible parameter vector that maximizes a scalar response score:

$$\theta^* = \arg \max_{\theta \in \Theta} \Phi(\theta), \quad (27)$$

where Θ denotes the physically admissible parameter domain and Φ is evaluated from the simulated S_{21} response. We cast this black-box optimization problem as a Markov decision process. At each step, the agent observes the current response and design context, applies a structured parameter update, and receives the improvement in Φ as its reward.

3.2 State Representation

The state combines three complementary information sources:

$$s_t = (\mathbf{X}_{\text{curve},t}, \mathbf{x}_{\text{param},t}, \mathbf{x}_{\text{summary},t}). \quad (28)$$

The curve tensor $\mathbf{X}_{\text{curve},t} \in \mathbb{R}^{5 \times 512}$ is sampled on a fixed normalized frequency grid. Its five channels contain normalized frequency, S_{21} in decibels, passband margin, stopband margin, and band-importance weight. The two margin channels explicitly encode the distance to the corresponding design limit, allowing one network to interpret different band specifications without reconstructing those relationships from raw frequency and transmission data alone.

The normalized parameter vector $\mathbf{x}_{\text{param},t}$ preserves the current structural configuration. The global performance summary is a ten-dimensional vector,

$$\mathbf{x}_{\text{summary},t} = [\bar{\Phi}_t, \bar{\Phi}_t^{\text{best}}, \Delta \bar{\Phi}_t, e_{\text{pb}}^{\text{max}}, e_{\text{sb}}^{\text{max}}, \mu_{\text{pb}}, \mu_{\text{sb}}, r_{\text{pb}}, e_c, \tau_t], \quad (29)$$

where the entries denote the normalized current score, best score in the current episode, latest score change, maximum passband and stopband violations, mean passband and stopband levels, passband ripple, center-frequency error, and normalized episode progress, respectively. This compact summary exposes the global optimization status while the curve branch retains local spectral detail. Its contribution is examined in Section 4.5.

3.3 Expert-Informed Discrete Action Space

Directly controlling every parameter independently leads to a high-dimensional action space and weakens exploration efficiency. We instead use a discrete library of coordinated parameter updates. The full action set is

$$\mathcal{A} = \{a_0, a_1, \dots, a_{15}\}. \quad (30)$$

Actions a_0 – a_{10} encode common filter-tuning operations, including center-frequency translation, bandwidth adjustment, passband-transmission enhancement, stopband-rejection enhancement, ripple smoothing, and upper-edge correction. Actions a_{11} – a_{15} are smaller local completion

actions that perturb selected resonator groups after the response approaches feasibility. Appendix A gives the exact update rules.

This design embeds domain knowledge without prescribing a fixed optimization trajectory. The agent still learns when and how often to apply each operation, while the coordinated updates reduce the effective search complexity and preserve physically meaningful parameter relationships.

3.4 Transition Dynamics and Response Score

After action a_t is selected, the environment updates the parameter vector subject to its physical bounds, evaluates the mBVD-ABCD simulator, and constructs the next state. Let $y_k(\theta)$ be the simulated $S_{21,\text{dB}}$ value at frequency sample k . Passband and stopband violations are defined against their respective masks and limits. The aggregate score used in this work is

$$\Phi(\theta) = -[\mathcal{L}_{\text{viol}} + \mathcal{L}_{\text{perf}} + 0.05\mathcal{L}_{\text{shape}} + 20\exp(e_c) - 20], \quad (31)$$

where $\mathcal{L}_{\text{viol}}$ penalizes specification violations, $\mathcal{L}_{\text{perf}}$ measures passband and stopband performance, $\mathcal{L}_{\text{shape}}$ regularizes undesirable response shape, and e_c is the normalized center-frequency error. The exponential center-frequency term strongly discourages solutions that satisfy local masks at the wrong spectral location.

The immediate reward is the score improvement,

$$r_t = \Phi(\theta_{t+1}) - \Phi(\theta_t), \quad (32)$$

which aligns the cumulative return with progressive improvement while avoiding a large task-dependent offset in the reward scale.

3.5 Topology-Aware Initialization

The initial parameters are generated deterministically from the target passbands. Let f_{base} be the mean of their center frequencies and let ξ_i be evenly spaced in $[0.998, 1.002]$. Series- and shunt-resonance frequencies are initialized as

$$f_{s,i}^{\text{se}} = f_{\text{base}}\xi_i, \quad f_{s,j}^{\text{sh}} = 0.98f_{\text{base}}\xi_j. \quad (33)$$

All resonators initially use $k_t^2 = 0.06$ and $Z = 50\Omega$, and all shunt compensation parameters are set to $h = 0$. This

Algorithm 1 Dueling-DQN training for BAW-filter optimization

Require: Simulator $\mathcal{M}$, action library $\mathcal{A}$, online network

Q_ϕ , target network Q_{ϕ^-}

- 1: Initialize replay buffer $\mathcal{D}$ and set $\phi^- \leftarrow \phi$
- 2: **for** episode = 1, ..., E **do**
- 3: Initialize θ_0 using Eq. (33) and construct s_0
- 4: **for** $t = 0, \dots, T-1$ **do**
- 5: Select a_t using an ϵ -greedy policy derived from Q_ϕ
- 6: Apply a_t , enforce bounds, and simulate the next response
- 7: Compute r_t using Eq. (32) and construct s_{t+1}
- 8: Store (s_t, a_t, r_t, s_{t+1}) in $\mathcal{D}$
- 9: Sample a minibatch from $\mathcal{D}$ and update ϕ
- 10: Periodically update $\phi^- \leftarrow \phi$
- 11: **end for**
- 12: **end for**

initialization places the resonances near the target operating region while retaining enough frequency staggering to avoid a fully symmetric starting point.

3.6 Multiscale Dueling DQN

Figure 2 summarizes the proposed network. The response branch consists of five multiscale residual convolutional blocks. Each block uses parallel one-dimensional kernels of sizes $\{3, 7, 15, 31\}$, together with dilation factors 2 and 4, to capture narrow edge behavior and broad inter-band trends simultaneously. Attention pooling converts the final feature map into a 128-dimensional curve embedding.

The parameter and global-summary vectors are processed by fully connected layers to form a second 128-dimensional context embedding. Concatenating both branches produces a 256-dimensional shared representation. Dueling heads then estimate the state value and action advantages, and combine them according to Eq. (7). This architecture separates the general quality of a filter state from the relative usefulness of each tuning operation.

Algorithm 1 outlines the training procedure. A replay buffer is used to break temporal correlations, and a target network stabilizes the bootstrapped value targets.

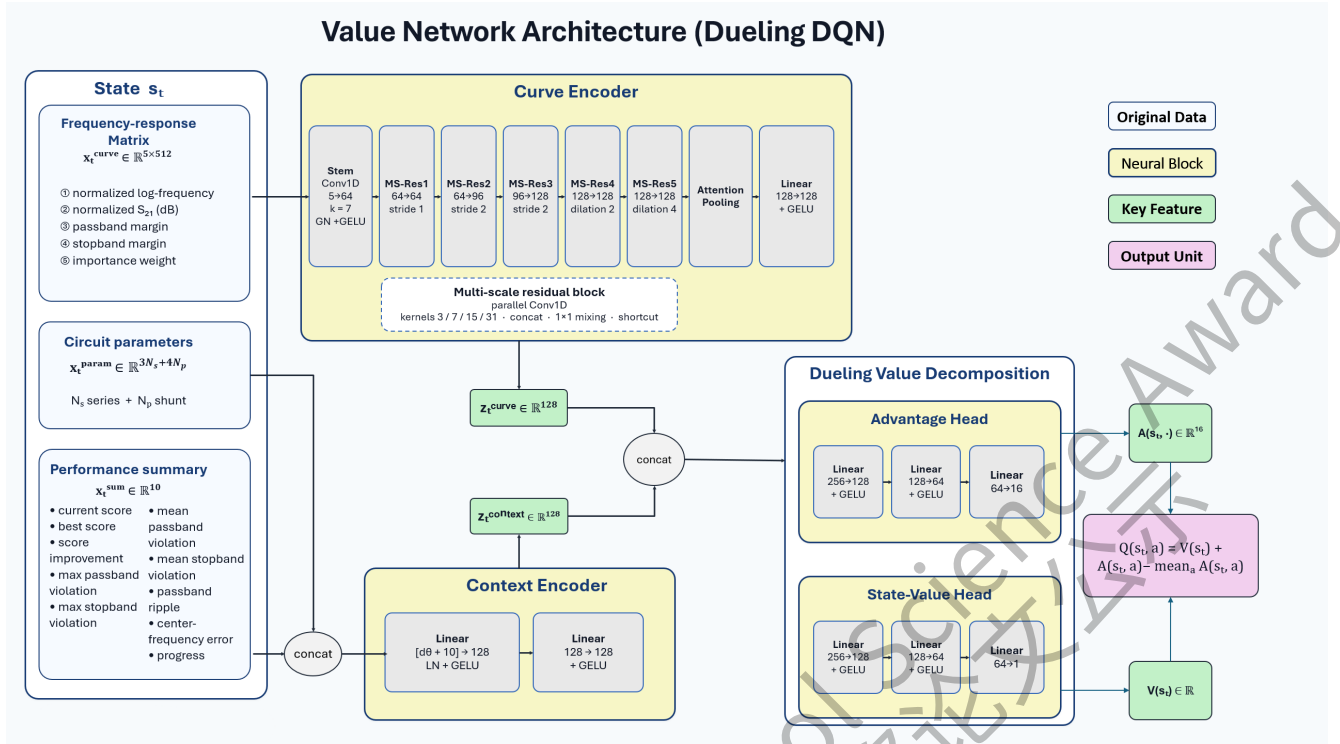


Figure 2: Multiscale dueling-DQN architecture. The curve branch extracts local and long-range spectral features, while the context branch encodes the current structure and global optimization status.

4 Experiments and Results

This section evaluates the proposed method on fixed-topology ladder BAW-filter design. The complete frequency-band specifications, default parameters, and initial response are provided in Appendix B. We first describe the task and training configuration, then present the DQN results, and finally examine the contributions of the dueling architecture, the expert-informed action space, and the global performance summary.

4.1 Task and Experimental Setup

The design targets are specified by passband and stopband tuples. The i th passband is

$$\mathcal{B}_i^p = (f_{i,\text{lo}}^p, f_{i,\text{hi}}^p, l_i^p), \quad (34)$$

where l_i^p is the minimum permissible $S_{21,\text{dB}}$ value. Likewise, the j th stopband is

$$\mathcal{B}_j^s = (f_{j,\text{lo}}^s, f_{j,\text{hi}}^s, l_j^s), \quad (35)$$

where l_j^s is the maximum permissible $S_{21,\text{dB}}$ value.

The principal experiment uses a 4s4p ladder filter. Each series resonator has three design variables, whereas

each shunt resonator has four; hence the design dimension is

$$d_\theta = 4 \times 3 + 4 \times 4 = 28. \quad (36)$$

The simulator spans the frequency range required by the masks in Appendix B. The action space contains 16 discrete actions, each episode contains at most 500 interaction steps, and training runs for 300 episodes. Unless otherwise stated, all methods use the same simulator, score definition, parameter bounds, initialization, and target specifications. A score closer to zero indicates better agreement with the design requirements.

4.2 Initial Parameter Setting

For a task with N_{PB} passbands, the center of passband i is

$$f_{c,i} = \frac{f_{i,\text{lo}}^p + f_{i,\text{hi}}^p}{2}, \quad (37)$$

and the reference frequency is the mean of all passband centers,

$$f_{\text{base}} = \frac{1}{N_{\text{PB}}} \sum_{i=1}^{N_{\text{PB}}} f_{c,i}. \quad (38)$$

The initial series- and shunt-resonator parameters are

$$\theta_j^{\text{se},0} = (f_{\text{base}}\xi_j^{\text{se}}, 0.06, 50\ \Omega), \quad (39)$$

$$\theta_j^{\text{sh},0} = (0.98f_{\text{base}}\xi_j^{\text{sh}}, 0.06, 50\ \Omega, 0), \quad (40)$$

where both ξ_j^{se} and ξ_j^{sh} are linearly distributed over $[0.998, 1.002]$. This setting places the initial response near the target band, establishes the frequency offset required by a ladder topology, and breaks exact symmetry through a $\pm 0.2\%$ frequency spread. The resulting 4s4p design has a score of approximately -33.9 and still violates several passband and stopband constraints. Exact values are given in Appendix B.

4.3 DQN Optimization Results

The best dueling-DQN design achieves a score of -3.99 , compared with approximately -33.9 for the initial structure. Training takes about 30 minutes on the experimental platform. Figure 3 compares the corresponding transmission responses. The optimized curve satisfies the prescribed masks while preserving a smooth passband and strong out-of-band rejection.

Figure 4 shows the best response score obtained in each episode. The rapid early improvement indicates that the expert-informed actions efficiently correct gross frequency and bandwidth errors. Later episodes focus on smaller edge, ripple, and rejection adjustments. This coarse-to-fine behavior is consistent with the intended division between the expert actions and the local completion actions.

4.4 Additional Design Specifications and Resonator Counts

Table 1 reports results for five ladder topologies and three modified passband specifications. For the topology study, the parameter dimension grows from 21 for 3s3p to 35 for 5s5p. The proposed method consistently reaches a score near -4.0 despite substantial differences in initial score and search-space dimension. It also remains effective when the passband is widened, narrowed, or shifted upward in frequency.

For the baseline specification, the best scores remain near -3.99 as the topology grows from 3s3p to 5s5p. The method also remains effective when the passband width or

center frequency changes. The narrower-passband task is slightly more difficult, most likely because passband ripple, band-edge placement, and stopband attenuation become more tightly coupled. These values are representative best results for the corresponding settings; repeated trials with multiple random seeds would be needed for a statistical comparison of training stability.

4.5 Comparison and Ablation Studies

Table 2 summarizes the complete method, the continuous-action PPO baseline, and the two ablation settings.

4.5.1 Removing the Dueling Architecture

To evaluate the contribution of the dueling decomposition, we trained a conventional DQN with the same inputs and action space. Its best score is -4.8 , and the corresponding curve in Fig. 4 exhibits larger oscillations and does not remain as close to the optimum. This behavior suggests that jointly learning absolute action values makes it harder to distinguish poor states from poor actions. Separating state quality from relative action advantage improves value discrimination and stabilizes learning.

4.5.2 Continuous-Action PPO Baseline

The PPO policy directly outputs 28 continuous parameter increments, modeled as independent Gaussian variables. After 500 episodes, its best score remains no better than approximately -21.5 , substantially below the -3.99 obtained by the proposed DQN. Under the present settings, random exploration in the raw high-dimensional parameter space is inefficient because the parameter effects are nonlinear and strongly coupled. The expert actions instead combine parameter changes into physically interpretable operations, thereby reducing ineffective exploration. This task-specific comparison should not be interpreted as a universal ranking of DQN and PPO.

4.5.3 Ablation of the Global Performance Summary

In this experiment, the current score, violation statistics, passband ripple, center-frequency error, and optimization progress are removed from the state, while the re-

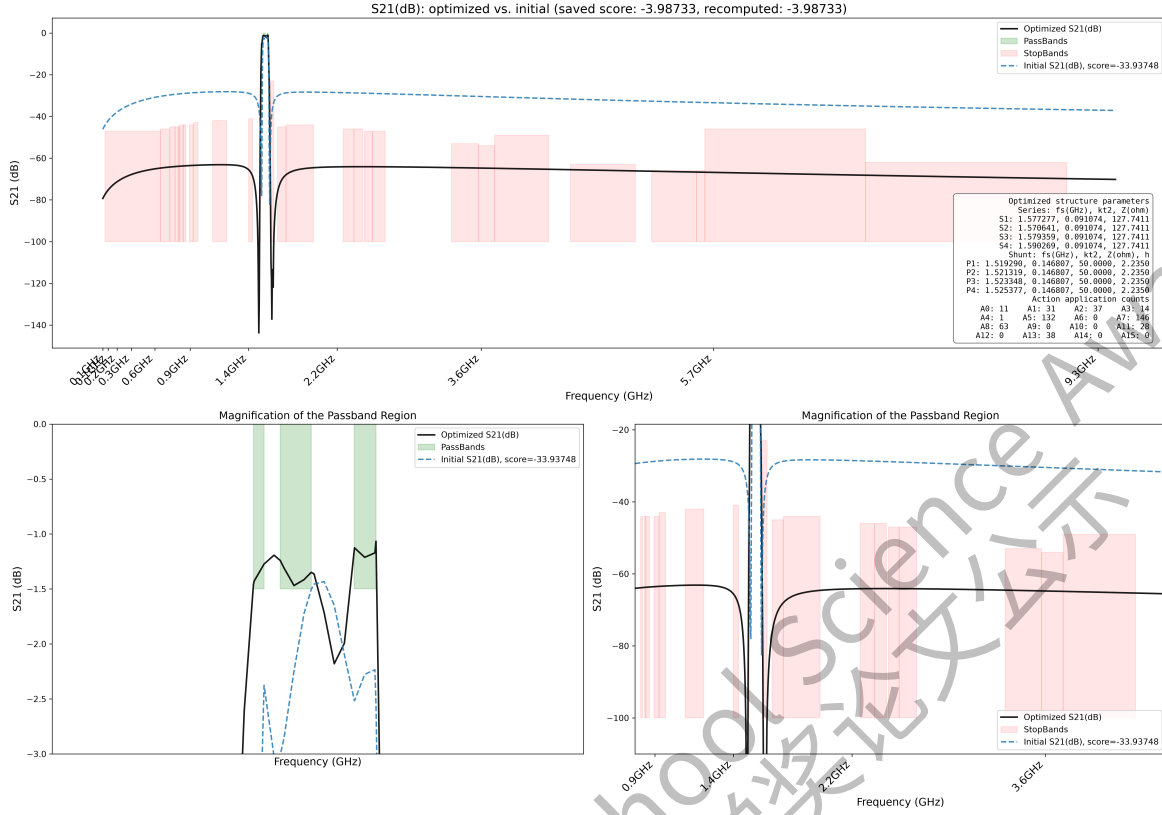


Figure 3: Simulated S_{21} responses of the initial and optimized 4s4p filters. The dashed curve denotes the initial design, and the solid curve denotes the best design found by dueling DQN.

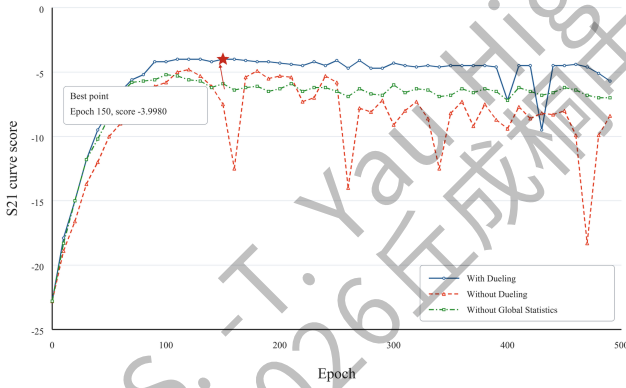


Figure 4: Episode-wise learning curve of the proposed dueling-DQN optimizer.

sponse curve and resonator parameters are retained. The best score decreases from -3.99 to approximately -5.2 . Although the omitted quantities can in principle be inferred from the full response, asking the convolutional encoder to reconstruct cross-band maxima, means, variations, and center position makes action-value learning more difficult. Explicit summary features therefore help the agent iden-

tify the dominant violation and the current optimization stage.

5 Conclusion

5.1 Main Findings

This paper presented an automated optimization framework for ladder BAW filters that combines a fast mBVD-ABCD simulator with an expert-informed dueling-DQN agent. The method represents each design through its full transmission response, structural parameters, and global optimization status, while a library of coordinated tuning actions reduces the difficulty of high-dimensional exploration. On the main 4s4p task, the score improves from approximately -33.9 to -3.99 , and the optimized response satisfies the specified passband and stopband limits. Experiments across multiple topologies and modified specifications show that the formulation remains effective as the parameter dimension and target bands change. The PPO comparison and ablation studies further indicate that

Table 1: Optimization results across filter topologies and design specifications.

Passband specification	N_{se}	N_{sh}	Dimension	Change from default	Initial score	Best score
Default passband	3	3	21	None	-33.3800	-3.9910
	4	3	24	None	-36.0011	-3.9918
	4	4	28	None	-33.9097	-3.9908
	5	4	31	None	-31.9720	-3.9931
	5	5	35	None	-29.6696	-3.9925
Wider passband	4	4	28	Width 120%	-42.1618	-4.0523
Narrower passband	4	4	28	Width 80%	-18.5345	-4.3532
Shifted passband	4	4	28	Center +30%	-34.2733	-3.9941

Table 2: Comparison and ablation results on the 4s4p task.

Method	Episodes	Best score
Proposed dueling DQN	300	-3.99
Continuous-action PPO (28-D)	500	-21.5
Dueling DQN without global summary	300	-5.2
DQN without dueling heads	300	-4.8

level results. A multi-fidelity workflow combining fast mBVD search with finite-element simulation, tolerance analysis, and measured data would improve physical fidelity and engineering transferability.

structured actions, the global performance summary, and the dueling architecture all improve optimization efficiency in the evaluated setting.

5.2 Limitations and Future Work

The current expert actions are derived mainly from engineering experience. Their effectiveness therefore depends on the topology and parameter range, and some actions are rarely used along the best trajectories. Future work may prune actions according to visit frequency and measured contribution, adapt their step sizes, or learn useful action combinations automatically.

The experiments cover several topologies and modified passband specifications, but they do not yet constitute a systematic study of generalization across random initial states and task distributions. Multi-task training and repeated evaluations with multiple random seeds should be used to report means, standard deviations, and confidence intervals.

Finally, the mBVD model omits some higher-order acoustic modes, process variation, and layout parasitics. The reported designs therefore remain equivalent-circuit-

A Expert-Informed Actions and Their Response Effects

A.1 Action-Space Design Principles

The design variables of a ladder BAW filter include the resonance frequency, effective electromechanical coupling coefficient, reference impedance, and compensation reactance of its series and shunt resonators. If every variable is adjusted independently and continuously, the action dimension grows with topology size and many random parameter combinations have little physical meaning. We therefore encode common tuning experience as

$$\mathcal{A} = \{a_0, a_1, \dots, a_{15}\}. \quad (41)$$

Actions a_0 – a_{10} are expert actions for center frequency, bandwidth, transmission, rejection, roll-off, and ripple; actions a_{11} – a_{15} provide local directions not covered by the global actions. Every action is a deterministic update $\theta_{t+1} = \mathcal{U}(\theta_t, a_t)$ followed by projection onto the feasible parameter domain. Because the parameter-to-response mapping is nonlinear, the effects described below are dominant trends near the baseline design rather than globally monotonic guarantees.

A.2 Parameter-Update Conventions

For a shunt resonator, the compensation parameter is mapped to an external reactive element as

$$\begin{aligned} h > 0 : \quad C &= h \times 10^{-12} \text{ F}, \\ h < 0 : \quad L &= (-h) \times 10^{-8} \text{ H}. \end{aligned} \quad (42)$$

Thus, increasing h adds capacitance or weakens existing inductive compensation, whereas decreasing h removes capacitance or strengthens inductive compensation. The notation $x \leftarrow (1 + \delta)x$ denotes a relative increase of $\delta \times 100\%$. Unless a subset is specified, an action is applied simultaneously to every resonator of the indicated type.

A.3 Expert Actions

A.3.1 Center-Frequency and Bandwidth Adjustment

Action a_0 multiplies all series- and shunt-resonance frequencies by 1.0003, translating the passband upward while

preserving their relative offsets. Action a_1 multiplies the same frequencies by 0.9997 and therefore shifts the passband downward. Action a_2 multiplies the series frequencies by 1.0003 and the shunt frequencies by 0.9997, increasing the series–shunt frequency separation and generally widening the low-loss region.

For action a_3 , define the mean frequency separation

$$\Delta \bar{f} = \frac{1}{N_{\text{se}}} \sum_i f_{s,i}^{\text{se}} - \frac{1}{N_{\text{sh}}} \sum_j f_{s,j}^{\text{sh}}. \quad (43)$$

The series frequencies are reduced by $0.005\Delta \bar{f}$ and the shunt frequencies are increased by the same amount, narrowing the passband around its center.

A.3.2 Passband Transmission and Stopband Rejection

Action a_4 applies $Z^{\text{se}} \leftarrow 0.991Z^{\text{se}}$ and $Z^{\text{sh}} \leftarrow 1.009Z^{\text{sh}}$, modifying the balance between the main path and shunt paths to reduce passband insertion loss. Action a_5 applies $k_t^{2,\text{sh}} \leftarrow 1.01k_t^{2,\text{sh}}$ and $h \leftarrow h + 0.005$, strengthening shunt-path rejection. Action a_6 reduces k_t^2 by 1% for every shunt resonator except the last, primarily correcting the upper band edge. Action a_7 increases series k_t^2 by 0.2% and decreases shunt k_t^2 by 0.2% to improve transmission toward the upper side of the passband.

A.3.3 Passband Shape and Baseline Restoration

Action a_8 jointly applies $k_t^{2,\text{se}} \leftarrow 1.01k_t^{2,\text{se}}$, $k_t^{2,\text{sh}} \leftarrow 0.99k_t^{2,\text{sh}}$, $Z^{\text{se}} \leftarrow 1.015Z^{\text{se}}$, and $h \leftarrow h + 0.025$ to smooth passband ripple. Action a_9 moves all resonance frequencies by 10% of their distance toward the target center and moves series impedances by 10% of their distance toward 50Ω ; it therefore serves as a restoring action when other updates move the design away from a useful region. Finally, a_{10} applies $k_t^{2,\text{se}} \leftarrow 1.006k_t^{2,\text{se}}$ to extend or improve the upper side of the passband.

A.4 Local Completion Actions

The expert actions mainly control global response properties. Five additional actions improve local reachability: a_{11} and a_{12} increase the first two series- and shunt-resonance frequencies, respectively, by 0.02%; a_{13} and a_{14} apply the same increase to the middle series and shunt

resonators; and a_{15} applies $h \leftarrow h - 0.001$. The last action supplies the compensation direction opposite to the h updates in a_5 and a_8 .

A.5 Action-Space Summary

The complete action library is summarized below. Let Δf denote the difference between the mean series- and shunt-resonance frequencies, and let f_{target} denote the target passband center.

Table 3: Expert actions and their principal purposes.

Action	Parameter update	Primary purpose
a_0	All $f_s \times 1.0003$	Shift center upward
a_1	All $f_s \times 0.9997$	Shift center downward
a_2	Series $f_s \times 1.0003$; shunt $f_s \times 0.9997$	Widen passband
a_3	Reduce mean series-shunt frequency separation	Narrow passband
a_4	Series $Z \times 0.991$; shunt $Z \times 1.009$	Improve transmission
a_5	Shunt $k_t^2 \times 1.01$; $h + 0.005$	Improve rejection
a_6	Shunt $k_t^2 \times 0.99$, except final stage	Correct upper roll-off
a_7	Series $k_t^2 \times 1.002$; shunt $k_t^2 \times 0.998$	Improve upper passband
a_8	Jointly update k_t^2 , series Z , and shunt h	Smooth ripple
a_9	Move f_s and series Z toward their baseline values	Restore baseline
a_{10}	Series $k_t^2 \times 1.006$	Extend upper passband

Table 4: Local completion actions and their principal purposes.

Action	Parameter update	Primary purpose
a_{11}	First two series $f_s \times 1.0002$	Local series correction
a_{12}	First two shunt $f_s \times 1.0002$	Local shunt correction
a_{13}	Middle series $f_s \times 1.0002$	Mid-ladder series search
a_{14}	Middle shunt $f_s \times 1.0002$	Mid-ladder shunt search
a_{15}	$h \leftarrow h - 0.001$	Reverse reactance adjustment

Figures 5–7 show the dominant response trends when each expert action is repeatedly applied to the same baseline design. Fifty repetitions are used only to make the change readily visible.

B Filter Specifications and Initial Design

B.1 Filter Specifications

The simulator covers 80–9435.3 MHz using 2000 frequency samples. Table 5 lists the three passbands, all of which use a -1.5 dB lower limit. Table 6 gives the stopband masks.

B.1.1 Passband Specifications

Table 5: Passband specifications.

Band	Start (MHz)	Stop (MHz)	Limit (dB)
P1	1569.42	1581.42	-1.5
P2	1598.06	1606.50	-1.5
P3	1559.05	1563.15	-1.5

B.1.2 Stopband Specifications

Table 6: Stopband specifications.

Band	Start (MHz)	Stop (MHz)	Limit (dB)
S1	100.0	612.0	-47
S2	612.0	698.0	-46
S3	699.0	748.0	-45
S4	748.0	777.0	-45
S5	777.0	787.0	-45
S6	788.0	803.0	-44
S7	815.0	824.0	-44
S8	824.0	849.0	-44
S9	880.0	915.0	-44
S10	915.0	960.0	-43
S11	1094.4	1224.0	-42
S12	1427.0	1463.0	-41
S13	1526.0	1536.0	-44
S14	1627.0	1660.0	-23
S15	1695.0	1772.0	-45
S16	1772.0	2025.0	-44
S17	2300.0	2400.0	-46
S18	2401.0	2482.0	-46
S19	2496.0	2570.0	-47
S20	2570.0	2690.0	-47
S21	3300.0	3550.0	-53
S22	3550.0	3700.0	-54
S23	3700.0	4200.0	-49
S24	4400.0	5000.0	-63
S25	5150.0	5566.0	-66
S26	5566.0	5641.0	-66
S27	5641.0	7125.0	-46
S28	7125.0	8986.0	-62

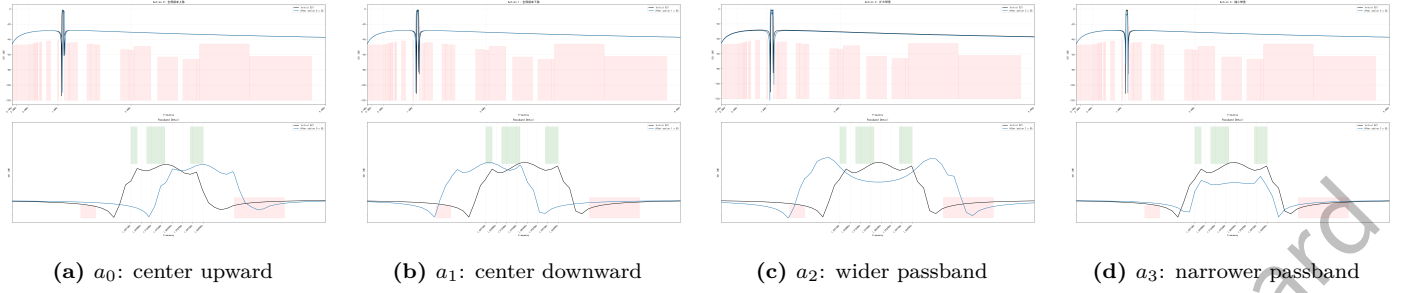


Figure 5: Frequency-response effects of the center-frequency and passband-width actions. Each action is repeated 50 times to magnify the trend.

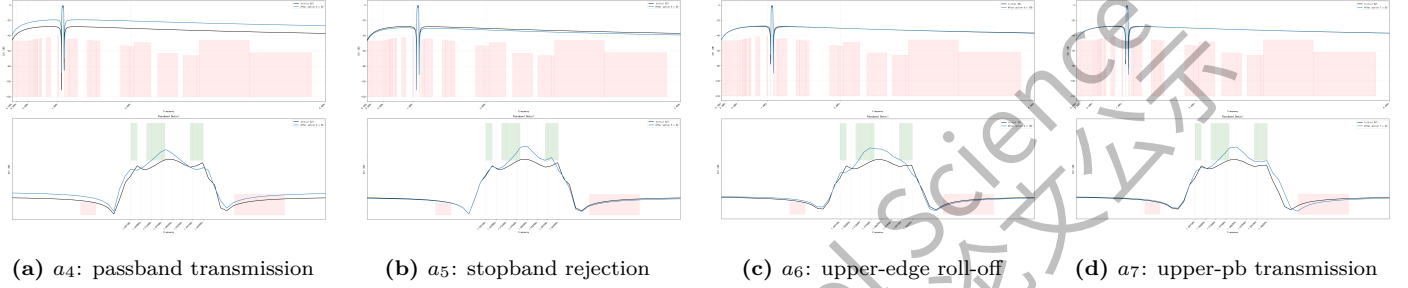


Figure 6: Frequency-response effects of the passband-transmission and stopband-rejection actions. Each action is repeated 50 times to magnify the trend.

B.2 Default Initial Equivalent-Circuit Parameters

The default topology contains four series and four shunt resonators. The mean of the three passband centers gives $f_{\text{base}} = 1579.6$ MHz. Series-resonance frequencies are spread around this value, whereas shunt-resonance frequencies are spread around $0.98f_{\text{base}}$. All initial coupling coefficients are 0.06, all reference impedances are 50Ω , and every initial shunt compensation parameter is zero.

B.2.1 Series-Resonator Parameters

Table 7: Default initial series-resonator parameters.

Resonator	f_s (MHz)	k_t^2	Z (Ω)
Series 1	1576.4408	0.06	50.0
Series 2	1578.5469	0.06	50.0
Series 3	1580.6531	0.06	50.0
Series 4	1582.7592	0.06	50.0

B.2.2 Shunt-Resonator Parameters

Table 8: Default initial shunt-resonator parameters.

Resonator	f_s (MHz)	k_t^2	Z (Ω)	h
Shunt 1	1544.9120	0.06	50.0	0.0
Shunt 2	1546.9760	0.06	50.0	0.0
Shunt 3	1549.0400	0.06	50.0	0.0
Shunt 4	1551.1040	0.06	50.0	0.0

B.3 Initial S_{21} Response and Score

With these specifications and default parameters, the response is simulated at 2000 frequency samples. The initial $S_{21,\text{dB}}$ curve is the dashed blue line in Fig. 3. The score is the negative of a composite loss formed from passband and stopband violations, the worst-case constraint error, the passband-to-stopband power ratio, passband variation, and center-frequency error. Consequently, scores are generally negative and values closer to zero indicate better agreement with the specifications. The default structure has

$$\Phi_{\text{initial}} = -33.937478. \quad (44)$$

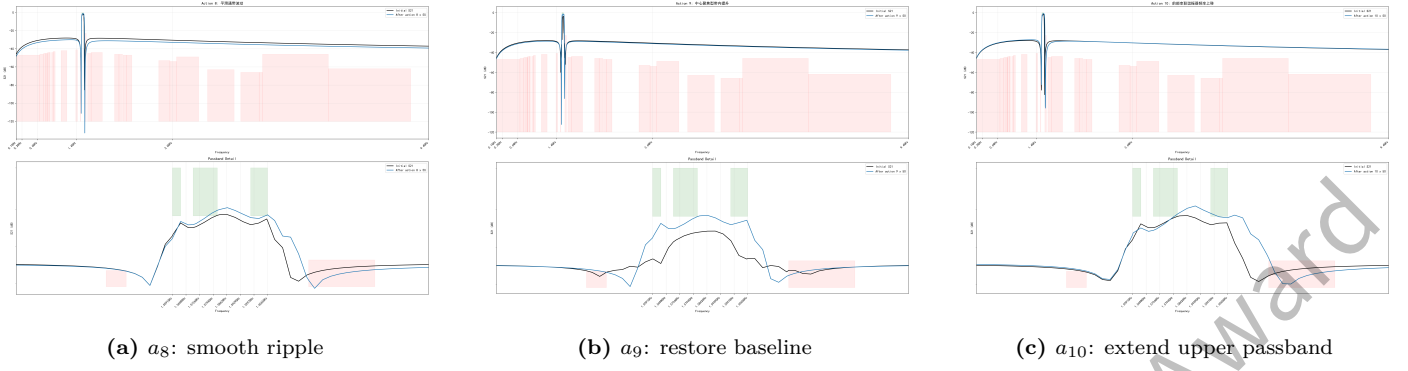


Figure 7: Frequency-response effects of the passband-shape and baseline-restoration actions. Each action is repeated 50 times to magnify the trend.

This value confirms that the initial response does not yet meet the target masks; the role of reinforcement learning is to increase the score and progressively satisfy the passband and stopband constraints.

References

- [1] Alfred Giménez, Jordi Verdú, and Pedro de Paco. General synthesis methodology for the design of acoustic wave ladder filters and duplexers. *IEEE Access*, 6:47969–47979, 2018.
- [2] Sung-Yuan Tseng et al. Synthesis of chebyshev/elliptic filters using minimum acoustic wave resonators. *IEEE Transactions on Microwave Theory and Techniques*, 2019.
- [3] Liyuan Xue, Bo Liu, Yang Yu, Qingsha S. Cheng, Muhammad Imran, and Tianrui Qiao. An unsupervised microwave filter design optimization method based on a hybrid surrogate model-assisted evolutionary algorithm. *IEEE Transactions on Microwave Theory and Techniques*, 71(3):1159–1170, 2023.
- [4] Aysu Belen and Mehmet A. Belen. Data-driven modeling of band-pass filter for sub-5g applications. *Microwave and Optical Technology Letters*, 65(8):2210–2216, 2023.
- [5] Yens Lindemans, Thijs Ullrick, Ivo Couckuyt, Dirk Deschrijver, and Tom Dhaene. Bayesian optimization of microwave filters: A physics-informed approach using the szego kernel. In *2025 IEEE 29th Workshop on Signal and Power Integrity (SPI)*, pages 1–4. IEEE, 2025.
- [6] Hanrui Wang, Kuan Wang, Jiacheng Yang, Linxiao Shen, Nan Sun, Hae-Seung Lee, and Song Han. Learning to design circuits. *arXiv preprint arXiv:1812.02734*, 2018.
- [7] Keertana Settaluri, Ameer Haj-Ali, Qijing Huang, Kourosh Hakhamaneshi, and Borivoje Nikolic. AutoCkt: Deep reinforcement learning of analog circuit designs. In *2020 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, pages 490–495. IEEE, 2020.
- [8] Zhe Zhao and Lihong Zhang. Deep reinforcement learning for analog circuit sizing. In *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5. IEEE, 2020.
- [9] Hanrui Wang, Kuan Wang, Jiacheng Yang, Linxiao Shen, Nan Sun, Hae-Seung Lee, and Song Han. GCN-RL circuit designer: Transferable transistor sizing with graph neural networks and reinforcement learning. In *Proceedings of the 57th ACM/IEEE Design Automation Conference (DAC)*, pages 1–6, 2020.
- [10] Weidong Cao, Mouhacine Benosman, Xuan Zhang, and Rui Ma. Domain knowledge-infused deep learning for automated analog/radio-frequency circuit parameter optimization. In *Proceedings of the 59th ACM/IEEE Design Automation Conference (DAC)*, pages 1015–1020, 2022.
- [11] Xi Wang Dai, Wen Hao Hu, Yang Hui Fu, Jie Liu, and Zhi Xi Chen. Automatic design of microwave filters by using relational induction neural networks. *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, 36(6):e3102, 2023.
- [12] Yunong Ye, Yifan Wu, Jiayu Chen, Guodong Su, Junchao Wang, and Jun Liu. Intelligent design of hairpin filters based on artificial neural network and proximal policy optimization. *Applied Sciences*, 13(16):9379, 2023.
- [13] Dongdong Fan, Shuai Ding, Haotian Zhang, Weihao Zhang, Qingsong Jia, Xu Han, Hao Tang, Zhaojun Zhu, and Yuliang Zhou. A novel proximal policy optimization approach for filter design. *Applied Computational Electromagnetics Society Journal*, 39(5):390–395, 2024.
- [14] Peng Gao, Tao Yu, Fei Wang, and Ru-Yue Yuan. Automated design and optimization of distributed filter circuits using reinforcement learning. *Journal of Computational Design and Engineering*, 11(5):60–76, 2024.
- [15] John D. Larson, Paul D. Bradley, Stephen Wartenberg, and Richard C. Ruby. Modified butterworth-van dyke circuit for fbar resonators and automated measurement system. In *Proceedings of the 2000 IEEE Ultrasonics Symposium*, volume 1, pages 863–868. IEEE, 2000.
- [16] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland,

Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.

- [17] Ziyu Wang, Tom Schaul, Matteo Hessel, Hado van Hasselt, Marc Lanctot, and Nando de Freitas. Dueling network architectures for deep reinforcement learning. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1995–2003, 2016.
- [18] Seunggeun Kim, Ziyi Wang, Sungyoung Lee, Youngmin Oh, Hanqing Zhu, Doyun Kim, and David Z. Pan. PPAAS: PVT and pareto aware analog sizing via goal-conditioned reinforcement learning. *arXiv preprint arXiv:2507.17003*, 2025.
- [19] Xin Zhang, Jian Wu, Yong-Qiang Chai, Shui Liu, and Yuan Peng. EM optimization of microwave filter based on long short-term memory-feedforward neural network and transfer functions. *Journal of Physics D: Applied Physics*, 57(48):485001, 2024.
- [20] Sha Luo, Jiteng Ma, Shuping Dang, and Andrew C. M. Austin. Deep learning-enabled rapid optimization for microwave filter design. In *2024 Asia-Pacific Microwave Conference (APMC)*, pages 1287–1289. IEEE, 2025.

Acknowledgments and Author Contribution Statement

This project investigated whether reinforcement learning can be used to automate the optimization of bulk acoustic wave filters. The project was completed independently by the author as an individual student research project. My responsibilities included defining the research topic, conducting the literature review, selecting and implementing the simulation model, formulating the optimization problem, designing the reinforcement-learning method, conducting and iterating the experiments, analyzing the results, and drafting and organizing the manuscript. More specifically, I independently constructed the state space, designed the expert-informed action set, scoring function, and reward function, selected and improved the reinforcement-learning algorithm, and summarized the methodology, experimental results, and conclusions.

Mr. Haobin Zhu is a subject teacher at my school and served as my school-based research supervisor for this project. He provided theoretical instruction in the operating principles of RF components, equivalent-circuit modeling, and reinforcement learning. He also offered research guidance by discussing the feasibility of the proposed methods, identifying potential problems in the experimental design, recommending relevant theoretical directions, and providing feedback on the organization and academic writing of the manuscript. His role was advisory: the selection of the research topic, acquisition of the experimental data, design and improvement of the algorithm, implementation and iteration of the experiments, analysis of the results, literature review, methodological development, and substantive preparation of the manuscript were completed independently by the author.

During the early stage of the project, I represented every parameter adjustment as a continuous action. Mr. Zhu helped me recognize that the resulting action space was excessively large and could prevent the neural network from converging reliably. He encouraged me to examine practical RF-circuit design cases and identify commonly used tuning operations. Based on this guidance, I independently summarized these operations and constructed an expert-informed discrete action set. This modification substantially reduced the action-space size, limited unproductive exploration, and improved the agent’s ability to identify effective filter designs.

In subsequent experiments, I observed that conventional DQN training remained unstable and that the circuit score fluctuated considerably. Mr. Zhu introduced the general principles and typical applications of the Dueling network architecture. After studying the relevant theory and literature, I independently incorporated the Dueling architecture into the value network and evaluated its effect experimentally. The resulting model exhibited substantially improved training stability. I am deeply grateful to Mr. Zhu for his theoretical instruction, research guidance, and constructive feedback on the manuscript.

All experimental data reported in this study were generated and collected by the author within the equivalent-circuit simulator developed for this project. No external experimental dataset, proprietary measurement dataset, or third-party training dataset was used. The reported results were obtained entirely from simulated filter responses and reinforcement-learning trajectories produced during the author’s experiments.

The project developed according to the following timeline. In January 2026, I read a news report about the use of artificial intelligence to automate the design of drug molecules. I later discussed this technological development with Mr. Zhu outside class. He explained that artificial intelligence could also be applied to the design of engineering systems, including circuits used in mobile communication devices. This discussion inspired me to explore AI-assisted circuit optimization.

From January to March 2026, I studied the role of RF filters in mobile communication systems and learned how the number, parameters, and topology of circuit components affect filter performance. I also learned that equivalent-circuit models provide a computationally efficient way to evaluate filter responses under different design configurations. In April, I studied the modified Butterworth–Van Dyke (mBVD) model and the ABCD-matrix cascade method for two-port networks.

In May, I conducted preliminary experiments using a continuous action space and the proximal policy optimization algorithm, but the results were unsatisfactory. Following Mr. Zhu's theoretical feedback concerning the difficulty of learning in an excessively unconstrained action space, I reviewed relevant circuit-design literature and independently developed the initial expert-informed action set in June. After receiving his suggestion that complementary operations were needed to preserve adequate exploration, I revised and completed the action-set design. By July, the resulting agent could identify promising filter configurations, although its training score continued to fluctuate. Mr. Zhu then recommended that I study the Dueling architecture as a possible way to improve value estimation. I subsequently completed the literature review, implemented the architecture, conducted the corresponding experiments, and analyzed the results. By August 2026, I had completed nearly all the experimental and manuscript preparation work for this project.

Statement on the Use of Artificial Intelligence

The artificial-intelligence tool used in this project was Grok Fast. Its use was limited to the preliminary learning stage of the project, during which it was consulted approximately once a week as an auxiliary learning tool to help locate, explain, and understand prerequisite knowledge. The topics studied with its assistance included the operating principles of RF filters, equivalent-circuit modeling, the fundamentals of reinforcement learning, and related background concepts.

Grok Fast was not used to determine the research topic, construct the theoretical or methodological framework, design the expert-informed action set, develop or improve the algorithm, write the experimental code, generate or collect experimental data, conduct the experiments, analyze the results, formulate the conclusions, or prepare the substantive content of the manuscript. All these core components were completed independently by the author under the academic guidance of the supervisor.

CV for Author

Hongyi Chen

Grade 10, Shanghai Foreign Language School

Achievements:

- Acceptance of paper "HalluciGuard: A Protocol-Centered Verification System for Educational AI" into The 2026 ACM International Joint Conference on Pervasive and Ubiquitous Computing / ISWC ACM International Symposium on Wearable Computers
- Member of the Inaugural Multilingual Science & Technology Innovation Talent Experimental Class at Shanghai Foreign Language School
- Third Prize in 2023 Shanghai Youth "Technology Without Limits" for independent research "Research on the Symmetry Axes of Regular Polygons"
- Selected for Fudan University's First "Into Mathematics" Summer Camp in 2022
- Two-time Recipient of the School "Weihong Foundation Award" (2022 and 2023 Academic Years)
- ESDP 2023 Champion in Shanghai Speech Final
- ESDP 2023 National Final "Future Star Award" with Supreme Awards in Speech, Debate & Interview and First Prizes in Writing & Hosting
- ESDP 2023 Second Prize in Shanghai Writing Final
- ESDP 2023 Supreme Award in Shanghai Debate Final
- CCF National Informatics CSP-J 2023 First Prize
- CCF National Informatics CSP-S 2024 Second Prize

- CCF LMCC Level 2 Certification (Youth Division)
- Advanced to Round 2 of the Second CCF LMCC Certification (September 2026) with a Round 1 Score of 94/100
- USACO Silver Division Contestant
- Selected for CUHK AI Preparatory Course and completed an image recognition project using OpenCV and robotic arm for ball classification
- Became a member of HKU Academy for the Talented and completed Flipbot Race Lab and 2026 Summer Program “The Mathematical Foundations Behind AI”