forum acusticum 2026
8-12 September · Graz, Austria

GENERALIZATION OF DEEP REINFORCEMENT LEARNING FOR THE DESIGN OF BROADBAND BANDGAP STRUCTURES

Semere B. Gebrekidan[1], Florian Toth[1]

[1] *Institute of Mechanics and Mechatronics, TU Wien, Austria*

This paper is a revised version of the authors' submitted manuscript that was peer-reviewed by at least two anonymous reviewers.

Abstract

Structures exhibiting broadband bandgaps play a critical role in controlling wave propagation for applications such as vibration isolation and acoustic filtering. To design such structures, topology optimization methods have been widely employed to maximize bandgap performance; however, these approaches are computationally intensive and often limited in scalability. Data-driven approaches can accelerate design discovery but remain constrained by the distribution of training data and frequently exhibit limited generalization to unseen design environments. In this work, we propose a discrete soft actor–critic (SAC)-based deep reinforcement learning framework to autonomously design elastic structures with broadband bandgaps. The agent is first trained to optimize the bandgap in a 4 cm × 4 cm design space and subsequently evaluated in a smaller 3 cm × 3 cm environment to assess its generalization capability. The results demonstrate that the proposed reinforcement learning framework successfully discovers structures with broadband bandgaps and generalizes to unseen design domains without additional training. Full-wave harmonic simulations confirm that the bandgap predictions obtained from unit-cell analysis are consistent with the frequency response of the corresponding periodic structures. Compared with conventional generative modeling approaches, the proposed method learns the underlying relationship between geometry and bandgap formation through environment interaction, enabling improved versatility and physics-aware design discovery.

Keywords: *optimization, bandgap, Deep reinforcement learning, generalize, broadband*

Corresponding author: *semere.gebrekidan@tuwien.ac.at.*

1. Introduction

Structures exhibiting bandgap characteristics can block the propagation of waves within specific frequency ranges, making them attractive for applications such as noise and vibration isolation, topological insulators, and related wave-control technologies. Bandgaps generally originate from Bragg scattering, where destructive interference occurs between incident and reflected waves, or from local resonance mechanisms. In particular, bandgap formation in plate-like elastic structures is primarily associated with local resonance, where elastic wave propagation is suppressed through localized deformation. This mechanism differs from the bandgap formation observed in conventional acoustic phononic crystals. Elastic bandgap structures capable of blocking various types of elastic waves, including longitudinal, transverse, and flexural waves, have been proposed [1]. The design of such structures commonly relies on Bloch wave analysis, in which a unit cell with Floquet periodic boundary conditions is considered, and the resulting dispersion curves are used to identify eigenfrequency ranges where wave propagation is prohibited. In this context, topology optimization has been widely employed to determine optimal material distributions that maximize bandgap widths at targeted frequencies. By applying Floquet boundary conditions to a unit-cell model, the gap between two consecutive dispersion curves can be maximized to achieve improved bandgap performance [2], [3], [4]. However, the flexibility of the method is preceded by the requirement of a significant computational cost, particularly when evaluating small changes in objective functions, for instance a change in bandgap frequency for a given design. To mitigate this computational challenge, deep learning methods have recently been integrated with topology optimization to accelerate design discovery [5], [6], [7], [8]. In addition, deep learning has been applied to both forward and inverse problems: predicting the bandgap of a given material structure [6], [9] and, conversely, generating structures that yield a desired bandgap [10], [11], [12]. However, these predictive models are generally constrained by the distribution of the training dataset, limiting their ability to discover fundamentally new structures.

Generative models, such as variational autoencoders (VAEs) and generative adversarial networks (GANs), aim to overcome this limitation by generating new structures through learned latent representations and extrapolation beyond the training data [11], [12]. However, their generalization capability remains limited, particularly when predicting bandgap characteristics for entirely new structures with dimensions that differ from those included in the training dataset. Deep reinforcement learning (DRL) has demonstrated enhanced generalization capability in complex and high-dimensional environments [13]. This concept has been extended to the design of mechanical metamaterials [14] and noise-mitigation structures, where successful generalization to unseen design conditions has been demonstrated [15].

In this work, we demonstrate that a DRL agent trained within a 4 cm × 4 cm grid environment can learn to construct designs that maximize bandgap performance and subsequently generalize to a completely new environment. Specifically, the trained agent is evaluated in a 3 cm × 3 cm design space, where it successfully discovers structures exhibiting broadband bandgaps without additional training. These results demonstrate the versatility and generalization capability of DRL for the autonomous discovery of bandgap structures.

2. Methods

Deep reinforcement learning is an unsupervised learning approach where an agent learns by interacting with an environment and selecting actions to maximize cumulative rewards. Recently, soft actor-critic(SAC) deep reinforcement learning algorithms have demonstrated strong generalization capabilities in complex environments [13]. Instead of directly maximizing the reward function, the SAC algorithm optimizes the entropy-regularized objective, encouraging exploration and improving generalization in complex environments. Due to the discrete nature of the grid-based environment, we adopt a discrete SAC formulation [16]. The SAC framework consists of an environment, states, actions, and rewards, which are described as follows.

Environment: The environment corresponds to the discrete grid representation of the design space, as shown in Figure 1(left). The 4 cm × 4 cm design domain is discretized into a 40 × 40 grid for training, while the 3 cm × 3 cm domain is discretized into a 30 × 30 grid to evaluate the generalization capability of the trained agent. By exploiting geometric symmetry, only one-eighth of the full grid is used to define the design environment. Since the diagonal symmetry line intersects the grid, some grid elements partially extend outside the triangular design region, as shown in Figure 1(middle). These elements are retained to ensure connectivity during the reconstruction of the complete geometry.

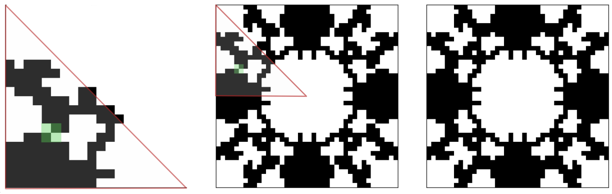

Figure 1: Representation of the design environment. The red triangle denotes the environment used by DRL agent (left). Diagonal, vertical, and horizontal symmetry operations are applied to reconstruct the complete grid (middle). The reconstructed grid is subsequently post-processed to eliminate point connections between adjacent black and white regions, as indicated by the green rectangle (right).

States: Each grid has two states represented in terms of position of the grids in x and y coordinates. The agent should be able to select a sequence of grids to maximum bandgap.

Actions: The action space consists of four directions, i.e., up, down, left, and right, where the agent moves to select the next grid based on the initial grid. Diagonal selection is prohibited. However, point connections may occur during the agent's exploration when previously selected grid elements are revisited or when grid elements along the diagonal symmetry line remain unselected before applying the full rotational symmetry operation. These undesired point connections are eliminated through a post-processing step using a 2×2 filter, which identifies point-connected regions and converts them into solid regions, as shown in Figure 1(right), thereby generating a fully connected solid structure for the simulation.

Rewards: The agent is allowed to explore the environment for 500 steps for each episode. Since the bandgap is evaluated at the end of every episode, we first define intermediate rewards to encourage exploration, corner selection, and diagonal connectivity. An intermediate reward of 0.5 is assigned for encouraging exploration, 0.4 for stepping to a corner to ensure periodicity along x and y directions, and 0.1 to ensure diagonal connectivity due to symmetry, while corresponding negative values are used to penalize the agent at each step.

After completing 500 steps, the selected grid elements are reconstructed using diagonal and vertical symmetries and a simulation is carried out in COMSOL Multiphysics to compute the dispersion curves of the resulting structure. The final reward function is formulated to maximize the bandgap between two consecutive bands, namely the lower j and upper $(j+1)$ bands, evaluated at 32 sampling points along the irreducible Brillouin zone, defined as,

$$J = \frac{2}{N}\left(\min \Omega^2_{j+1,k} - \max \Omega^2_{j,k}\right)/1000, \tag{1}$$

$$\text{reward} = \begin{cases} 10J, & \text{if } J > 0 \\ J, & \text{otherwise} \end{cases} \tag{2}$$

where Ω denotes the angular frequency, j is the band index, k represents the sampling points along the irreducible Brillouin zone, and N is the total number of bands considered ($N = 20$). A higher weighting factor is assigned to the final reward to encourage the agent to explore states that maximize the bandgap, while the intermediate rewards guide the agent toward generating physically meaningful and geometrically valid structures. In this work, the objective is to maximize the overall bandgap for the first 20 bands without targeting a specific frequency range. The SAC framework consists of an actor network, which selects actions according to the learned policy, and a critic network, which evaluates the value of the state-action pair.

Agent interaction with environment: The agent starts randomly from the lower grid. Based on the current state, at each step it selects an action from the action space and transitions to the next state, where intermediate rewards are used to encourage the agent to explore the environment and form a valid well-connected geometry for 500 steps, which is one episode. As shown in Figure 1(middle), at the end of each episode the sequence of grids selected by the agent is rotated along the symmetry to form a full geometry. To ensure connectivity, a 2 by 2 filter is used to convert point-connected neighboring air and solid grids into solid grids, as shown in Figure 1(right). Using the `mph` Python library, which enables access to COMSOL functionality via Python, the grids are translated into the COMSOL environment. An eigenfrequency analysis is then performed to compute the first 20 dispersion bands along the boundary of the reduced Brillouin zone. A total of 32 sampling points (denoted by k) are used for each band(j), and periodic boundary conditions are applied in both the x and y directions. The resulting bandgaps are used to calculate the final reward. This interaction process is carried out for 3000 episodes, in which each simulation in COMSOL takes approximately 2 minutes. For the case of testing on the 3 cm $\times$ 3 cm design domain, we carried out 1000 episodes to verify the agents generalization capability.

Network architecture: The SAC algorithm consists of one actor network, two critic networks, and two target critic networks. The actor network comprises three fully connected hidden layers with 80 neurons each. A ReLU activation function is used in the hidden layers, while a softmax activation function is applied in the output layer to predict action probabilities. The critic networks share a similar architecture, except that the final layer uses a linear activation function to estimate the Q-values. Experiences are stored in a replay buffer, where 80% of the sampled transitions are prioritized from experiences with positive rewards. A batch size of 128 is used for training, and the agent performs five parameter updates per episode. The Adam optimizer is employed with a learning rate of 6×10^{-4}. The target networks are updated using soft updates with a coefficient $\tau = 10^{-3}$.

3. Results and Discussions

The hyperparameters of the agent are selected based on trial and error. For optimizing the bandgap of the 4 cm $\times$ 4 cm structure shown in Figure 2a, the agent explores the environment for the first 1000 episodes (i.e., randomly selecting actions from the action space) and learns for the remaining episodes (i.e., using the actor network to predict actions). Positive rewards are associated with structures exhibiting broadband bandgaps. The moving average of the reward over 3000 episodes demonstrates that the agent successfully learns to design structures with positive rewards. Further inspection of the optimized designs shows that the agent is able to generate structures with broadband bandgaps, as shown in Figure 2b and c.

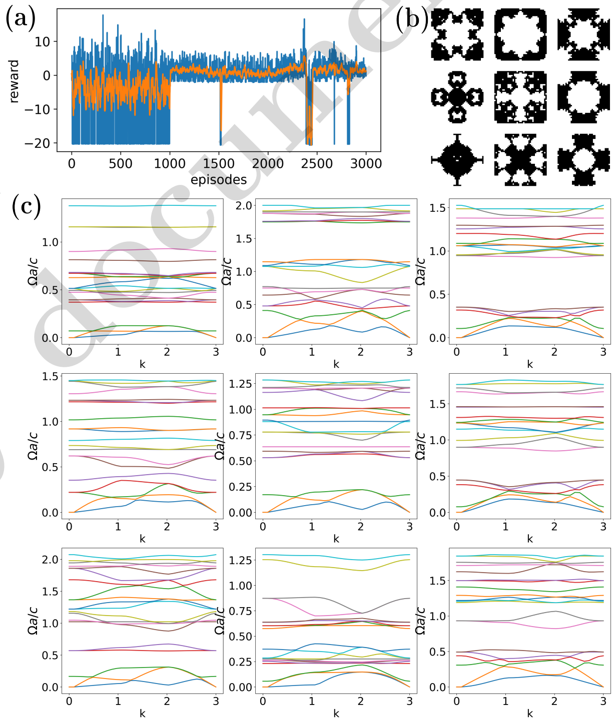

Figure 2: Results for the 4 cm $\times$ 4 cm design domain: (a) the average reward per episode and the moving average reward over 10 episodes (shown in orange); (b) designs with positive reward; and (c) their corresponding normalized dispersion curves. a and c represent the size of the design domain and the speed of sound in air (343 m/s), respectively.

To analyze the generalization capability of the trained agent, we evaluate its performance in discovering structures with broadband bandgaps in the 3 cm $\times$ 3 cm design space. The moving average of the reward demonstrates positive values, as shown in Figure 3a, indicating that the agent successfully adapts to the new environment. Despite being trained in a different environment, the agent is able to construct designs with broadband bandgaps. Randomly selected structures and their corresponding bandgaps are shown in

Figure 3b and c, demonstrating that the agent learns meaningful structural configurations that yield positive rewards. The generated designs exhibit similar geometric features, particularly cross-shaped patterns, with variations primarily occurring in the material distribution near the central region. These results suggest that the trained agent identifies that material placement near the diagonal symmetry regions contributes to broader bandgaps through local resonance mechanisms enabled by thin connecting features. The ability to discover similar effective configurations in a different design domain demonstrates the generalization capability of the proposed approach and highlights its potential for the design of vibration-mitigation structures. Furthermore, due to the biased sampling strategy introduced during experience replay, the agent tends to converge toward similar design configurations, resulting in limited design diversity.

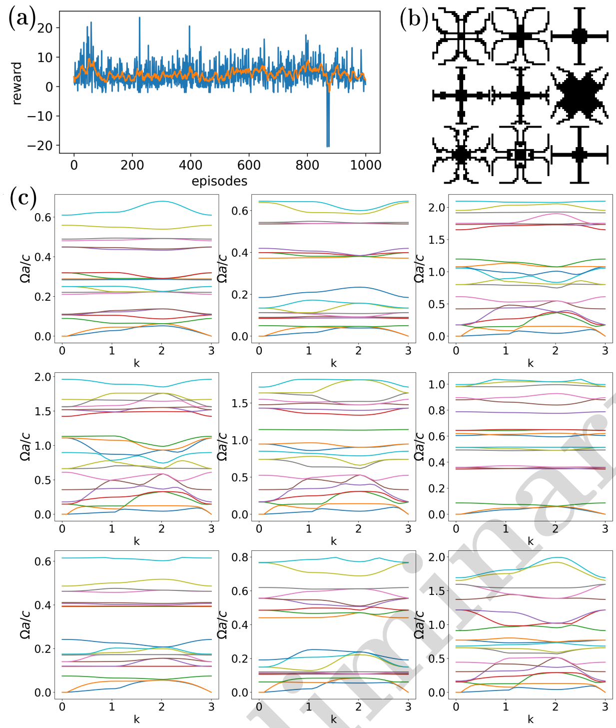

Figure 3: Validation in a new environment for the 3 cm × 3 cm design space. (a) the average reward per episode and the moving average reward over 10 episodes (shown in orange); (b) design samples with positive reward; and (c) the corresponding normalized dispersion curves of the designs. a and c represent the size of the design domain and the speed of sound in air (343 m/s), respectively.

To verify the bandgaps predicted from unit-cell dispersion analysis, full-wave harmonic simulations are performed for both design environments. The unit cells are periodically arranged along the x and y directions, with four unit cells repeated in each direction, as shown in Figure 4a and b (middle). The structure is excited by applying a prescribed displacement of 0.01 m along the left boundary using the solid mechanics module in COMSOL Multiphysics, generating

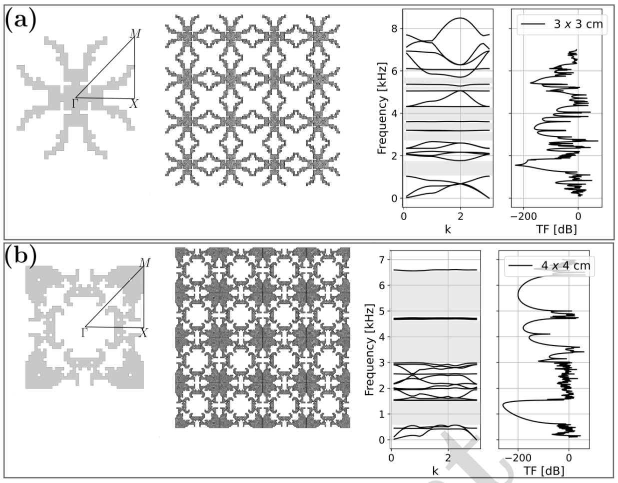

Figure 4: From left to right: unit cells and their representation in the irreducible Brillouin zone, periodically arranged unit cells for full-wave simulations, the dispersion curves of the unit cells, and the transfer functions obtained from full-wave simulations for (a) the 3 cm × 3 cm design space and (b) the 4 cm × 4 cm design space.

in-plane propagating longitudinal waves. The displacement magnitude at the right boundary is then used to evaluate the transfer function. Periodic boundary conditions are applied to the upper and lower boundaries. The harmonic analysis is conducted over a frequency range from 100 Hz to 7000 Hz with a frequency increment of 5 Hz. The logarithmic transfer function (TF), defined as $20\log(d_{\text{out}}/d_{\text{in}})$, is used to quantify wave transmission and verify the predicted bandgaps. Here, d_{in} and d_{out} represent the input displacement at the left boundary and the output displacement at the right boundary, respectively. As shown in Figure 4a (right), the full-wave harmonic analysis demonstrates negligible transmission within the bandgap regions predicted by the unit-cell analysis for 3 cm × 3 cm unit cell. Similarly, for the 4 cm × 4 cm design, the TF shows a significant reduction of propagating wave at the predicted bandgap frequencies (Figure 4b , middle), which is consistent with the unit-cell band structure results. It should be noted that, for finite structures, the observed bandgap frequencies may differ from those obtained under periodic boundary conditions due to the influence of boundary conditions. Furthermore, the predicted bandgaps are based on the assumption of purely in-plane displacement, where out-of-plane motion is neglected. For three-dimensional thin structures, additional deformation modes, such as bending-induced out-of-plane displacement, may influence the wave propagation characteristics. Therefore, further investigation is required to evaluate these effects. Although the proposed framework accelerates the discovery of bandgap structures, the current training and testing procedures still require repeated multiphysics simulations. Replacing these simulations with surrogate models could further reduce the computational cost and enable more efficient design exploration.

4. Conclusions

In summary, we demonstrate that the discrete SAC-based deep reinforcement learning framework is capable of autonomously designing elastic structures with broadband bandgaps and generalizing to unseen design environments without additional training. Compared with conventional generative modeling and other deep learning-based approaches, the proposed method provides enhanced versatility and generalization capability for discovering structures beyond the original training domain. Unlike data-driven generative methods, whose performance is often constrained by the distribution of the training dataset and relies on extrapolation from previously observed designs, the proposed reinforcement learning framework learns the relationship between structural geometry and bandgap formation through direct interaction with the design environment. This enables more flexible, physics-informed discovery of structures with a wide bandgap beyond the limitations imposed by predefined datasets.

References

[1] S. Zouari, J. Brocail, and J.-M. Génevaux, "Flexural wave band gaps in metamaterial plates: A numerical and experimental study from infinite to finite models," *Journal of Sound and Vibration*, vol. 435, pp. 246–263, 2018.

[2] O. Sigmund and J. Søndergaard Jensen, "Systematic design of phononic band–gap materials and structures by topology optimization," *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 361, no. 1806, pp. 1001–1019, 2003.

[3] D. Giannini, M. Schevenels, and E. P. Reynders, "Topology optimization design of multimodal resonators for metamaterial panels with maximized broadband vibroacoustic attenuation," *Journal of Sound and Vibration*, vol. 595, p. 118 691, 2025.

[4] V. Cool, O. Sigmund, and N. Aage, "Metamaterial design with vibroacoustic bandgaps through topology optimization," *Computer Methods in Applied Mechanics and Engineering*, vol. 436, p. 117 744, 2025.

[5] N. A. Kallioras, G. Kazakis, and N. D. Lagaros, "Accelerated topology optimization by means of deep learning," *Structural and Multidisciplinary Optimization*, vol. 62, no. 3, pp. 1185–1212, 2020.

[6] X. Li, S. Ning, Z. Liu, Z. Yan, C. Luo, and Z. Zhuang, "Designing phononic crystal with anticipated band gap through a deep learning based data-driven method," *Computer Methods in Applied Mechanics and Engineering*, vol. 361, p. 112 737, 2020.

[7] C. Deng, Y. Wang, C. Qin, Y. Fu, and W. Lu, "Self-directed online machine learning for topology optimization," *Nature communications*, vol. 13, no. 1, p. 388, 2022.

[8] A. Chandrasekhar and K. Suresh, "Tounn: Topology optimization using neural networks," *Structural and Multidisciplinary Optimization*, vol. 63, no. 3, pp. 1135–1149, 2021.

[9] Z. Shi, E. Tsymbalov, M. Dao, S. Suresh, A. Shapeev, and J. Li, "Deep elastic strain engineering of bandgap through machine learning," *Proceedings of the National Academy of Sciences*, vol. 116, no. 10, pp. 4117–4122, 2019.

[10] A. C. Ogren, B. T. Feng, K. L. Bouman, and C. Daraio, "Gaussian process regression as a surrogate model for the computation of dispersion relations," *Computer Methods in Applied Mechanics and Engineering*, vol. 420, p. 116 661, 2024.

[11] T. V. Tran, S. Nanthakumar, and X. Zhuang, "Deep learning-based framework for the on-demand inverse design of metamaterials with arbitrary target band gap," *npj Artificial Intelligence*, vol. 1, no. 1, p. 2, 2025.

[12] Z. Wang, W. Xian, M. R. Baccouche, H. Lanzerath, Y. Li, and H. Xu, "A gaussian mixture variational autoencoder-based approach for designing phononic bandgap metamaterials," in *International design engineering technical conferences and computers and information in engineering conference*, American Society of Mechanical Engineers, vol. 85390, 2021, V03BT03A002.

[13] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *International conference on machine learning*, Pmlr, 2018, pp. 1861–1870.

[14] N. K. Brown, A. P. Garland, G. M. Fadel, and G. Li, "Deep reinforcement learning for the rapid on-demand design of mechanical metamaterials with targeted nonlinear deformation responses," *Engineering Applications of Artificial Intelligence*, vol. 126, p. 106 998, 2023.

[15] S. B. Gebrekidan and S. Marburg, "Autonomous design of noise-mitigating structures using deep reinforcement learning," *The Journal of the Acoustical Society of America*, vol. 156, no. 1, pp. 151–163, 2024.

[16] P. Christodoulou, "Soft actor-critic for discrete action settings," *arXiv preprint arXiv:1910.07207*, 2019.