XI ECCOMAS Thematic Conference on Smart Structures and Materials
SMART 2025
M. Krommer, M. Nader and M. Schagerl (Eds)

A CASE STUDY IN SEMI-ACTIVE STRUCTURAL CONTROL BASED ON MULTI-AGENT REINFORCEMENT LEARNING

A. BARBAY, D. PISARSKI, G. MIKUŁOWSKI, B. BŁACHOWSKI, Ł. JANKOWSKI*

∗ Institute of Fundamental Technological Research, Polish Academy of Sciences
Pawińskiego 5B, 02-106, Warsaw, Poland
e-mail: ajedlins@ippt.pan.pl, ljank@ippt.pan.pl

Abstract: This study applies the machine learning technique of multi-agent reinforcement learning for semi-active structural control. The considered structure is a high-rise shear-type building subjected to seismic excitation, where actuators function as viscous dampers with a controllable level of damping. The problem formulation is inherently nonlinear due to the bilinear nature of the control concept. The analytical derivation of optimal semi-active control solutions is seldom feasible, leading to many practical control algorithms being suboptimal and/or heuristic in their formulation. In the framework proposed here, the control algorithm is developed through interaction with the controlled system by applying actions, observing results, and optimizing effects. For this purpose, a multi-agent reinforcement learning Q-network architecture is employed. Verification is conducted through a numerical experiment, utilizing a finite element model of a structure equipped with a tuned mass damper and a controllable viscous damper. The results demonstrate that the proposed method outperforms a conventional, optimally tuned TMD. This study also includes a comparison of the obtained results with those achieved using the typical single-agent approach. The key contribution is to demonstrate a significant improvement in control performance through the application of a multi-agent policy.

Key words: Structural control, Reinforcement Learning, Multi-agent, Damping, Vibration

1 INTRODUCTION

Recent years have seen significant advances in the field of reinforcement learning (RL), which has been very successful in solving various sequential decision problems in machine learning. Most successful applications of RL, e.g., in games [1], robotics [2], autonomous driving [3], engineering [4], and logistics [5] require the participation of one or more agents, which naturally belongs to the field of multi-agent RL (MARL), which has recently emerged due to advances in single-agent RL techniques. Despite its empirical success, the theoretical foundations of multi-agent approaches are relatively scarce in the literature, especially in the field of structural control. The single-agent approach is more widely researched [6]. This contribution presents research on a collaborative multi-agent control approach of an engineering structure experiencing seismic excitation under increasing levels of measurement noise (0–50%) and compares it to the performance of seven individual agents.

Vibrations in civil engineering are dangerous and can negatively influence the operation of engineering structures. One of the most dangerous phenomena is an earthquake, which may bring a very destructive range of mechanical vibrations [7]. Buildings experience damage or destruction, with the associated risks for people. Various techniques for structural control and the mitigation of structural vibrations have been investigated. Active methods have high efficiency and are widely used in many industries [8]. Active control requires the generation of external forces, which creates significant power demands. In civil engineering, this demand can become too great to be met during a major seismic event. A typical approach is passive control, which focuses on mitigating vibrations through structural optimization and energy dissipation mechanisms. This approach relies on passive devices such as Tuned Mass Absorbers (TMAs) and Tuned Mass Dampers (TMDs), which are mechanical resonance-based components designed to reduce vibration. These devices take the vibration energy and dissipate it internally, reducing excessive motion. Among the key advantages of TMAs and TMDs is that they do not require altering an existing structure or disrupting its basic functionality, making them an optimal choice for upgrades and troubleshooting. An intermediate solution between the active and passive techniques is semi-active control methods, which focus mainly on dissipating energy rather than generating large external forces. These approaches work by locally adjusting the mechanical properties of specific components, such as modifying viscous damping or stiffness. The generated forces are purely dissipative and opposed to the motion of the structure, large power sources are thus not required, which reduces the risk of instability. Nevertheless, due to their nonlinear nature, deriving optimal control can be analytically complex.

This paper describes research on the use of a semi-active tuned mass damper (TMD) as an actuator and switching control as a structural control strategy. The aim is to improve control quality by using a multi-agent reinforcement learning approach. A TMD is a device installed on structures to reduce mechanical vibrations, consisting of a mass mounted on one or more damped springs [9]. Numerous modifications of the classical TMD approach have been introduced and tested [10]. The effectiveness of the TMD in vibration control can be increased by adjusting several key parameters, such as the mass ratio, the natural frequency ratio between the TMD and the main structure, and the TMD damping ratio. In the case considered here, the TMD is controlled by switching its viscous damping coefficient in an on/off manner (between very low and very high damping). This results in the control signal directly affecting the damping matrix in the equation of motion in a linear manner, which corresponds to bilinear control. Analytical solutions for this type of control are often not available. For open-loop control, a bang-bang strategy is often optimal, which involves switching between two extreme states. In the proposed methodology, the switching points are determined by reinforcement learning (RL) machine learning algorithms. The model learns by interacting with the environment, i.e., through a series of trial and error in a simulated structure subjected to random seismic excitation. The mentioned approach remains novel in the context of using multiple agents. Good results have been achieved in the approach of using a single agent [11]. Also, research is available on the robustness under measurement noise, where the performance was checked in a noisy environment simulating the natural measurement processes [12]. This has shown the potential for further development in this approach, as the evaluation of the model consisted of testing its performance in an environment closer to reality than an ideal mathematical model. The tests considered here examine the effectiveness of 7 separate agents and a multi-agent voting-based architecture in a range of different levels of measurement noise.

The target function quantifies the effectiveness of the control, and it is based on the maximum inter-story displacement and the total energy of the structure. The results show the higher effectiveness of the multi-agent approach as compared to the single-agents operating alone. Applied together, they form an integrated model that allows the structure to be controlled in a more effective way.

2 REINFORCEMENT LEARNING

In contrast to such popular machine learning paradigms as supervised and unsupervised learning, reinforcement learning (RL) is a distinct approach based on sequential decision-making and trial-and-error interactions with the environment [13]. When applied to the adaptive control of highly complex and dynamic systems, the idea appears to be extremely effective. Formally speaking, RL can be defined as a Markov decision process (MDP). Reinforcement learning algorithms encourage the AI agent to explore an optimal decision chain and define "correct behaviour" in the considered environment.

This research explores the potential of reinforcement learning (RL) in improving semi-active structural control. In this study, an RL agent is utilized, which employs a dense artificial neural network to learn and encode the Q-function [14]. Q-learning and DQN [15] are popular RL methods and have been previously applied in multi-agent settings [16]. Q-Learning uses an action-value function under the policy π defined as the following expected value [17]:

$$Q^{\pi}(s, a) = E[R/s^t = s, a^t = a] \tag{1}$$

where R is a total reward received after taking action a in state s and s^t, a^t are the state and action at a specific timestep t. However, this Q function can be rewritten recursively as an expectation over the next state s' and next action a',

$$Q^{\pi}(s, a) = E_{S'}[r(s, a) + \gamma E[Q^{\pi}(s', \pi(s'))]] \tag{2}$$

where $r(s, a)$ is the immediate reward received when taking action a in state s, and γ is a discount factor (between 0 and 1) that controls how much future rewards are valued compared to immediate rewards. The DQN learns the action-value function Q (maximum estimated Q-value for the next state s'), which corresponds with the optimal policy by a loss minimization. The formula below (Loss function of the Deep Q-Network) shows how different the predicted Q-values are from the target y:

$$L(\theta) = E_{s,a}[(Q(s, a/\theta) - y)^2], \qquad \text{where} \qquad y = r + \gamma \max_{a'} \bar{Q}(s', a'), \tag{3}$$

where y as a target value that follows the Bellman equation, where future rewards are recursively considered. It contains the maximum estimated Q-value for the next state s'. $\bar{Q}$ is a target Q-function with parameters updated periodically with the latest θ to help stabilize learning, where θ are the neural network parameters. Another key element in stabilizing DQN is the use of an experience replay buffer D containing tuples of (s, a, r, s'). By directly applying Q-Learning to multi-agent settings, each agent learns its optimal Q function independently [15]. It is important to note that the approach presented here is inherently probabilistic, as the expected value is drawn from a probability distribution. However, by selecting the maximum value, we effectively transform this method into a deterministic variant [15].

3 MULTI-AGENT REINFORCEMENT LEARNING APPROACH

There are a number of important applications that involve interaction between multiple agents, where emergent behaviour and complexity arise from agents cooperating together. Multi-agent reinforcement learning (MARL) uses a variety of training strategies, often using centralized, decentralized, or hybrid approaches [18]. The training process typically involves the interaction of multiple agents in a shared environment, who learn simultaneously or in sequential phases, however, this approach is not always appropriate for the application.

Centralized MARL is based on agents who are trained together, sharing experiences and learning from each other's experiences. This approach allows agents to learn from the actions and rewards of other agents, which promotes coordinated behaviour and develops cooperative strategies. The centralized approach requires more resources due to shared data and joint policy optimization. The policy example can be a Multi-Agent Deep Deterministic Policy Gradient (MADDPG) – which uses centralized critics during training but decentralized policies during execution [17]. Another example is the QMIX – a value-based approach where a mixing network combines agent-specific Q-values into a global Q-value [19]. The applications of centralized MARL are usually connected with a need for the joint policy of agents (e.g., robotic swarm [20], coordination of multi-drone navigation [21].

In decentralized MARL, each agent learns independently using its observations and rewards without direct access to other agents' information. Each agent optimizes its policy based on its local observations and rewards. The approach is more efficient when dealing with a large number of agents, however, if agents do not have a global knowledge, they might develop conflicting policies. The main examples of policies are: Independent Q-Learning (IQL) – each agent learns its Q-function without considering other agents [22] and Decentralized PPO (Proximal Policy Optimization) – used in environments where agents act independently [23]. Decentralized training is usually used when agents operate in highly dynamic environments (e.g., competitive games, self-driving cars in traffic) and when individual agent policies need to be robust to environmental changes.

Some methods combine centralized and decentralized approaches to balance coordination and scalability. These include: Meta-Learning Strategies [24], where agents learn how to switch between centralized and decentralized modes, and Hierarchical MARL [25], where a central controller guides decentralized agents in complex tasks.

In this contribution, the decentralized MARL approach was used. Agents were trained independently, achieving their own optimal polices over time. Then, they were used in one environment to collaborate and execute the optimal control algorithm with majority voting as a decision strategy.

4 STRUCTURE

Any interaction between the agent and the environment requires continuous feedback from the structure. For testing purposes, it was necessary to create a numerical environment reflecting the nature of the structure and its properties. Due to security reasons, the tests could not be performed on the actual physical structure. As a numerically modeled structure is used, deviations from the physical model are expected. The dynamics of the structure is approximate, and the model's robustness to measurement errors and noise must be taken into account. As shown in [12], the control approach developed here is relatively insensitive to these errors. This contribution also shows how the multi-agent approach manages the measurement noise case. The structure is modeled as a 35-story shear-type building with a single degree of freedom (DOF) per story, equipped with a semi-active TMD at the top. Three areas of measurement are specified as floors 19, 35, and TMD, where the measurements are the displacement and velocity of each pointed floor (Fig. 1).

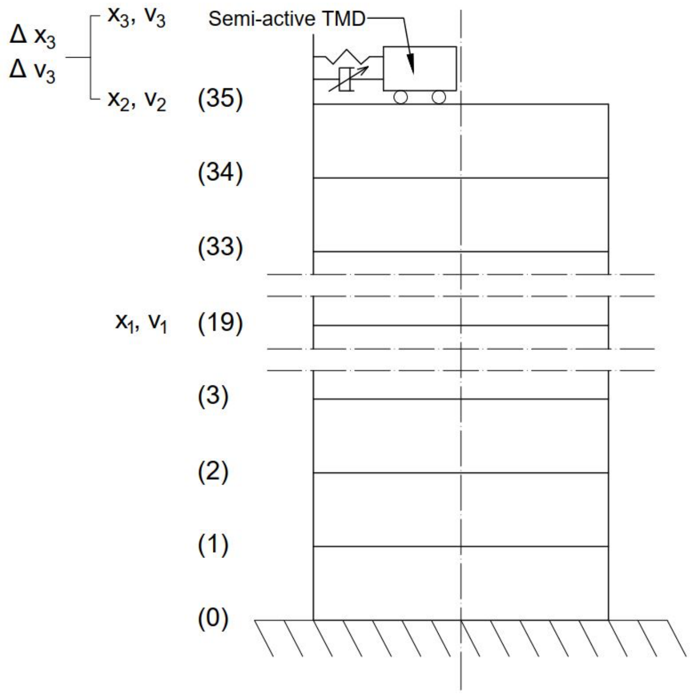

Figure 1: The investigated 35-DOF shear structure with a TMD at the top level

The equation of motion is defined using 36 degrees of freedom (DOF), the seismic excitation being modeled as lateral acceleration, a(t). It is expressed as follows:

$$[M]\{u\} + [C]\{u\} + [K]\{u\} = -[M]\{r\}a(t) \tag{4}$$

The matrices *[M]*, *[C]*, and *[K]* correspond to mass, damping, and stiffness, respectively. The control strategy applied is bang-bang control, which regulates the TMD dashpot by alternating between two extreme states—zero and maximum damping—while *[C]* also considers the inherent damping of the building. A material damping model is used to establish the critical damping ratio for the building's first natural vibration mode without TMD, which is set at 2%.

The lateral load distribution is described by a 36-element vector filled with ones. The Tuned Mass Damper (TMD) comprises 3% of the building's total mass. The floor masses and stiffness values are taken from a reference structure mentioned in the literature [26]. The natural frequencies of the structure range from 0.887 Hz to 14.917 Hz.

Artificial intelligence models have a tendency to overfit specific patterns found in the data. To ensure that the model learns to control the structure effectively, the applied excitation must be diverse and cover a broad frequency spectrum, resembling real earthquake scenarios. To avoid the control system adapting to just one ground motion pattern, seismic excitation was modeled as white noise at each time step. The ground acceleration value a(t) was continuously sampled from a uniform distribution centered around zero. The feedback signal employed 6 measurements shown in Fig. 1 (displacements and velocities at floor 19, 35, and the TMD). This approach was used to evaluate the performance of both the control system and the reinforcement learning algorithm.

5 CONTROL WITH AN INDIVIDUAL AND MULTI-AGENT STRATEGY

Each agent receives local observations, including displacement and velocity measurements. Seven agents were trained using observations from 6 sensors (x_1, v_1, x_2, v_2, Δx_3, Δv_3) shown in Fig. 1. The target function quantifies the maximum inter-story displacement and the total energy of the structure, while the reward additionally promotes the restriction of damping control during high relative velocities between the TMD and the top floor. All the considered values are normalized with respect to the same quantities obtained with an optimally tuned passive TMD for the same seismic excitation. Noise levels range from 0% to 50% in 5% increments. When simulating with a multi-agent approach, agents make independent decisions and then apply a majority vote. The simulation phase of each agent is carried out using a dense ANN consisting of an input layer, 2 dense hidden layers (each with 15 neurons), and an output layer. A graphical representation of the ANN is presented in Fig. 2. The activation function used in the neural network is a leaky rectified linear unit (Leaky ReLU). Each test consisted of 1000 test episodes at each noise level. Metrics are the mean target function values for each agent and the multi-agent system and a standard deviation for the multi-agent system.

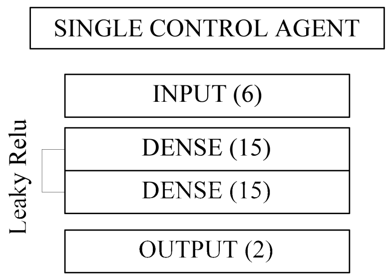

Figure 2: Single RL control agent artificial neural network architecture

7 RESULTS

7.1 Mean target function values across different noise levels

The results plotted in Fig. 3 show the mean target function values across different noise levels:

- The multi-agent architecture consistently achieves better (lower) target function values compared to individual agents, indicating a better performance.
- The standard deviation band (±1σ) of the multi-agent system contains individual agents' results, but the multi-agent approach remains below them, which confirms its effectiveness.
- As noise increases, all systems remain stable, but the multi-agent system consistently outperforms all individual agents.

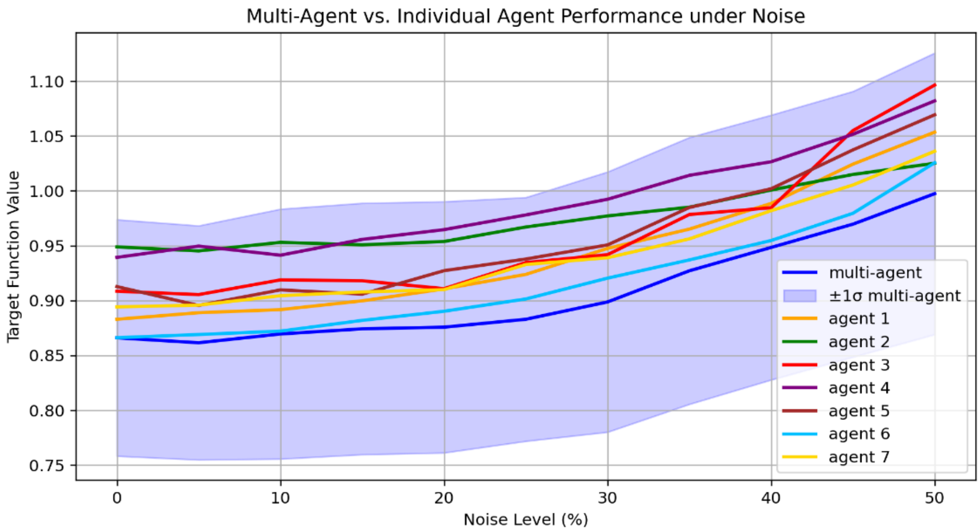

Figure 3: Multi-agent vs. individual agent performance under noise

7.2. Performance distribution of the multi-agent system and individual agents

The box plots in Fig. 4 and Fig. 5 compare the performance distribution of the multi-agent system and individual agents at 0% noise and 50% noise. In Fig. 4, agents' performance distribution for 0% noise level is represented:

- The multi-agent system achieves the lowest median target function value, meaning it outperforms individual agents in a noise-free environment.
- The variance in performance is relatively small, suggesting consistent behaviour.
- Individual agents have their distributions slightly overlapping with the multi-agent system, but all their median values are higher.

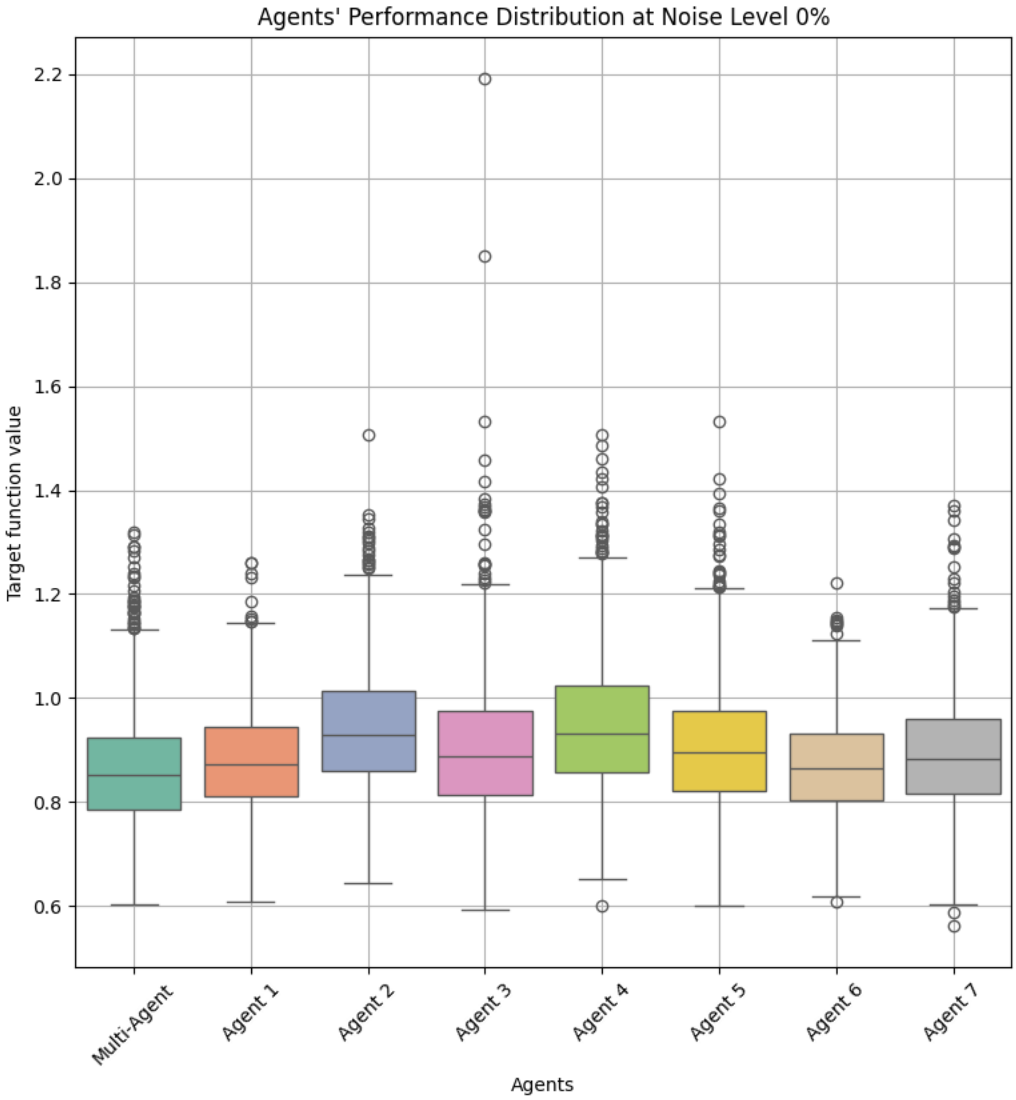

Figure 4: Agents' performance distribution for 0% noise level

In Fig. 5, agents' performance distribution for 50% noise level is represented:

- The multi-agent system maintains its advantage, achieving the lowest median evaluation value.
- The spread of individual agents' performance increases, showing that stability of some agents is lower.
- The multi-agent system remains more stable (less variance), indicating robustness to noise, while some individual agents exhibit high outliers, meaning their performance is getting worse under uncertainty.

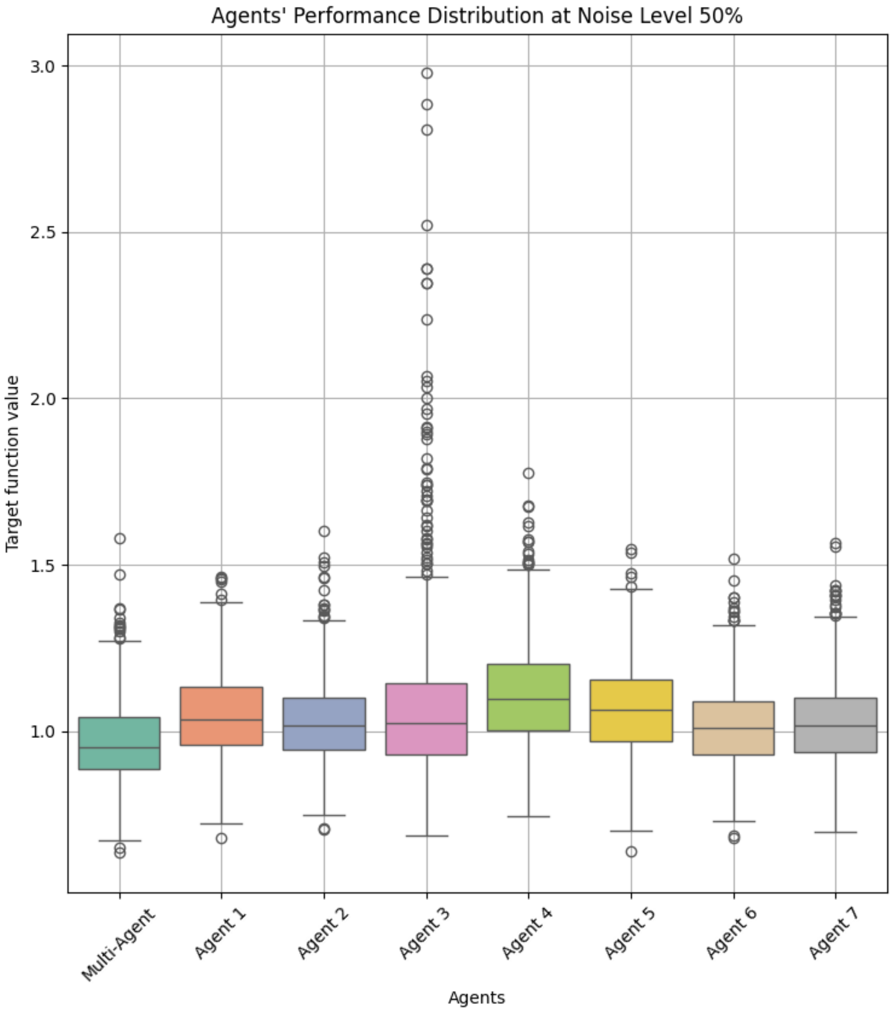

Figure 5: Agents' performance distribution for 50% noise level

8 SUMMARY

The results suggest that the multi-agent approach is both more robust and more effective at optimizing performance across different levels of noise. This study demonstrated that multi-agent reinforcement learning provides better structural control compared to a single-agent approach. We found that utilizing the combined knowledge of agents improves the control performance, and that the multi-agent system provides more robustness under noise conditions. Future work will explore cooperative learning strategies to further enhance system robustness and adaptability.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support of the National Science Centre, Poland, granted under the grant agreement 2020/39/B/ST8/02615. For the purpose of Open Access, the

authors have applied a CC-BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission.

REFERENCES

[1] D. Silver, A. Huang, C.J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, et al., "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, pp. 484-489, 2016.

[2] J. Kober, J.A. Bagnell, and J. Peters, "Reinforcement learning in robotics: A survey," *International Journal of Robotics Research*, vol. 32, no. 11, pp. 1238-1274, 2013.

[3] S. Shalev-Shwartz, S. Shammah, and A. Shashua, "Safe, multi-agent, reinforcement learning for autonomous driving," *arXiv preprint*, arXiv:1610.03295, 2016.

[4] A. Bernard and I.F. Smith, "Reinforcement learning for structural control," *Journal of Computing in Civil Engineering*, vol. 22, no. 2, pp. 133-139, 2008.

[5] B. Rolf, I. Jackson, M. Müller, S. Lang, T. Reggelin, and D. Ivanov, "A review on reinforcement learning algorithms and applications in supply chain management," *International Journal of Production Research*, vol. 61, no. 20, pp. 7151-7179, 2023.

[6] D. Pisarski and Ł. Jankowski, "Reinforcement learning-based control to suppress the transient vibration of semi-active structures subjected to unknown harmonic excitation," *Computer-Aided Civil and Infrastructure Engineering*, vol. 38, no. 12, pp. 1605-1621, 2023.

[7] F. da S. Brandão and L. F. F. Miguel, "Vibration control in buildings under seismic excitation using optimized tuned mass dampers," *Frattura ed Integrità Strutturale*, vol. 54, pp. 66-87, 2020.]

[8] S. Korkmaz, "A review of active structural control: Challenges for engineering informatics," *Computers and Structures*, vol. 89, no. 23-24, pp. 2113-2132, 2011.

[9] J. Ormondroyd and J.P. Den Hartog, "The Theory of the Dynamic Vibration Absorber," *Journal of Fluids Engineering, Transactions of the ASME*, vol. 49-50, no. 2, pp. 021007, 1928.

[10] S. Nagarajaiah and H.-J. Jung, "Smart tuned mass dampers: Recent developments," *Smart Structures and Systems*, vol. 13, no. 2, pp. 173-176, 2014.

[11] A. Jedlińska, D. Pisarski, G. Mikułowski, B. Błachowski, Ł. Jankowski, "Semi-active structural control using viscous dampers and reinforcement learning," In *Proc. of the 10th ECCOMAS Thematic Conference on Smart Structures and Materials (SMART)*, 2023.

[12] A. Jedlińska, D. Pisarski, G. Mikulowski, B. Blachowski, and Ł. Jankowski, "Semi-Active Control of a Shear Building Based on Reinforcement Learning: Robustness to Measurement Noise and Model Error," In *Proc. of the 18th Conference on Computer Science and Intelligence Systems (FedCSIS 2023)*, Warsaw, Poland, 2023, pp. 1007-1010.

[13] R.S. Sutton and A.G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed., MIT Press, 2018.

[14] H. Van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-Learning," In *Proc. of the 30th AAAI Conference on Artificial Intelligence (AAAI 2016)*,

Phoenix, AZ, USA, Feb. 12–17, 2016, pp. 2094-2100.
[15] B. Jang, M. Kim, G. Harerimana, and J.W. Kim, "Q-Learning Algorithms: A Comprehensive Classification and Applications," *IEEE Access*, vol. 7, pp. 133653-133667, 2019.
[16] G. Tesauro, "Extending Q-learning to general adaptive multi-agent systems," In *Proc. Advances in Neural Information Processing Systems*, vol. 16, pp. 871-878, 2003.
[17] R. Lowe, Y. Wu, A. Tamar, J. Harb, P. Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," In *Proc. of the 31st Conference on Neural Information Processing Systems (NeurIPS 2017)*, vol. 30, 2017
[18] K. Zhang, Z. Yang, and T. Başar, "Multi-agent reinforcement learning: A selective overview of theories and algorithms," in M.W. Mahoney and R.S. Sutton, Eds., *Handbook of Reinforcement Learning and Control*, vol. 284, pp. 321-384, Springer, 2021.
[19] T. Rashid, M. Samvelyan, C. Schroeder de Witt, G. Farquhar, J. Foerster, and S. Whiteson, "Monotonic value function factorisation for deep multi-agent reinforcement learning," *Journal of Machine Learning Research*, vol. 21, no. 178, pp. 1-51, 2020.
[20] H. Tang, H. Zhang, Z. Shi, X. Chen, W. Ding, and X.-P. Zhang, "Autonomous swarm robot coordination via mean-field control embedding multi-agent reinforcement learning," In *Proc. 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Detroit, MI, USA, pp. 8820-8826, 2023.
[21] T. Gavin, S. LacroiX, and M. Bronz, "Multi-agent reinforcement learning based drone guidance for N-view triangulation," In *Proc. 2024 International Conference on Unmanned Aircraft Systems (ICUAS)*, Chania, Crete, Greece, pp. 578–585, 2024.
[22] Z. Zhang and D. Wang, "Adaptive individual Q-learning—A multiagent reinforcement learning method for coordination optimization," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 4, pp. 7739-7750, 2025.
[23] J. Liu, F. Li, J. Wang, and H. Han, "Proximal policy optimization based decentralized networked multi-agent reinforcement learning," In *Proc. 2024 IEEE 18th International Conference on Control & Automation (ICCA)*, Reykjavík, Iceland, pp. 839-844, 2024.
[24] Y.Q. Ji, "Application of meta-learning in multi-agent reinforcement learning – A survey," *Trends in Social Sciences and Humanities Research*, 2024. https://doi.org/10.61784/tsshr3075
[25] M. Ibrahim and A. Fayad, "Hierarchical strategies for cooperative multi-agent reinforcement learning," *arXiv Preprint*, arXiv:2212.07397, 2022.
[26] S. Pourzeynali, H.H. Lavasani, and A.H. Modarayi, "Active control of high-rise building structures using fuzzy logic and genetic algorithms," *Engineering Structures*, vol. 29, no. 3, pp. 346-357, 2007.