Name:	Hongyi Chen
School:	Shanghai Foreign Language School Affiliated to Shanghai International Studies University
Province:	Shanghai
Country/Region:	China
Supervisor:	Haobin Zhu
Supervisor's Affiliation:	Shanghai Foreign Language School Affiliated to Shanghai International Studies University
Paper Title:	*Expert-Informed Dueling Deep Q-Learning for mBVD Equivalent-Circuit Optimization of BAW Filters*

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

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

Hongyi Chen

Abstract

To address the high parameter dimensionality, complex coupling, and substantial repeated-simulation costs associated with the design of bulk acoustic wave (BAW) filters with a fixed ladder topology, this paper proposes an automated parameter optimization method based on the modified Butterworth–Van Dyke (mBVD) equivalent model and Dueling DQN. First, the mBVD model describes the frequency-dependent impedance of each resonator, and cascaded two-port ABCD matrices enable rapid calculation of the filter's $S_{21,\mathrm{dB}}$ response. Second, parameter tuning under a fixed topology is formulated as a Markov decision process. Frequency-response curves, resonator parameters, and a global performance summary are jointly used to construct the state, while high-dimensional continuous parameter adjustments are compressed into a discrete action space comprising expert actions and local complementary actions. To extract local and global patterns from the frequency response and address the strong dependence of the performance score on the state itself, a deep value network is constructed using multiscale residual convolutions, dilated convolutions, attention pooling, and dueling value decomposition. Experimental results show that the proposed method improves the performance score from approximately -33.9 to -3.99, with the optimized frequency response satisfying the prescribed passband and stopband constraints. Under the current experimental settings, it outperforms a PPO baseline that directly uses high-dimensional continuous actions. Ablation results further indicate that the global performance summary and the dueling architecture improve action-value assessment. This study demonstrates that combining equivalent-circuit simulation with reinforcement learning enables rapid, low-cost parameter optimization of BAW filters.

Keywords: bulk acoustic wave filters; mBVD model; automated circuit design; deep reinforcement learning; Dueling DQN; parameter optimization

2026 S. -T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

Contents

Abstract i

1 Introduction 1

2 Theoretical Background and Related Work 2
- 2.1 Reinforcement Learning, DQN, and the Dueling Architecture . . . 2
- 2.2 Reinforcement Learning-Driven Automated Circuit Design . . . 3
- 2.3 Fast Simulator Design . . . 3

3 Equivalent-Circuit Simulation System for BAW Filters 4
- 3.1 The mBVD Model . . . 4
- 3.2 Two-Port Networks and ABCD Matrices . . . 4

4 Learning Algorithm for Automated BAW Filter Design 5
- 4.1 Problem Formulation . . . 5
- 4.2 State Space Design . . . 5
- 4.3 Action Space Design . . . 6
- 4.4 State Transitions and Performance Evaluation . . . 7
- 4.5 Parameter Initialization . . . 7
- 4.6 Multiscale Dueling DQN Network . . . 7

5 Experiments and Results 8
- 5.1 Experimental Task and Settings . . . 8
- 5.2 DQN Optimization Results . . . 9
- 5.3 Additional Design Specifications and Resonator Counts . . . 9
- 5.4 Comparison and Ablation Analysis . . . 9

6 Conclusions and Outlook 11
- 6.1 Conclusions . . . 11
- 6.2 Limitations and Future Work . . . 12

A Expert Action Space Design and Response Effects 13
- A.1 Action Space Design Principles . . . 13
- A.2 Parameter-Update Conventions . . . 13
- A.3 Expert Actions . . . 13
- A.4 Local Complementary Actions . . . 17
- A.5 Summary of the Action Space . . . 18

B Experimental Filter Specifications and Initial Response 18
- B.1 Filter Specifications . . . 18
- B.2 Default Initial Equivalent-Circuit Parameters . . . 18
- B.3 Initial S_{21} Curve and Score . . . 20

References 21

Acknowledgments and Author Contributions 22

Statement on the Use of Artificial Intelligence 23

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

1 Introduction

Artificial intelligence (AI)-assisted engineering design is a rapidly developing area of research and application, with important advances in protein structure research and drug molecule design. Automated radio-frequency (RF) device design is another promising direction, and parameter optimization of bulk acoustic wave (BAW) filters is a representative problem. As a core component of the RF front end, a BAW filter provides channel selection and interference suppression within limited spectral resources, and its performance directly affects signal transmission quality in communication systems. Such filters typically comprise multiple BAW resonators, each of which converts electrical energy into acoustic energy and vice versa through electromechanical coupling. Different parameters jointly affect the resonance frequencies and frequency response, giving rise to nonlinear and nonconvex parameter optimization. The coupling of dozens of parameters within a single system increases the design difficulty, and repeated parameter adjustments are typically required to meet the target specifications.

BAW filters have a relatively well-established theoretical foundation, and commercial simulation software can analyze the frequency responses of different structures and parameter settings using relevant physical models. A common design workflow combines optimization methods, such as genetic algorithms and particle swarm optimization, with repeated simulator calls to evaluate and select candidate designs. However, this workflow may incur substantial software licensing and computational costs, particularly when individual simulations are time-consuming and numerous iterative evaluations are required. Moreover, conventional optimization workflows often have difficulty directly transferring previous optimization experience to new design tasks, so changes in the target specifications generally require a new parameter search.

To reduce the costs of performance evaluation and parameter search, recent studies have introduced data-driven modeling and reinforcement learning into circuit and filter design. Surrogate modeling methods use parameter combinations and their corresponding frequency responses obtained from simulations or measurements to learn the mapping between design parameters and performance, thereby reducing the number of expensive simulations required during optimization. Reinforcement learning instead learns parameter-adjustment policies through interaction with a design environment and explores the possibility of applying previous optimization experience to different design objectives. These two approaches can also be combined: a surrogate model provides rapid performance evaluation, while a reinforcement learning agent learns a sequential optimization policy. Related studies investigate reinforcement learning-based circuit optimization, data-driven modeling, and sample-efficient optimization from different perspectives [1–5], but do not all follow the same technical approach.

For design workflows that rely on data-driven surrogate models, obtaining sufficiently representative and reliable training data is an important prerequisite. Generating such data generally requires simulation or measurement resources, making data acquisition a potentially substantial barrier to research. Under the conditions of this study, no public dataset directly applicable to modeling the target BAW filters was obtained. Furthermore, the mapping from physical parameters, such as geometric dimensions and material properties, to the frequency response involves complex multiphysics coupling. Even when the simulation model has a clear physical basis, translating engineers' tuning experience directly into effective learning actions in such parameter spaces remains challenging. Without appropriate action constraints, reinforcement learning may also suffer from inefficient exploration.

To systematically analyze BAW filter performance, a resonator can be represented by an equivalent circuit consisting of resistors, inductors, and capacitors, allowing its electromechanically coupled resonant behavior to be expressed as a circuit response. This representation facilitates analysis of how parameter changes affect the frequency response and enables engineering tuning experience to be expressed as parameter-adjustment operations with clear circuit-level meaning. Incorporating such experience into reinforcement learning action design may therefore reduce exploration of parameter-adjustment combinations that lack physical meaning, narrow the range of actions available at each step, and thereby improve learning efficiency and optimization performance.

Motivated by these considerations, this study uses equivalent-circuit models of BAW resonators as its basic building blocks, connects multiple resonators in a ladder topology, and calculates the overall filter S_{21} curve through cascaded two-port ABCD matrices. A performance score is then computed according to the target specifications, yielding a low-cost BAW filter simulation environment. Building on this environment and engineering tuning experience, the continuous adjustment of dozens of parameters is represented as a sequential decision process with 16 discrete actions to reduce the exploration difficulty of reinforcement learning. Experimental results show that the resulting agent can obtain designs satisfying the prescribed specifications within a relatively short training time. Due to limitations in commercial simulation software access and data availability, direct performance comparisons with existing methods under a common high-fidelity simulation or measurement framework have not been conducted. Rather than seeking a filter with globally optimal performance, this study explores a computationally inexpensive approach to BAW filter parameter optimization that does not depend on commercial full-wave simulation software.

The main contributions of this paper are summarized as follows:

1. An equivalent-circuit-model-driven reinforcement learning framework for BAW filter design is proposed. It in-

2026 SCIENCE PAPER ONLINE 录用于2026年在线科学论文 Science Paper Online

tegrates the mBVD model, cascaded ABCD networks, and DQN-based parameter optimization into a unified closed loop for rapid frequency-response calculation and iterative optimization.
2. An expert-knowledge-guided discrete composite action space is constructed, transforming high-dimensional continuous parameter optimization into a finite set of tuning operations with physical meaning for filters. Local complementary actions maintain basic reachability in the parameter space, thereby reducing the exploration difficulty of reinforcement learning.
3. A multisource-feature Dueling DQN network is designed for filter frequency responses. Multiscale residual convolutions, dilated convolutions, and attention pooling extract curve features, which are fused with resonator parameters and a global performance summary to enhance action-value representation.

The remainder of this paper is organized as follows. Section 2 introduces the theoretical foundations, including reinforcement learning, DQN, the dueling architecture, and equivalent-circuit modeling of BAW filters. Section 3 describes the BAW filter performance simulation system. Section 4 presents the automated filter design agent based on the simulation system and reinforcement learning, including the problem definition, state space, action space, state transitions, performance scoring, parameter initialization, and value network. Section 5 presents the experimental settings, optimization results, comparisons, and ablation analyses. Section 6 concludes the paper and discusses future research directions. The appendices provide details of the action space, experimental specifications, and initial parameters.

2 Theoretical Background and Related Work

With the rapid development of deep reinforcement learning for high-dimensional decision-making, continuous control, and combinatorial optimization, introducing reinforcement learning (RL) into electronic design automation (EDA) has become an important research direction. Unlike conventional rule-based or gradient-based automated circuit design methods, reinforcement learning does not require explicit analytical gradients of circuit performance with respect to the design variables. Instead, it repeatedly interacts with a circuit simulation environment and learns a sequential decision policy from design trajectories to determine how the circuit should be modified in its current state. This property makes reinforcement learning particularly suitable for the high-dimensional, strongly coupled, nonconvex, and simulation-driven black-box optimization problems commonly encountered in analog and RF circuit design [1, 4, 6, 7].

2.1 Reinforcement Learning, DQN, and the Dueling Architecture

Reinforcement learning models the interaction between an agent and its environment as a Markov decision process (MDP),

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

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

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

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

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

Thus, once Q^* is obtained, the optimal action in the current state can be selected as

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

A deep Q-network (DQN) approximates the optimal action-value function using a neural network $Q(s, a; \theta)$ parameterized by θ, making it suitable for decision problems with high-dimensional states and discrete actions [8]. For an experience sample (s_t, a_t, r_t, s_{t+1}), DQN computes the temporal-difference target using a target network:

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

The value network parameters are updated by minimizing

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

where θ^- denotes the target network parameters. Experience replay reduces correlations between samples by randomly sampling historical interaction data, while the target network improves training stability through delayed updates. Training typically uses an ϵ-greedy policy to balance random exploration against selection of the currently best action.

Standard DQN directly estimates the Q value of each action, whereas the dueling architecture decomposes it into a state-value function $V(s)$ and an advantage function $A(s, a)$ [9]:

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

2025 S.T. Yau High School Science Award
收稿于2026年成稿中学科学奖公示

Here, $V(s)$ measures the overall value of the current state, while $A(s,a)$ describes the value of each action relative to the average. Mean-centering the advantage function removes the non-uniqueness of the decomposition into $V(s)$ and $A(s,a)$. This architecture enables the network to learn state quality and relative differences among actions separately, making it suitable for parameter optimization problems in which several discrete actions have similar values but the overall performance differs substantially across states.

2.2 Reinforcement Learning-Driven Automated Circuit Design

A typical reinforcement learning-based circuit design process can be expressed as the following closed loop. First, given the target specifications spec, initial parameters are obtained from expert experience, an existing design, or random sampling and denoted by θ_0. The simulator is then invoked to obtain

$$y_0 = \mathrm{sim}(\theta_0), \tag{8}$$

from which the initial state is constructed as

$$s_0 = (\theta_0, y_0, \mathrm{spec})\,. \tag{9}$$

At decision step t, the policy network generates a parameter-adjustment action based on the current state:

$$a_t \sim \pi_\phi(\cdot \mid s_t). \tag{10}$$

The environment updates the circuit according to the action mapping $\mathcal{U}$:

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

The performance simulation model is then invoked to obtain

$$y_{t+1} = \mathrm{sim}(\theta_{t+1}), \tag{12}$$

and the reward is computed as

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

The environment also checks whether the target specifications have been met, a simulation failure has occurred, or the maximum number of design steps has been reached. This yields a design trajectory

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

During training, the interaction data in Eq. (14) are used to update the policy parameters. DQN or discrete-policy actor–critic methods can be used for discrete parameter tuning, whereas DDPG, TD3, SAC, or continuous-action PPO are more natural choices for continuous parameter optimization. This closed loop transforms a one-off static parameter search into a sequential optimization process that accumulates interaction experience. When target specifications are included in the state and the policy is trained over a task distribution comprising multiple specifications, reinforcement learning may further yield a goal-conditioned optimization policy and transfer tuning experience across related tasks. AutoCkt, GCN-RL, and subsequent goal-conditioned reinforcement learning studies exemplify this direction [1, 2, 10]. This paper focuses on fixed-topology parameter optimization under a single prescribed specification and does not further discuss cross-specification generalization.

2.3 Fast Simulator Design

During automated optimization, the optimization algorithm typically calls a simulator repeatedly to evaluate device performance under different design parameters and update the design accordingly. However, high-accuracy simulations based on physics-based models, such as finite element analysis, are often computationally expensive, limiting search efficiency. Researchers have therefore explored neural-network-based surrogate models that learn the mapping between design parameters and device performance and replace some of the expensive simulations required during optimization.

Simon et al. [11] trained neural networks using one- and two-dimensional modeling data for FBAR resonators to construct a digital twin for resonance-frequency prediction, replacing repeated device simulations and accelerating resonator parameter optimization. Almalkawi and Caron [12] used wafer-level resonator measurements to train a multilayer perceptron (MLP) that maps resonator area, series resonance frequency, and sweep frequency to the complex reflection coefficient. Circuit cascading was then used to construct a rapid response model for ladder filters, supporting filter performance analysis and RF module co-design. He et al. [13] trained a three-dimensional FBAR neural-network surrogate model using 5000 random electrode shapes and their finite-element-simulated frequency responses. The model enables rapid frequency-response prediction for varying electrode geometries, with a prediction speed approximately 10 times that of conventional finite element methods, providing an efficient tool for resonator geometry optimization. Zhao et al. [14] further represented BAW filter circuit topologies as graphs and developed a filter performance surrogate model using a graph attention network (GAT). Combined with a hybrid NSGA-II and PSO optimization algorithm, this approach enables rapid filter parameter optimization under multiple objectives and constraints.

Although these methods no longer rely on simulation software during subsequent optimization and enable rapid performance estimation, training their surrogate models depends on large volumes of sample data generated by simulation software. The overall methodology therefore remains dependent on such software. Given the limitations of the available research resources, this study constructs a fast theoretical simulation model based on the fundamental equivalent-circuit theory of BAW filters, neglecting

2026 J. Yau High School Science Award 应用于2026立成科学奖公示

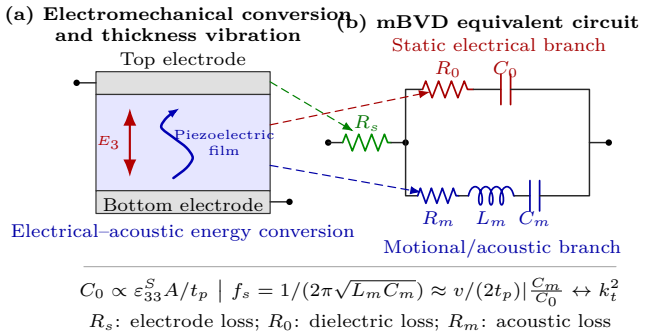


Figure 1: Physical structure of a BAW resonator and its equivalent-circuit representation. The electrode–piezoelectric material stack enables electromechanical conversion and supports longitudinal thickness vibration. In the mBVD model, (C_0, R_0) form the static electrical branch, representing static capacitance and dielectric loss, respectively; (L_m, C_m, R_m) form the motional branch, representing the equivalent inertia, compliance, and acoustic loss of mechanical vibration, respectively; and R_s represents electrode and interconnect losses.

higher-order factors with relatively minor effects.

3 Equivalent-Circuit Simulation System for BAW Filters

3.1 The mBVD Model

Bulk acoustic wave (BAW) resonators exploit electromechanical coupling in piezoelectric materials to convert electrical energy into acoustic energy and vice versa. When the excitation frequency approaches the resonance frequency, pronounced thickness-direction acoustic vibrations occur in the piezoelectric layer, producing resonance and antiresonance characteristics. Cascading multiple BAW resonators in a ladder topology and appropriately configuring the resonance and antiresonance frequencies of the series and shunt resonators allow low-loss signal transmission within the passband, while series impedance and shunt bypass paths suppress out-of-band signals, thereby producing a bandpass response [15, 16].

Figure 1 summarizes the two levels of correspondence required by the circuit-level model used in this study. First, the electrode–piezoelectric material–electrode stack undergoes longitudinal thickness vibration under an applied electric field and converts the acoustic motion back into an electrical response. Second, this physical process corresponds to two parallel branches in the mBVD model: C_0 and R_0 describe the clamped capacitance and dielectric loss; L_m, C_m, and R_m describe the equivalent acoustic mass, compliance, and damping, respectively; and the series resistance R_s primarily represents electrode and interconnect losses. Approximately, $C_0 \propto \varepsilon A/t_p$, $f_s = 1/(2\pi\sqrt{L_m C_m}) \approx v/(2t_p)$, and C_m/C_0 is related to the electromechanical coupling coefficient k_t^2.

Finite element models can describe the internal acoustic field of a resonator with relatively high accuracy, but at substantial computational cost. To improve the efficiency of circuit-level analysis and optimization, this study adopts the modified Butterworth–Van Dyke (mBVD) equivalent model for BAW resonators [17]. At angular frequency ω, the equivalent impedance used by the simulator is

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

where

$$\begin{aligned} Z_{\mathrm{static}} &= R_0 + \frac{1}{j\omega C_0}, \\ Z_{\mathrm{motional}} &= R_m + j\omega L_m + \frac{1}{j\omega C_m}. \end{aligned} \tag{16}$$

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

3.2 Two-Port Networks and ABCD Matrices

A two-port network is a circuit module with an input port and an output port, whose external characteristics can be described by the voltages and currents at these ports. For a ladder filter comprising multiple resonators, it is unnecessary to solve individually for the voltages and currents at all internal nodes. Instead, each series or shunt branch can be represented as an independent two-port element, with network parameters describing its input–output relationship.

This study uses the ABCD matrix, also known as the transmission matrix, to represent a two-port network. Under the convention that output-port current is positive when flowing out of the network, the input and output variables are related by

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

where V_1 and I_1 denote the input-port voltage and current, respectively; V_2 and I_2 denote the output-port voltage and current, respectively; and A, B, C, and D are frequency-dependent network transmission parameters.

For a series element with impedance $Z(\omega)$, the ABCD matrix is

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

For a shunt element with admittance $Y(\omega)$, the ABCD matrix is

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

In a BAW ladder filter, series resonators lie along the main signal transmission path, whereas shunt resonators connect transmission nodes to ground. Substituting the

2025 S.T. Yau High School Science Award 优秀获奖科学论文公开

resonator impedance $Z_{\mathrm{mBVD}}(\omega)$ calculated by the mBVD model into Eq. (18) or Eq. (19) yields the two-port matrix of the corresponding resonator element.

ABCD matrices are particularly suitable for analyzing cascaded networks. Suppose a filter comprises N two-port elements along the signal propagation direction, with transmission matrices $\mathbf{M}_1, \mathbf{M}_2, \ldots, \mathbf{M}_N$. The ABCD matrix of the complete filter is then

$$\begin{aligned}\mathbf{M}_{\mathrm{total}}(\omega) &= \prod_{i=1}^{N} \mathbf{M}_i(\omega) \\ &= \mathbf{M}_1(\omega)\mathbf{M}_2(\omega)\cdots\mathbf{M}_N(\omega) = \begin{bmatrix} A(\omega) & B(\omega) \\ C(\omega) & D(\omega) \end{bmatrix}.\end{aligned} \tag{20}$$

Because matrix multiplication is not commutative, the multiplication order must match the actual cascade order.

Scattering parameters describe the relationships among incident, reflected, and transmitted waves in an RF network. In particular, S_{21} is the forward transmission coefficient measured at port 2 when a signal is applied at port 1. With the same real reference impedance Z_0 at both ports and port 2 matched, S_{21} can be calculated from the overall ABCD parameters as

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

RF systems typically use $Z_0 = 50\,\Omega$. The corresponding decibel representation is

$$S_{21,\mathrm{dB}}(\omega) = 20\log_{10}\left|S_{21}(\omega)\right|. \tag{22}$$

Thus, the complete filter's S_{21} frequency response is obtained by calculating the frequency-dependent impedance of each BAW resonator, constructing the ABCD matrices of the series and shunt elements, and cascading them in topological order. This method avoids repeatedly solving the complete circuit's nodal equations and offers a unified computational formulation, straightforward cascade extension, and high simulation efficiency. It is therefore suitable for filter parameter optimization requiring numerous frequency-response evaluations.

4 Learning Algorithm for Automated BAW Filter Design

For the parameter optimization of BAW filters with a fixed ladder topology, this study constructs a reinforcement learning environment based on the mBVD equivalent circuit and uses Dueling DQN to learn the mapping from filter states to effective parameter-adjustment actions. This section presents the formal problem definition, state space, action space, state transitions and performance evaluation, parameter initialization, and value network architecture.

4.1 Problem Formulation

Consider a filter comprising N_{se} series resonators and N_{sh} shunt resonators. The design parameters of the ith series resonator are expressed as

$$\theta_i^{\mathrm{se}} = \left(f_{s,i}^{\mathrm{se}}, k_{t,i}^{2,\mathrm{se}}, Z_i^{\mathrm{se}}\right), \tag{23}$$

where f_s, k_t^2, and Z denote the series resonance frequency, effective electromechanical coupling coefficient, and reference impedance, respectively. The parameters of the jth shunt resonator are expressed as

$$\theta_j^{\mathrm{sh}} = \left(f_{s,j}^{\mathrm{sh}}, k_{t,j}^{2,\mathrm{sh}}, Z_j^{\mathrm{sh}}, h_j\right), \tag{24}$$

where h_j denotes an additional reactance parameter in the shunt branch, with its sign indicating an added capacitance or inductance. The complete parameter vector is

$$\theta = \left[\theta_1^{\mathrm{se}}, \ldots, \theta_{N_{\mathrm{se}}}^{\mathrm{se}}, \theta_1^{\mathrm{sh}}, \ldots, \theta_{N_{\mathrm{sh}}}^{\mathrm{sh}}\right], \tag{25}$$

with dimensionality

$$d_\theta = 3N_{\mathrm{se}} + 4N_{\mathrm{sh}}. \tag{26}$$

The experiments use a ladder structure with $N_{\mathrm{se}} = N_{\mathrm{sh}} = 4$, giving $d_\theta = 28$.

Given the passband and stopband specifications, the mBVD–ABCD simulation model calculates the frequency response from the design parameters θ:

$$y_\theta(f) = S_{21,\mathrm{dB}}(f;\theta). \tag{27}$$

The corresponding parameter optimization objective can be expressed as

$$\theta^* = \arg\max_{\theta\in\Theta} \Phi\left(\theta;\mathrm{spec}\right), \tag{28}$$

where Θ is the feasible parameter domain, spec denotes the target frequency-band specifications, and Φ is the composite performance scoring function.

To solve Eq. (28) using reinforcement learning, the parameter-adjustment process is formulated as a finite-horizon Markov decision process:

$$\mathcal{M} = (\mathcal{S}, \mathcal{A}, P, R, \gamma). \tag{29}$$

The agent selects a parameter-adjustment action a_t in state s_t. The environment updates the resonator parameters accordingly, recalculates the frequency response, and returns the next state and immediate reward. By maximizing the cumulative discounted reward, the agent learns a mapping from the current filter state to an effective direction of parameter adjustment.

4.2 State Space Design

The filter state comprises three components: the frequency-response curves, resonator parameters, and global performance summary:

$$s_t = (X_t^{\mathrm{curve}}, x_t^{\mathrm{param}}, x_t^{\mathrm{sum}}). \tag{30}$$

2026 S. Yau High School Science Award 参赛中文论文 仅用于2026丘成桐中学科学奖

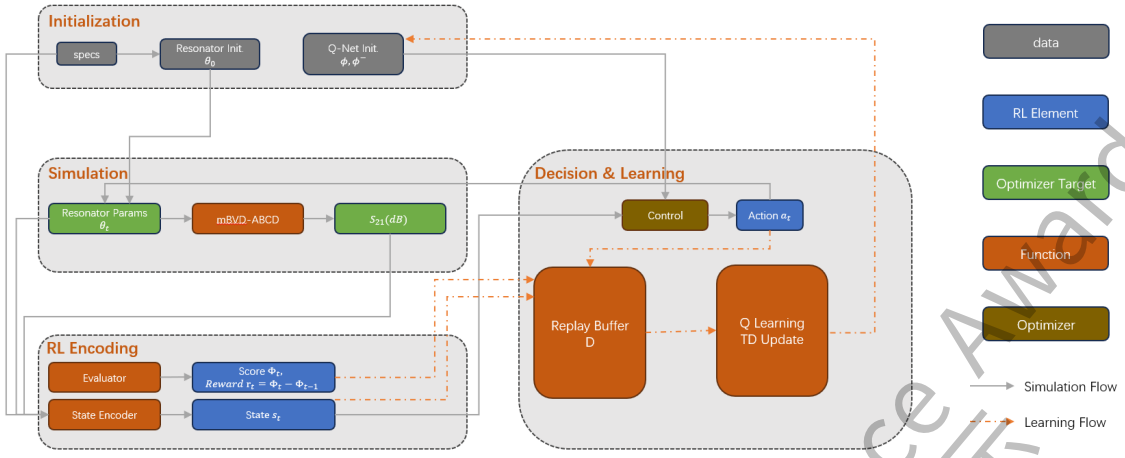


Figure 2: Simulation and reinforcement learning workflow

The frequency-response features $X_t^{\text{curve}} \in \mathbb{R}^{5\times N_c}$ are sampled along a logarithmic frequency axis, with $N_c = 512$ in this study. The five channels are

$$X_t^{\text{curve}} = \left[x^f,\, x^{S_{21}},\, x^{\text{pass}},\, x^{\text{stop}},\, x^w\right], \tag{31}$$

namely, normalized frequency, normalized $S_{21,\text{dB}}$, passband specification margin, stopband specification margin, and frequency-band importance weight. The passband margin indicates whether the current response exceeds the minimum passband transmission, while the stopband margin indicates whether it remains below the maximum stopband transmission. Frequency-band weights emphasize passband edges and critical stopbands. Consequently, the network observes both the shape of the frequency response and the direction of deviation from the target specifications at each frequency point.

The parameter features $x_t^{\text{param}} \in \mathbb{R}^{d_\theta}$ are obtained by concatenating all current resonator parameters. To reduce scale differences among physical quantities, the resonance frequencies, electromechanical coupling coefficients, reference impedances, and additional reactance parameters are normalized by their respective characteristic scales.

The global performance summary $x_t^{\text{sum}} \in \mathbb{R}^{10}$ is defined as

$$x_t^{\text{sum}} = \left[\bar{\Phi}_t,\, \bar{\Phi}_t^{\text{best}},\, \Delta\bar{\Phi}_t,\, v_{\text{p,max}},\, v_{\text{s,max}},\, \bar{v}_{\text{p}},\, \bar{v}_{\text{s}},\, \rho_{\text{p}},\, e_f,\, \eta_t\right]. \tag{32}$$

where $\bar{\Phi}_t$ is the current normalized score, $\bar{\Phi}_t^{\text{best}}$ is the best historical score within the current episode, and $\Delta\bar{\Phi}_t$ is the score change between consecutive states; $v_{\text{p,max}}$ and $v_{\text{s,max}}$ are the maximum passband and stopband violations, respectively, and $\bar{v}_{\text{p}}$ and $\bar{v}_{\text{s}}$ are the corresponding mean violations; ρ_{p} denotes passband ripple, e_f denotes center-frequency deviation, and η_t denotes the current optimization progress.

The global performance summary complements the frequency-response curves. Although metrics such as maximum violation, mean violation, passband ripple, and center-frequency deviation can be computed from the raw curves, requiring the convolutional network to derive cross-band statistics implicitly increases the difficulty of value learning. Explicitly supplying these metrics enables the agent to directly assess the main deficiencies, overall quality, and optimization stage of the current design, thereby providing clearer global information for action-value estimation.

4.3 Action Space Design

Directly requiring the policy to output continuous increments for all resonator parameters produces a high-dimensional continuous action space. Because the parameters and the S_{21} curve exhibit substantial nonlinear coupling, random continuous actions often fail to provide effective exploration. This study therefore adopts an expert-knowledge-guided discrete composite action space:

$$\mathcal{A} = \{a_0, a_1, \ldots, a_{15}\}. \tag{33}$$

The first 11 actions are expert actions targeting overall filter performance. They include center-frequency shifting, passband-width adjustment, passband transmission enhancement, stopband attenuation enhancement, band-edge roll-off improvement, passband ripple smoothing, and parameter restoration toward a reference design. Each expert action simultaneously adjusts a group of related parameters, giving the changes a clear circuit-level interpretation. This expert action set was designed by combining filter theory, candidate operations suggested by a large language model, and multiple rounds of simulation tests. Within the BAW filter automated-design literature reviewed in this study, no work using the same composite action design was identified.

The remaining 5 actions are local complementary actions, primarily used to adjust the frequencies of specific series or shunt resonators and additional reactance parameters, thereby covering local search directions not addressed by the expert actions.

2025 Si-Ti Yan Xidian School Science Award 仅用于2026年西电科学奖公示

This action space reduces the complexity of action search while preserving the necessary local adjustment capability. The specific parameter increments and frequency-response effects of each action are provided in Appendix A.

4.4 State Transitions and Performance Evaluation

After an action is selected, the environment first updates the corresponding resonator parameters according to the action definition and constrains the results to the predefined physically feasible domain. It then uses the mBVD model to calculate the frequency-dependent impedance of each resonator, converts these impedances into series or shunt two-port elements, and cascades their ABCD matrices in ladder-topology order to obtain the new $S_{21,\mathrm{dB}}$ curve. Finally, the state is reconstructed from the updated frequency response, structural parameters, and optimization history. In the absence of random device variations and simulation noise, this state-transition process is deterministic.

To evaluate the current design, let $y(f) = S_{21,\mathrm{dB}}(f)$. For a passband $\mathcal{B}_i^{\mathrm{p}}$ with minimum permitted transmission l_i^{p} and a stopband $\mathcal{B}_j^{\mathrm{s}}$ with maximum permitted transmission l_j^{s}, the pointwise violations are defined as

$$v_i^{\mathrm{p}}(f) = \max\left(0, l_i^{\mathrm{p}} - y(f)\right), \tag{34}$$

$$v_j^{\mathrm{s}}(f) = \max\left(0, y(f) - l_j^{\mathrm{s}}\right). \tag{35}$$

The specification-violation loss L_{viol} combines weighted mean violations and worst-point violations within each frequency band. Passband terms are assigned higher weights, and frequency points near passband and stopband edges also receive greater importance. In addition to specification violations, the scoring function includes three performance terms:

- L_{perf}: a loss based on the ratio of mean passband and stopband powers, used to increase power separation between the two;
- L_{shape}: the variance of $S_{21,\mathrm{dB}}$ within the passband, used to suppress passband ripple;
- e_c: the deviation of the actual response center from the target passband, set to zero when the response peak lies within the target passband.

The composite performance score is defined as

$$\Phi(\theta) = -\left[L_{\mathrm{viol}} + L_{\mathrm{perf}} + 0.05 L_{\mathrm{shape}} + 20\exp(e_c) - 20\right]. \tag{36}$$

A higher score indicates that the current filter frequency response more closely matches the target specifications. The scoring function is constructed according to the filter design requirements, with coefficients 0.05 and 20 adjusting the numerical scales and relative contributions of the relevant terms.

The score increment between consecutive states is used as the immediate reward:

$$r_t = \Phi(\theta_{t+1}) - \Phi(\theta_t). \tag{37}$$

When a parameter adjustment improves filter performance, $r_t > 0$; otherwise, $r_t < 0$. Compared with directly using the absolute score, the difference-based reward more directly reflects the marginal effect of the current action and provides continuous feedback with a clear direction.

4.5 Parameter Initialization

For each design task, reference parameters are first constructed from the target passbands. Suppose there are N_{PB} target passbands, with the ith passband spanning $[f_{i,\mathrm{lo}}^{\mathrm{p}}, f_{i,\mathrm{hi}}^{\mathrm{p}}]$. The reference frequency is defined as the mean of the passband center frequencies:

$$f_{\mathrm{base}} = \frac{1}{N_{\mathrm{PB}}} \sum_{i=1}^{N_{\mathrm{PB}}} \frac{f_{i,\mathrm{lo}}^{\mathrm{p}} + f_{i,\mathrm{hi}}^{\mathrm{p}}}{2}. \tag{38}$$

The initial resonance frequencies of the series and shunt resonators are set to

$$f_{s,i}^{\mathrm{se},0} = f_{\mathrm{base}}\xi_i^{\mathrm{se}}, \qquad f_{s,j}^{\mathrm{sh},0} = 0.98 f_{\mathrm{base}}\xi_j^{\mathrm{sh}}, \tag{39}$$

where ξ_i^{se} and ξ_j^{sh} are linearly distributed over $[0.998, 1.002]$. The remaining parameters are initialized as

$$k_t^2 = 0.06, \qquad Z = 50\,\Omega, \qquad h = 0. \tag{40}$$

This initialization follows three principles: locating the response near the passband center, offsetting the series and shunt frequencies, and slightly spreading the frequencies of resonators of the same type. The reference frequency positions the initial response near the target passband. The lower shunt resonance frequencies provide the basic series–shunt frequency arrangement required for a ladder network to produce a bandpass response. The small linear spread breaks exact symmetry among resonators of the same type, preserving degrees of freedom for subsequent adjustment of bandwidth, passband ripple, and band-edge roll-off.

During training, the environment also applies small random perturbations around these reference parameters while maintaining the basic configuration in which the shunt resonance frequency does not exceed that of the corresponding series resonator. This procedure increases the diversity of initial states, enabling the policy to learn parameter optimization behavior in the neighborhood of the reference design rather than adapting only to a single initial state.

4.6 Multiscale Dueling DQN Network

DQN is used to solve the discrete sequential optimization problem described above [8]. Training mechanisms, including experience replay, a target network, and temporal-difference updates, follow standard DQN. This subsection focuses on the value network designed for the filter state. The network input is

$$s_t = (X_t^{\mathrm{curve}}, x_t^{\mathrm{param}}, x_t^{\mathrm{sum}}), \tag{41}$$

2026 Ti Yau High School Science Award 文件于2026年1月提交科学奖文件

and the output comprises the values of all discrete actions:

$$\boldsymbol{Q}(s_t) = [Q(s_t, a_0), \ldots, Q(s_t, a_{15})]\,. \tag{42}$$

The network consists of a frequency-response encoder, a context encoder, and a dueling value-decomposition module, as shown in Fig. 3.

The frequency-response encoder uses one-dimensional convolutions to process the five-channel curves. To capture both short-range features, such as passband ripple and local resonance peaks, and long-range features, such as passband width, center frequency, and stopband extent, the encoder contains five multiscale residual convolutional blocks. Each block uses four parallel kernel sizes:

$$\mathcal{K} = \{3, 7, 15, 31\} \tag{43}$$

Some blocks use strided convolutions to reduce resolution along the frequency axis, while deeper blocks use dilated convolutions with dilation rates of 2 and 4 to expand the receptive field. Attention pooling is applied to the convolutional output, enabling the network to focus on passband edges, severe violation points, and critical stopbands, ultimately producing a 128-dimensional curve feature vector z_t^{curve}.

The resonator parameters and global performance summary are first concatenated and then encoded by a two-layer fully connected network into a 128-dimensional context feature vector z_t^{context}. The two feature types are subsequently fused as

$$z_t = \left[z_t^{\text{curve}}; z_t^{\text{context}}\right] \in \mathbb{R}^{256}. \tag{44}$$

This late-fusion scheme separately preserves the sequential structure of the frequency response and the physical meaning of the resonator parameters.

At the output, a dueling architecture decomposes the state–action value into state value and action advantage [9]:

$$Q(s, a) = V(s) + A(s, a) - \frac{1}{|\mathcal{A}|} \sum_{a' \in \mathcal{A}} A(s, a'). \tag{45}$$

The state-value stream outputs a scalar $V(s)$ to assess the overall quality of the current filter state, while the advantage stream outputs $A(s, a)$ to measure the gain or loss of each parameter-adjustment action relative to the average action. Subtracting the mean advantage eliminates the non-uniqueness of the value decomposition and gives $V(s)$ a clear interpretation as the average state value.

During filter optimization, the overall performance can vary substantially across parameter states, whereas the values of several small parameter-adjustment actions within the same state are often similar. Standard DQN must directly learn the absolute value of each action and is therefore susceptible to the overall quality of the state. The dueling architecture separately estimates whether the current design is good and which action is relatively better. This allows the network to first establish a stable assessment of state quality and then distinguish the marginal effects of different actions, improving discrimination among similar-valued actions and enhancing training stability.

5 Experiments and Results

The proposed method is evaluated through a case study of BAW filter parameter design with a fixed ladder topology. The frequency-band specifications, default initial parameters, and initial response are provided in Appendix B. This section first describes the experimental task and training settings, then analyzes the DQN optimization results, and finally examines the roles of the expert action space and global performance summary through comparison with continuous-action PPO and state-feature ablation.

5.1 Experimental Task and Settings

The passband and stopband specifications of the experimental task are externally prescribed. The ith passband specification is expressed as

$$\mathcal{B}_i^{\text{p}} = \left(f_{i,\text{lo}}^{\text{p}}, f_{i,\text{hi}}^{\text{p}}, l_i^{\text{p}}\right), \tag{46}$$

where $f_{i,\text{lo}}^{\text{p}}$ and $f_{i,\text{hi}}^{\text{p}}$ are the lower and upper passband frequency limits, respectively, and l_i^{p} is the minimum permitted $S_{21,\text{dB}}$ within the passband. Similarly, the jth stopband specification is expressed as

$$\mathcal{B}_j^{\text{s}} = \left(f_{j,\text{lo}}^{\text{s}}, f_{j,\text{hi}}^{\text{s}}, l_j^{\text{s}}\right), \tag{47}$$

where l_j^{s} is the maximum permitted $S_{21,\text{dB}}$ within the stopband.

The experiments use a ladder filter comprising 4 series resonators and 4 shunt resonators. Each series resonator has three parameters: resonance frequency, electromechanical coupling coefficient, and reference impedance. Each shunt resonator additionally includes a reactance parameter. The total number of parameters to be optimized is therefore

$$d_\theta = 4 \times 3 + 4 \times 4 = 28. \tag{48}$$

The filter frequency response is calculated using the mBVD equivalent circuit and cascaded ABCD matrices. The agent selects a parameter-adjustment operation from 16 discrete composite actions and receives an immediate reward equal to the difference in composite scores between consecutive states. Dueling DQN is trained for 300 episodes, with up to 500 parameter-adjustment steps per episode. A performance score closer to 0 indicates better agreement between the frequency response and the target specifications. Under the current settings, the initial structural parameters yield a performance score of approximately -33.9.

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

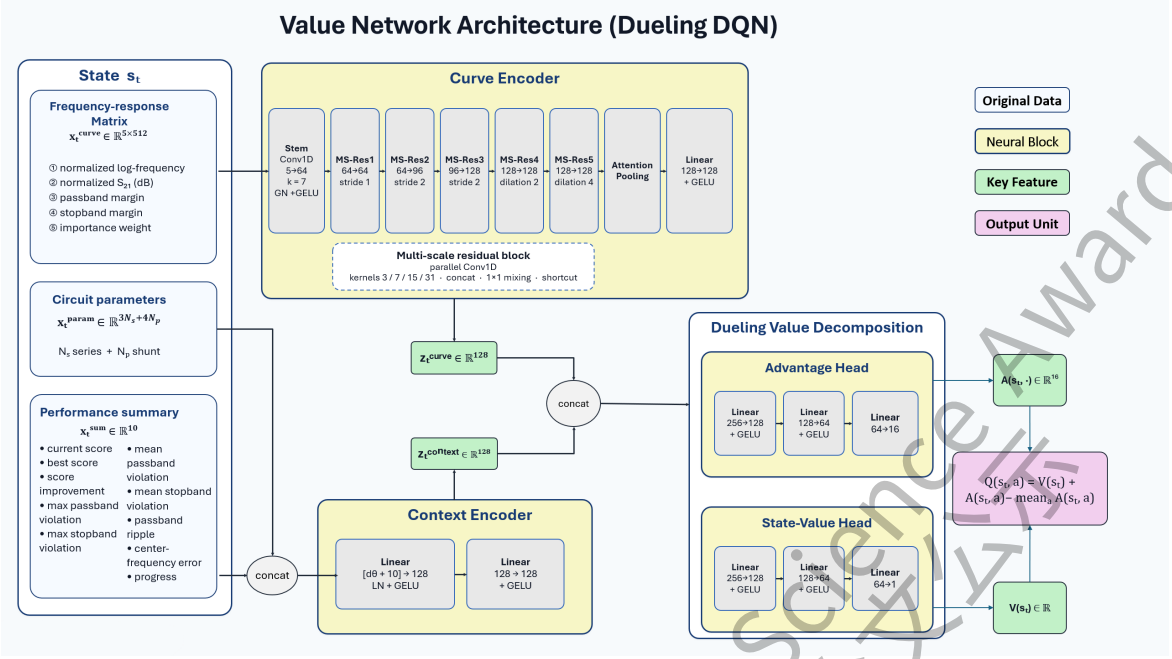


Figure 3: Architecture of the multiscale Dueling DQN value network

5.2 DQN Optimization Results

After 300 training episodes, the best parameter set found by DQN achieves a score of −3.99, an improvement of approximately 29.9 over the initial design. The corresponding $S_{21,\mathrm{dB}}$ curve is shown in Fig. 4.

The optimized frequency response satisfies all passband and stopband thresholds, demonstrating that the agent can effectively adjust the frequency arrangement, coupling strengths, and impedance relationships of the series and shunt resonators through discrete composite actions. The remaining gap between the best score and 0 is mainly attributable to the passband ripple term, while the principal constraints on passband transmission and stopband attenuation are all satisfied.

Figure 5 shows the best score achieved in each training episode. As training progresses, the agent gradually accumulates action-value information across different states and continues to discover better parameter combinations. On the experimental platform used in this study, the complete 300-episode training process takes approximately 200 minutes, indicating that the fast mBVD–ABCD simulation model can support the large number of repeated performance evaluations required by reinforcement learning.

5.3 Additional Design Specifications and Resonator Counts

The results in Table 1 show that, under the default target specifications, as the filter topology is expanded from 3s3p to 5s5p, the number of optimization parameters increases from 21 to 35, yet the best scores achieved by the different topologies remain consistently around −3.99. This indicates that the proposed state representation, expert-guided discrete action space, and Dueling DQN can accommodate topology-size variations within a certain range and achieve similar final optimization performance across different parameter dimensions.

For the fixed 4s4p topology, when the target passband width is increased to 120% of its original value, the performance score improves from −42.1618 to −4.0523. When the target passband is shifted upward in frequency by 30%, the best score reaches −3.9941, which is essentially consistent with the result under the default specifications. These results indicate that the proposed method is applicable to filter topologies of different sizes and can effectively optimize design tasks with changes in passband location and width.

By contrast, when the passband width is reduced to 80% of its original value, the best score is −4.3532, slightly lower than those obtained in the other experiments. This suggests that the coupling among band-edge locations, passband ripple, and stopband attenuation may become stronger under narrow-passband conditions, increasing the difficulty of coordinated parameter optimization.

5.4 Comparison and Ablation Analysis

Table 2 summarizes the results of the complete method, continuous-action PPO, and state ablation experiments.

5.4.1 Removing the Dueling Architecture

To evaluate the role of the dueling architecture in the overall algorithm, a comparative experiment was conducted using a neural network without this architecture. Its learning

2026 Yan Hua High School Science Award 延华2026年科学奖

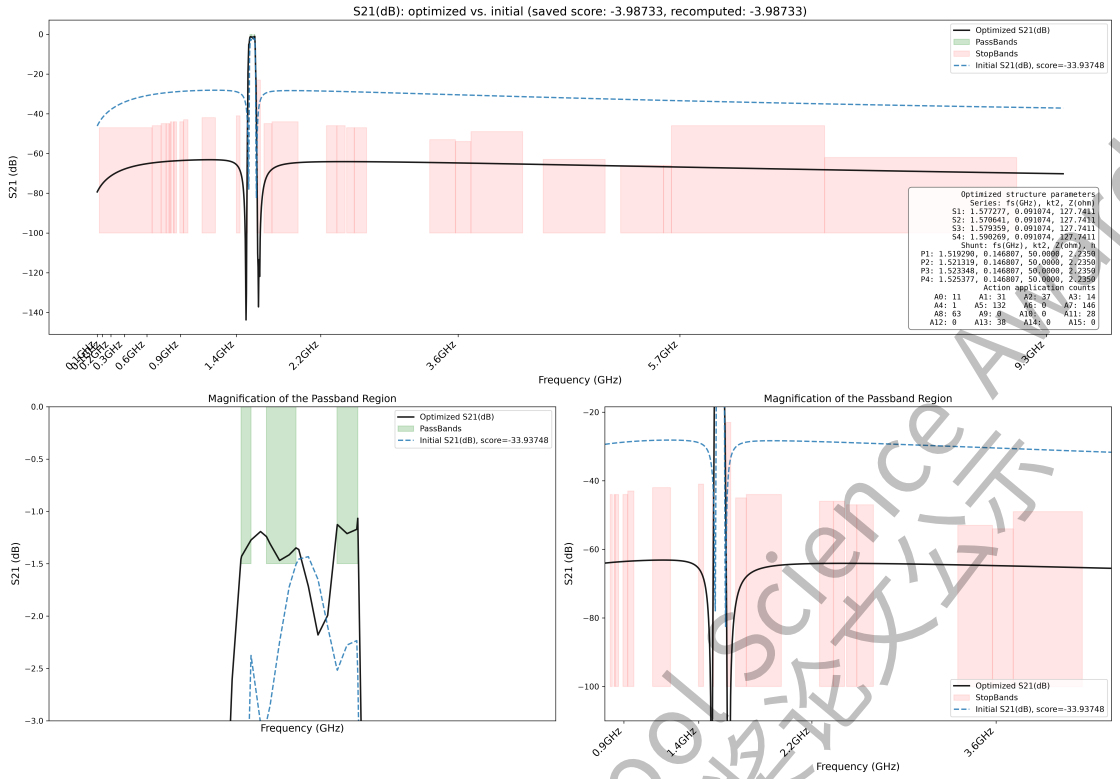


Figure 4: The $S_{21,\mathrm{dB}}$ curves of the initial design and the best design obtained by DQN. The blue dashed line denotes the initial frequency response, while the green and red markers indicate the passband and stopband specifications, respectively.

curve (*Without Dueling* in Fig. 5) exhibits substantial oscillations and does not stabilize near the optimum. This indicates that the dueling architecture helps improve training stability under the experimental settings of this study.

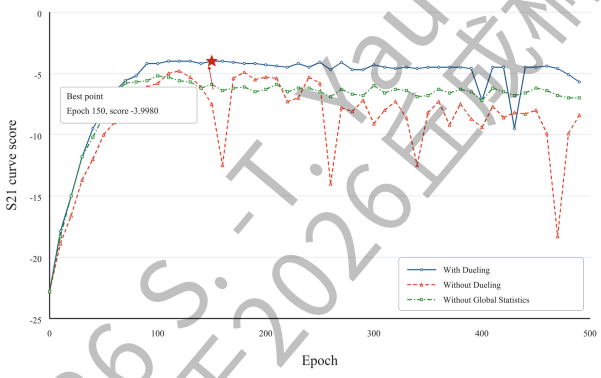


Figure 5: Best performance score per episode during training for different algorithm variants

5.4.2 Comparison with Continuous-Action PPO

To analyze the role of the expert-guided discrete action space, continuous-action PPO is additionally used for comparison. In this setting, the policy directly outputs continuous adjustments to the 28 structural parameters, with each action dimension assumed to follow an independent Gaussian distribution.

After 500 training episodes, continuous-action PPO fails to achieve a substantial performance improvement, with a best score no higher than approximately -21.5, well below the -3.99 obtained by DQN. This result indicates that, under the current experimental settings, direct random exploration in the high-dimensional continuous parameter space is inefficient. The effects of different parameters on the frequency response are strongly nonlinear and coupled, so many continuous actions fail to produce consistent performance improvements and may rapidly move the current design away from an effective parameter region.

In contrast, expert actions combine multidimensional parameter changes into physically meaningful operations, such as center-frequency shifting, bandwidth adjustment, passband transmission enhancement, and stopband attenuation enhancement. These operations reduce ineffective

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

Table 1: Comparison of training results under different target specifications and resonator topologies

Target specification	N_s	N_p	Parameter dimension D	Initial score	Best score
Default specifications	3	3	21	−33.38	−3.9910
Default specifications	4	3	24	−36.0011	−3.9918
Default specifications	4	4	28	−33.9097	−3.9908
Default specifications	5	4	31	−31.9720	−3.9931
Default specifications	5	5	35	−29.6696	−3.9925
Passband width increased to 120%	4	4	28	−42.1618	−4.0523
Passband width reduced to 80%	4	4	28	−18.5345	−4.3532
Passband shifted upward in frequency by 30%	4	4	28	−34.2733	−3.9941

Note: N_s and N_p denote the numbers of series and shunt resonators, respectively. The parameter dimension is calculated as $D = 3N_s + 4N_p$. Initial scores are inferred from the best scores and the score improvements recorded during training.

Table 2: Experimental results for different optimization methods and state configurations

Method	Action or state configuration	Training episodes	Best score
Dueling DQN	Discrete expert actions and full state	300	−3.99
PPO	28-dimensional continuous actions	500	−21.5
Dueling DQN	Global performance summary removed	300	−5.20
DQN	Dueling architecture removed	300	−4.82

exploration and provide the agent with clearer search directions. This experiment reflects differences between the two action spaces under the task and training settings used here; it does not imply that DQN outperforms PPO in all continuous parameter optimization problems.

5.4.3 Ablation of the Global Performance Summary

To analyze the role of the global performance summary, statistical features including the current score, maximum violation, mean violation, passband ripple, center-frequency deviation, and optimization progress are removed from the DQN state while all other settings remain unchanged. Only the frequency-response curves and resonator parameters are retained.

After removing the global performance summary, DQN achieves a best score of approximately −5.2 (*Without Global Statistics* in Fig. 5), below the −3.99 obtained with the full state. Although some global performance information can in principle be inferred from the complete $S_{21,\mathrm{dB}}$ curve, requiring the convolutional network to implicitly calculate cross-band maxima, means, variances, and center locations increases the difficulty of action-value learning. Explicitly providing these statistics helps the agent quickly identify the main types of specification violations and the overall optimization stage, thereby improving parameter search performance.

6 Conclusions and Outlook

6.1 Conclusions

This paper proposes an automated design method for parameter optimization of BAW filters with a fixed ladder topology, combining fast mBVD equivalent-circuit simulation, an expert-guided discrete action space, and a multi-scale Dueling DQN. The mBVD–ABCD model provides a low-cost mapping from resonator parameters to the $S_{21,\mathrm{dB}}$ response. The composite state jointly describes the frequency response, physical parameters, and global performance. Expert actions transform high-dimensional continuous parameter changes into discrete tuning operations with circuit-level meaning. Multiscale convolutions, attention pooling, and dueling value decomposition are used to extract frequency-domain patterns, emphasize critical frequency locations, and distinguish state quality from relative action advantages, respectively.

In the experimental task comprising 4 series resonators and 4 shunt resonators, the proposed method improves the composite performance score from approximately −33.9 to −3.99, with the optimized response satisfying the prescribed passband and stopband thresholds. Under the current settings, continuous-action PPO achieves a best score no higher than approximately −21.5, indicating the difficulty of direct exploration in a 28-dimensional continuous action space. Removing the global performance summary reduces the best DQN score to approximately −5.2, demonstrating the benefit of explicit global performance summaries for action-value assessment. Overall, the experimental results support the feasibility of combining fast equivalent-circuit simulation with deep reinforcement learning for BAW filter parameter optimization.

In addition, the method discussed in this paper addresses how to investigate automated BAW filter design without access to commercial simulation software or ex-

2026 S. Yau High School Science Award
仅用于2026年成相中学科学奖论文公示

tensive records of its simulation runs.

6.2 Limitations and Future Work

This study has the following limitations. First, the discrete expert actions are primarily designed from human experience, and their effectiveness depends on the specific topology and parameter ranges. Some actions are also used infrequently along the optimal trajectories. Future research may investigate action pruning, adaptive step sizes, and automatic action composition based on action visitation frequencies, advantage distributions, and actual performance contributions.

Second, the mBVD model neglects some higher-order acoustic modes, process variations, and layout parasitics, so the results remain at the equivalent-circuit optimization level. Future work may combine fast mBVD-based search with finite element simulation, manufacturing-tolerance analysis, and measurement data to construct a multifidelity optimization workflow, thereby improving physical credibility and engineering transferability.

2026 S. - T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

A Expert Action Space Design and Response Effects

A.1 Action Space Design Principles

The design parameters of a BAW ladder filter include the resonance frequencies, electromechanical coupling coefficients, reference impedances, and additional reactance parameters of the series and shunt resonators. If all parameters are adjusted independently and continuously, the action dimensionality increases with the number of resonators, and many random parameter combinations lack clear physical meaning. To reduce the exploration difficulty of reinforcement learning, commonly used filter-tuning experience is encoded into a finite set of discrete composite actions:

$$\mathcal{A} = \{a_0, a_1, \ldots, a_{15}\}. \tag{49}$$

Here, a_0–a_{10} are expert actions that primarily adjust center frequency, passband width, passband transmission, stopband attenuation, band-edge roll-off, and passband ripple. Actions a_{11}–a_{15} are local complementary actions that cover local parameter-adjustment directions not addressed by the expert actions. Each action corresponds to a deterministic parameter-update operator

$$\theta_{t+1} = \mathcal{U}(\theta_t, a_t), \tag{50}$$

and the updated parameters must remain within the predefined physically feasible domain.

The filter parameters and the S_{21} curve are nonlinearly coupled. Accordingly, the action effects described below represent the dominant response trends near the reference design and do not imply strictly monotonic effects at every possible parameter state.

A.2 Parameter-Update Conventions

Let the series and shunt resonator parameters be

$$\theta_i^{\mathrm{se}} = \left(f_{s,i}^{\mathrm{se}}, k_{t,i}^{2,\mathrm{se}}, Z_i^{\mathrm{se}}\right), \tag{51}$$

and

$$\theta_j^{\mathrm{sh}} = \left(f_{s,j}^{\mathrm{sh}}, k_{t,j}^{2,\mathrm{sh}}, Z_j^{\mathrm{sh}}, h_j\right). \tag{52}$$

Here, h_j is a scaled control parameter for an additional passive branch, rather than a physical quantity directly expressed in farads or henries. When $h_j > 0$, it corresponds to an additional shunt capacitance

$$C_j = h_j \times 10^{-12}\ \mathrm{F}; \tag{53}$$

and when $h_j < 0$, it corresponds to an additional shunt inductance

$$L_j = (-h_j) \times 10^{-8}\ \mathrm{H}. \tag{54}$$

Thus, increasing h_j increases capacitive compensation or reduces existing inductive compensation, whereas decreasing h_j reduces capacitive compensation or increases inductive compensation.

For relative updates, the notation

$$x \leftarrow (1+\delta)x \tag{55}$$

denotes an increase of $\delta\times100\%$ in parameter x. All composite actions act simultaneously on the specified set of resonators.

A.3 Expert Actions

A.3.1 Center-Frequency and Bandwidth Adjustment

Action a_0: Upward center-frequency shift Increase the series resonance frequencies of all series and shunt resonators by 0.03%:

$$f_s^{\mathrm{se}} \leftarrow 1.0003 f_s^{\mathrm{se}}, \qquad f_s^{\mathrm{sh}} \leftarrow 1.0003 f_s^{\mathrm{sh}}. \tag{56}$$

This action approximately preserves the relative frequency arrangement of the series and shunt resonators and primarily shifts the entire filter passband toward higher frequencies.

Action a_1: Downward center-frequency shift Decrease the series resonance frequencies of all series and shunt resonators by 0.03%:

$$f_s^{\mathrm{se}} \leftarrow 0.9997 f_s^{\mathrm{se}}, \qquad f_s^{\mathrm{sh}} \leftarrow 0.9997 f_s^{\mathrm{sh}}. \tag{57}$$

This action primarily shifts the entire filter passband toward lower frequencies.

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

Action a_2: Passband widening Increase the resonance frequencies of all series resonators by 0.03% while decreasing those of all shunt resonators by 0.03%:

$$f_s^{\mathrm{se}} \leftarrow 1.0003 f_s^{\mathrm{se}}, \qquad f_s^{\mathrm{sh}} \leftarrow 0.9997 f_s^{\mathrm{sh}}. \tag{58}$$

This action increases the frequency separation between the series and shunt resonators and typically widens the low-loss transmission region near the reference parameters.

Action a_3: Passband narrowing Define the difference between the mean series and shunt resonance frequencies as

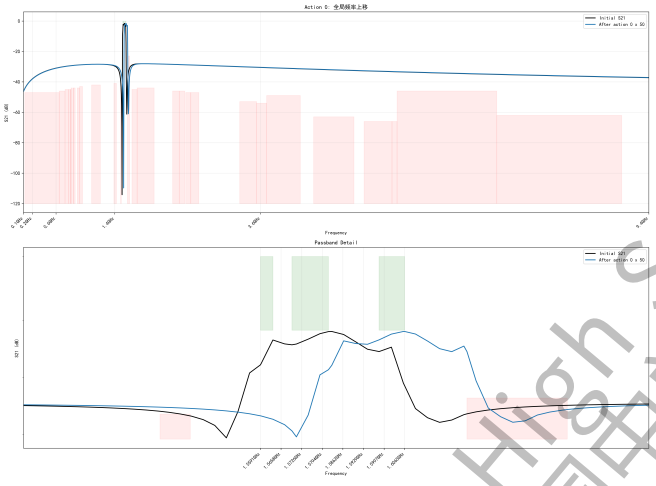
$$\Delta \bar{f} = \frac{1}{N_{\mathrm{se}}} \sum_i f_{s,i}^{\mathrm{se}} - \frac{1}{N_{\mathrm{sh}}} \sum_j f_{s,j}^{\mathrm{sh}}. \tag{59}$$

Then update

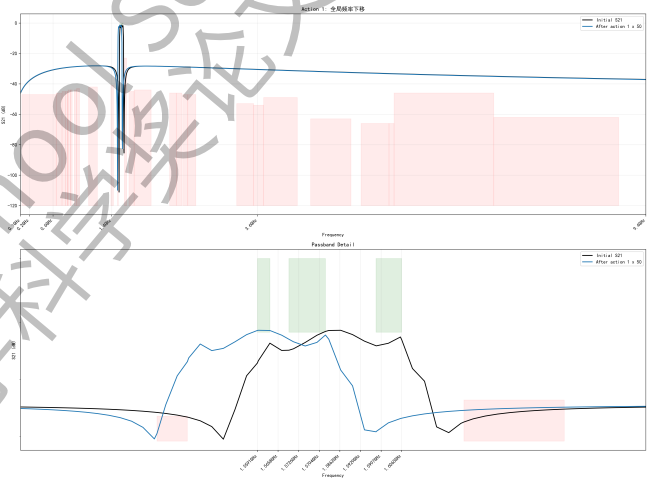
$$f_{s,i}^{\mathrm{se}} \leftarrow f_{s,i}^{\mathrm{se}} - 0.005 \Delta \bar{f}, \tag{60}$$

$$f_{s,j}^{\mathrm{sh}} \leftarrow f_{s,j}^{\mathrm{sh}} + 0.005 \Delta \bar{f}. \tag{61}$$

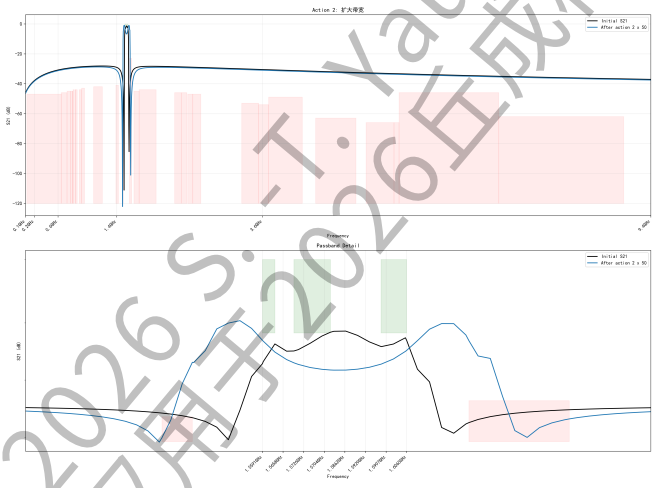
This action reduces the frequency separation between the series and shunt resonators, causing the passband to contract toward its center within the current parameter neighborhood.



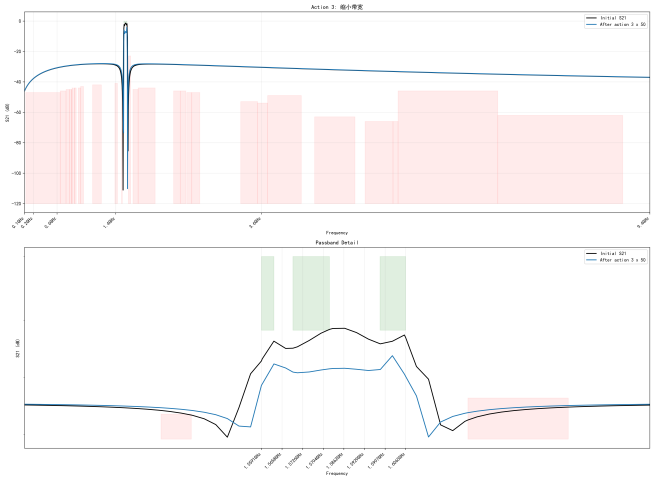
(a) Action a_0: Upward center-frequency shift



(b) Action a_1: Downward center-frequency shift



(c) Action a_2: Passband widening



(d) Action a_3: Passband narrowing

Figure 6: Frequency-response effects of center-frequency and passband-width adjustment actions. For clarity, each action is applied 50 consecutive times to amplify the observed trend.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

A.3.2 Passband Transmission and Stopband Attenuation Adjustment

Action a_4: Passband transmission enhancement Decrease the reference impedances of all series resonators by 0.9% while increasing those of all shunt resonators by 0.9%:

$$Z^{\mathrm{se}} \leftarrow 0.991 Z^{\mathrm{se}}, \qquad Z^{\mathrm{sh}} \leftarrow 1.009 Z^{\mathrm{sh}}. \tag{62}$$

This action changes the relative impedances of the main transmission path and the shunt bypass paths and typically helps reduce passband insertion loss near the reference design.

Action a_5: Stopband attenuation enhancement Increase the electromechanical coupling coefficients of all shunt resonators by 1%, while increasing capacitive compensation or reducing inductive compensation in the shunt branches:

$$k_t^{2,\mathrm{sh}} \leftarrow 1.01 k_t^{2,\mathrm{sh}}, \qquad h^{\mathrm{sh}} \leftarrow h^{\mathrm{sh}} + 0.005. \tag{63}$$

This action primarily strengthens signal bypassing through the shunt branches over the relevant frequency range, thereby improving stopband rejection.

Action a_6: Upper band-edge roll-off improvement Decrease the electromechanical coupling coefficients of all shunt resonators except the last one by 1%:

$$k_{t,j}^{2,\mathrm{sh}} \leftarrow 0.99 k_{t,j}^{2,\mathrm{sh}}, \qquad j = 1, \ldots, N_{\mathrm{sh}} - 1. \tag{64}$$

By adjusting the distribution of coupling among the shunt resonators, this action primarily improves the shape of the upper passband edge.

Action a_7: Upper-passband transmission enhancement Increase the electromechanical coupling coefficients of all series resonators by 0.2% while decreasing those of all shunt resonators by 0.2%:

$$k_t^{2,\mathrm{se}} \leftarrow 1.002 k_t^{2,\mathrm{se}}, \qquad k_t^{2,\mathrm{sh}} \leftarrow 0.998 k_t^{2,\mathrm{sh}}. \tag{65}$$

This action changes the relative coupling strengths of the series and shunt resonators and primarily improves transmission in the high-frequency region of the passband.

A.3.3 Passband Shape and Reference-Parameter Adjustment

Action a_8: Passband ripple smoothing Simultaneously adjust the electromechanical coupling coefficients of the series and shunt resonators, the reference impedances of the series resonators, and the compensation parameters of the shunt branches:

$$k_t^{2,\mathrm{se}} \leftarrow 1.01 k_t^{2,\mathrm{se}}, \tag{66}$$

$$k_t^{2,\mathrm{sh}} \leftarrow 0.99 k_t^{2,\mathrm{sh}}, \tag{67}$$

$$Z^{\mathrm{se}} \leftarrow 1.015 Z^{\mathrm{se}}, \tag{68}$$

$$h^{\mathrm{sh}} \leftarrow h^{\mathrm{sh}} + 0.025. \tag{69}$$

By jointly adjusting multiple parameters, this action changes the relative strengths of the resonant modes within the passband and primarily reduces passband ripple. Because several coupled parameters are involved, its actual effect depends strongly on the current filter state.

Action a_9: Restoration toward the reference design Let f_c be the target passband center frequency. Move the frequency of every resonator toward f_c by 10% of its current difference from f_c:

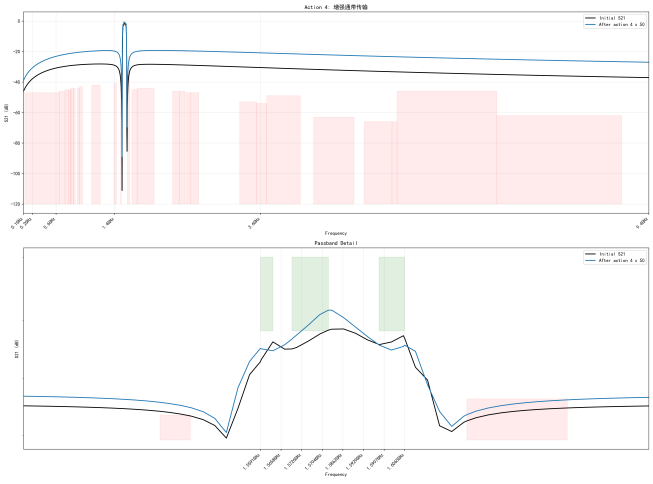
$$f_s \leftarrow f_s + 0.1(f_c - f_s). \tag{70}$$

At the same time, move the reference impedance of each series resonator toward $50\,\Omega$ by 10% of the current difference:

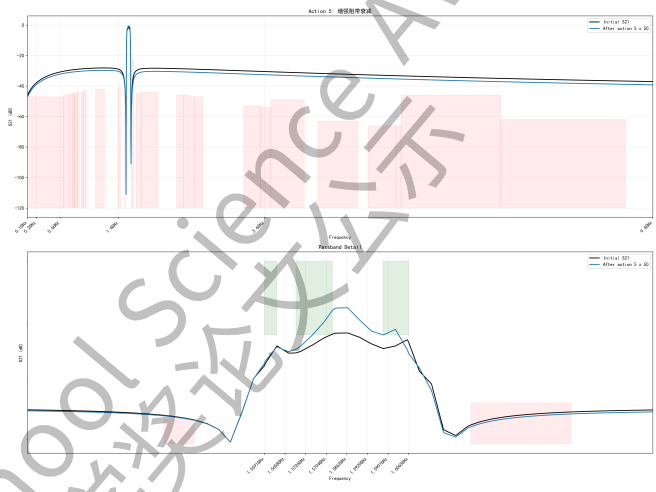
$$Z^{\mathrm{se}} \leftarrow Z^{\mathrm{se}} + 0.1\,(50\,\Omega - Z^{\mathrm{se}})\,. \tag{71}$$

This action provides a restoration mechanism. When other actions move the parameters away from a reasonable region, it gradually returns the design toward the reference parameters.

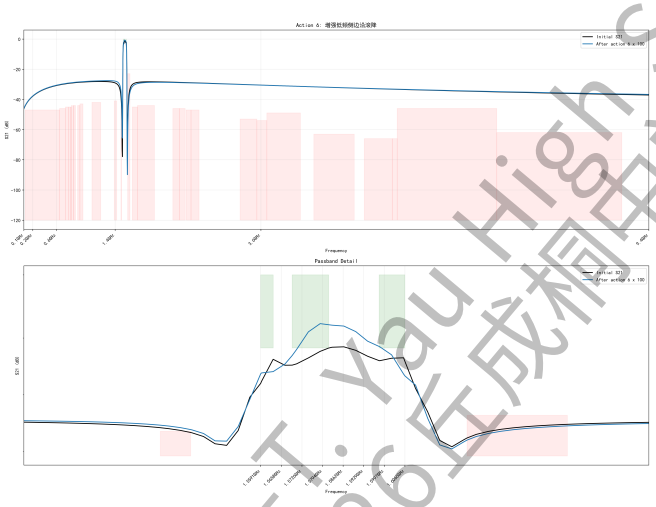
2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示



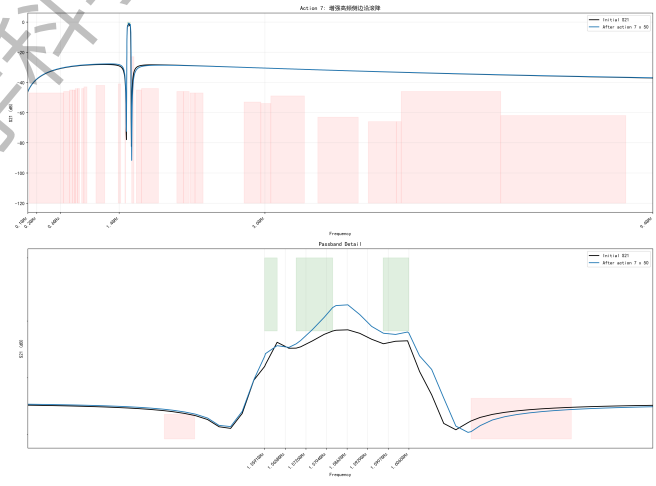
(a) Action a_4: Passband transmission enhancement



(b) Action a_5: Stopband attenuation enhancement



(c) Action a_6: Upper-edge roll-off improvement



(d) Action a_7: Upper-passband transmission enhancement

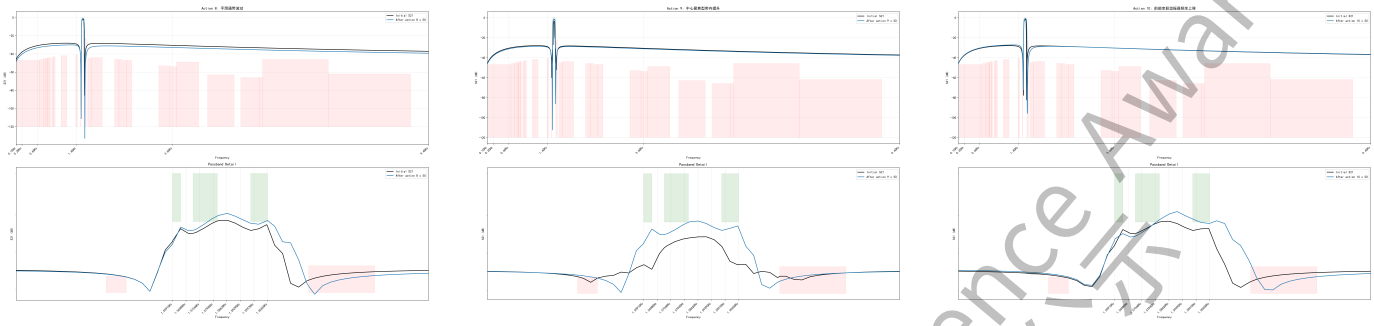
Figure 7: Frequency-response effects of passband transmission and stopband attenuation adjustment actions. Each action is applied 50 consecutive times to amplify the observed trend.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

Action a_{10}: Upper-passband extension Increase the electromechanical coupling coefficients of all series resonators by 0.6%:

$$k_t^{2,\mathrm{se}} \leftarrow 1.006 k_t^{2,\mathrm{se}}. \tag{72}$$

This action primarily adjusts the separation between the resonance and antiresonance frequencies of the series resonators and can improve or extend the upper-passband response within the current parameter neighborhood.



(a) Action a_8: Passband ripple smoothing (b) Action a_9: Reference-parameter restoration (c) Action a_{10}: Upper-passband extension

Figure 8: Frequency-response effects of passband-shape and reference-parameter adjustment actions. Each action is applied 50 consecutive times to amplify the observed trend.

A.4 Local Complementary Actions

Expert actions primarily adjust the overall parameters of resonators of the same type, effectively controlling the macroscopic filter response. However, they do not readily cover subtle frequency-response variations caused by individual resonators or local groups. Five additional complementary actions are therefore introduced to improve reachability within the local parameter space.

Action a_{11}: Frequency increase of the first series resonators Increase the resonance frequencies of the first two series resonators by 0.02%:

$$f_{s,i}^{\mathrm{se}} \leftarrow 1.0002 f_{s,i}^{\mathrm{se}}, \qquad i \in \{1, 2\}. \tag{73}$$

Action a_{12}: Frequency increase of the first shunt resonators Increase the resonance frequencies of the first two shunt resonators by 0.02%:

$$f_{s,j}^{\mathrm{sh}} \leftarrow 1.0002 f_{s,j}^{\mathrm{sh}}, \qquad j \in \{1, 2\}. \tag{74}$$

Action a_{13}: Frequency increase of the middle series resonators Increase the frequencies of the series resonators in the middle of the ladder network by 0.02%:

$$f_{s,i}^{\mathrm{se}} \leftarrow 1.0002 f_{s,i}^{\mathrm{se}}, \qquad i \in \mathcal{I}_{\mathrm{mid}}^{\mathrm{se}}. \tag{75}$$

Action a_{14}: Frequency increase of the middle shunt resonators Increase the frequencies of the shunt resonators in the middle of the ladder network by 0.02%:

$$f_{s,j}^{\mathrm{sh}} \leftarrow 1.0002 f_{s,j}^{\mathrm{sh}}, \qquad j \in \mathcal{I}_{\mathrm{mid}}^{\mathrm{sh}}. \tag{76}$$

Here, $\mathcal{I}_{\mathrm{mid}}^{\mathrm{se}}$ and $\mathcal{I}_{\mathrm{mid}}^{\mathrm{sh}}$ denote the index sets of the middle series and shunt resonators, respectively.

Action a_{15}: Increased inductive or reduced capacitive compensation Decrease the shunt-branch control parameter by 0.001:

$$h^{\mathrm{sh}} \leftarrow h^{\mathrm{sh}} - 0.001. \tag{77}$$

When $h > 0$, this action reduces the added capacitance; when $h < 0$, it increases the added inductance. Its adjustment direction is opposite to the increases in h used in a_5 and a_8, providing bidirectional search capability for the additional reactance parameter.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

A.5 Summary of the Action Space

Tables 3 and 4 summarize the parameter updates and primary design objectives of the discrete actions.

Table 3: Expert actions and their primary effects

Action	Parameter update	Primary design objective
a_0	Increase f_s of all series and shunt resonators by 0.03%	Shift the center frequency upward
a_1	Decrease f_s of all series and shunt resonators by 0.03%	Shift the center frequency downward
a_2	Increase series f_s by 0.03% and decrease shunt f_s by 0.03%	Widen the passband
a_3	Reduce the mean frequency separation between series and shunt resonators	Narrow the passband
a_4	Decrease series Z by 0.9% and increase shunt Z by 0.9%	Enhance passband transmission
a_5	Increase shunt k_t^2 by 1% and increase h by 0.005	Enhance stopband rejection
a_6	Decrease shunt k_t^2 by 1% except in the final stage	Improve upper band-edge roll-off
a_7	Increase series k_t^2 by 0.2% and decrease shunt k_t^2 by 0.2%	Enhance upper-passband transmission
a_8	Jointly adjust series and shunt k_t^2, series Z, and shunt h	Smooth passband ripple
a_9	Restore f_s toward the target center frequency and series Z toward $50\,\Omega$	Return toward the reference design
a_{10}	Increase series k_t^2 by 0.6%	Extend the upper passband

Table 4: Local complementary actions and their primary effects

Action	Parameter update	Primary design objective
a_{11}	Increase f_s of the first two series resonators by 0.02%	Locally adjust series resonance frequencies
a_{12}	Increase f_s of the first two shunt resonators by 0.02%	Locally adjust shunt resonance frequencies
a_{13}	Increase f_s of the middle series resonators by 0.02%	Supplement frequency search for middle series resonators
a_{14}	Increase f_s of the middle shunt resonators by 0.02%	Supplement frequency search for middle shunt resonators
a_{15}	Decrease shunt-branch h by 0.001	Reduce capacitive or increase inductive compensation

Overall, expert actions regulate macroscopic performance, including center frequency, bandwidth, passband transmission, and stopband attenuation, while local complementary actions provide finer-grained parameter adjustments. Together, they transform the original high-dimensional continuous parameter optimization into a discrete sequential decision problem with clear circuit-level meaning while retaining a basic capability to correct local frequency-response features.

B Experimental Filter Specifications and Initial Response

B.1 Filter Specifications

The filter design objectives used in this study are jointly defined by passband and stopband constraints. Passband constraints require $S_{21,\mathrm{dB}}$ within specified frequency ranges to remain at or above the prescribed lower limits, whereas stopband constraints require $S_{21,\mathrm{dB}}$ to remain at or below the prescribed upper limits. The simulation frequency range is obtained by extending the minimum starting frequency and maximum ending frequency across all specified bands. In this experiment, it spans 80 MHz to 9435.3 MHz.

B.1.1 Passband Specifications

B.1.2 Stopband Specifications

B.2 Default Initial Equivalent-Circuit Parameters

The default structure comprises 4 series resonators and 4 shunt resonators. The default center reference frequency is the mean of the three passband center frequencies, namely 1579.6 MHz. The initial resonance frequencies of the series

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

Table 5: Passband specifications of the experimental filter

ID	Start frequency/MHz	End frequency/MHz	S_{21} lower limit/dB
P1	1569.42	1581.42	-1.5
P2	1598.06	1606.50	-1.5
P3	1559.05	1563.15	-1.5

Table 6: Stopband specifications of the experimental filter

ID	Start frequency/MHz	End frequency/MHz	S_{21} upper limit/dB
S1	100.00	612.00	-47
S2	612.00	698.00	-46
S3	699.00	748.00	-45
S4	748.00	777.00	-45
S5	777.00	787.00	-45
S6	788.00	803.00	-44
S7	815.00	824.00	-44
S8	824.00	849.00	-44
S9	880.00	915.00	-44
S10	915.00	960.00	-43
S11	1094.40	1224.00	-42
S12	1427.00	1463.00	-41
S13	1526.00	1536.00	-44
S14	1627.00	1660.00	-23
S15	1695.00	1772.00	-45
S16	1772.00	2025.00	-44
S17	2300.00	2400.00	-46
S18	2401.00	2482.00	-46
S19	2496.00	2570.00	-47
S20	2570.00	2690.00	-47
S21	3300.00	3550.00	-53
S22	3550.00	3700.00	-54
S23	3700.00	4200.00	-49
S24	4400.00	5000.00	-63
S25	5150.00	5566.00	-66
S26	5566.00	5641.00	-66
S27	5641.00	7125.00	-46
S28	7125.00	8986.00	-62

2026 S.T. Yau High School Science Award
仅用于2026年丘成桐中学科学奖论文公示

resonators are linearly spread around this reference frequency, while those of the shunt resonators are linearly spread around 0.98 times the reference frequency. All initial coupling coefficients k_t^2 are 0.06, all reference impedances Z are $50\,\Omega$, and all additional shunt-branch parameters h are initialized to 0.

B.2.1 Series Resonator Parameters

Table 7: Default initial series resonator parameters

ID	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

ID	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} Curve and Score

Simulation is performed at 2000 frequency samples using the above specifications and default initial structural parameters. The initial $S_{21,\mathrm{dB}}$ response is shown as the blue dashed line in Fig. 4. The current scoring function first constructs loss terms from passband violations, stopband violations, worst-point constraints, the passband-to-stopband power ratio, passband fluctuations, and center-frequency deviation, and then takes their negative as the composite performance score. Scores are therefore generally negative, with values closer to 0 indicating closer agreement with the specifications.

The composite performance score of this default initial structure is

$$\Phi_{\mathrm{initial}} = -33.937478. \tag{78}$$

This result indicates that the default initial parameters do not yet satisfy the target specifications. Subsequent reinforcement learning training aims to improve this composite performance score through interactive parameter tuning and gradually bring the response into compliance with the passband and stopband constraints.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] Yens Lindemans, Thijs Ullrick, Ivo Couckuyt, Dirk Deschrijver, and Tom Dhaene. Bayesian optimization of microwave filters: A physics-informed approach using the szegő 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] 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.

[8] 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.

[9] 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.

[10] 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.

[11] Gergely Simon, Gergely B. Hantos, Mihir S. Patel, Andrew Tweedie, and Gerald Harvey. Machine learning enabled FBAR digital twin for rapid optimization. In *2020 IEEE International Ultrasonics Symposium (IUS)*. IEEE, 2020.

[12] Mohammad Almalkawi and Josh Caron. Domain decomposition based artificial neural networks (ANNs) modeling of acoustic wave resonators and filters. In *2021 IEEE 21st Annual Wireless and Microwave Technology Conference (WAMICON)*. IEEE, 2021.

[13] Xi He, Xing Haw Marvin Tan, Ma Chen, Feixuan Huang, Fengyuan Yang, Qinghua Ren, Yiming Ma, Jianlin Chen, and Nan Wang. Neural network assisted 3d FBAR modeling with random electrode shapes. In *2024 IEEE MTT-S International Conference on Microwave Acoustics and Mechanics (IC-MAM)*, pages 13–16. IEEE, 2024.

[14] Zhengcai Zhao, Danyu Mu, Rui Ding, Weipeng Xuan, F. Gao, J. Luo, and S. Dong. Fast design and verification of BAW filters based on graph neural networks. In *2026 IEEE 21st International Conference on Nano/Micro Engineered and Molecular Systems (NEMS)*, pages 281–286. IEEE, 2026.

[15] 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.

[16] Sung-Yuan Tseng et al. Synthesis of chebyshev/elliptic filters using minimum acoustic wave resonators. *IEEE Transactions on Microwave Theory and Techniques*, 2019.

[17] 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.

2026 S. T. Yau High School Science Award 优胜作品 2026丘成桐中学科学奖获奖论文公示

Acknowledgments and Author Contributions

This project investigates whether reinforcement learning can be used for automated parameter optimization of bulk acoustic wave filters. It was independently completed by the author as an individual student research project. My work included selecting the research topic, reviewing the literature, selecting and implementing the simulation model, formulating the optimization problem, designing the reinforcement learning method, conducting and iterating on experiments, analyzing the results, and drafting and organizing the manuscript. More specifically, I independently constructed the state space, designed the expert action set, scoring function, and reward function, selected and improved the reinforcement learning algorithm, and summarized the methods, experimental results, and conclusions.

This project investigates whether reinforcement learning can be used for automated parameter optimization of bulk acoustic wave filters and was undertaken by the author as an individual student research project under the supervisor's guidance. The author's work included selecting the research topic, reviewing the literature, selecting and implementing the simulation model, formulating the optimization problem, designing the reinforcement learning method, conducting experiments and iteratively refining the approach, analyzing the results, and drafting and organizing the manuscript. Specifically, the author constructed the state space, designed the expert action set, scoring function, and reward function, selected and improved the reinforcement learning algorithm, and completed the methodological description, experimental analysis, and summary of the research conclusions.

Mr. Haobin Zhu is a subject teacher at the author's school and served as the school-based supervisor for this project. He taught the principles of RF components, equivalent-circuit modeling, and reinforcement learning, and provided research guidance by discussing the feasibility of proposed approaches, identifying potential issues in experimental design, recommending relevant theories and research directions, and offering suggestions on manuscript structure and academic expression. Mr. Zhu's primary responsibilities were theoretical instruction and research supervision. The author carried out the selection of the research topic, acquisition of experimental data, algorithm design and improvement, experimental implementation and iteration, result analysis, literature review, methodological development, and writing of the substantive manuscript content under his guidance.

In subsequent experiments, I observed that training with a conventional DQN remained unstable, with substantial fluctuations in circuit scores. Mr. Zhu introduced the general principles of the dueling network architecture and its typical applications. After studying the relevant theory and literature, I independently incorporated the dueling architecture into the value network and experimentally evaluated its effects. The modified model exhibited improved training stability. I sincerely thank Mr. Zhu for his guidance and constructive suggestions on theoretical learning, research direction, and manuscript revision.

All experimental data reported in this paper were generated and collected by the author using the equivalent-circuit simulator developed for this project. No external experimental datasets, proprietary measurement datasets, or third-party training data were used. All results were derived from the simulated filter responses and reinforcement learning trajectories generated during the author's experiments. Apart from the supervisor, no other teachers or students participated in experimental design, code development, data analysis, or manuscript preparation.

The project was conducted as follows. In January 2026, I read news reports about AI-driven drug molecule design and subsequently discussed this technological development with Mr. Zhu outside class. He explained that AI could also be used to design a wide range of engineering systems, including circuits in mobile communication devices. This discussion motivated me to explore AI-assisted circuit optimization.

From January to March 2026, I learned about the role of RF filters in mobile communication systems and how the number, parameters, and topology of circuit components affect filter performance. I also learned that equivalent-circuit models can evaluate filter responses under different design settings at relatively low computational cost. In April, I studied the modified Butterworth–Van Dyke (mBVD) model and the ABCD matrix cascading method for two-port networks.

In May, I conducted preliminary experiments using a continuous action space and proximal policy optimization, but the results were unsatisfactory. After Mr. Zhu pointed out that an overly unconstrained action space could increase learning difficulty, I reviewed the relevant circuit design literature and independently developed an initial expert action set in June. He subsequently suggested adding complementary operations to ensure sufficient exploration of the parameter space, and I revised and finalized the action set accordingly. By July, the agent could identify promising filter parameters, but training scores still fluctuated substantially. Mr. Zhu recommended further study of the dueling architecture to improve value estimation. I then completed the relevant literature study, network implementation, comparative experiments, and result analysis. By August 2026, the experimental work and manuscript preparation were largely complete.

2026 S. T. Yau High School Science Award 仅用于2026丘成桐中学科学奖公开

Statement on the Use of Artificial Intelligence

The AI tool used in this project was ChatGPT 5.6. Its use was concentrated primarily in the early learning stage, where it served as an auxiliary learning tool to help retrieve, explain, and understand prerequisite knowledge, including RF filter principles, equivalent-circuit modeling, reinforcement learning fundamentals, and related background concepts. During manuscript preparation, AI was also used to assist with translation, grammar checking, and language polishing. In addition, some operations suggested by a large language model were adopted during expert action design, followed by simulation testing and adjustment of the specific parameter-update magnitudes.

ChatGPT 5.6 was not used to determine the research topic, construct the theoretical or methodological framework, develop or improve the algorithm, write experimental code, generate or collect experimental data, conduct experiments, analyze results, formulate conclusions, or write the substantive content of the paper on the author's behalf. These core tasks were independently completed by the author under the academic guidance of the supervisor.

2026 S. -T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示