Proceedings of the Design Society, Volume 6: DESIGN 2026
https://doi.org/10.1017/pds.2026.10445

Enhancing design adaptation through an information-enriched reinforcement learning state

Yannick Utz✉, Sandro J. Wartzack and Stefan Goetz

Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

✉ utz@mfk.fau.de

ABSTRACT: The applicability and scalability of design adaptations utilizing reinforcement learning can be broadened by using graph-based approaches instead of rigid vector- or grid-based ones. However, graph-based approaches often require a high number of simulations to converge. To reduce the simulation effort in the mechanical optimisation, the reinforcement learning setup is enriched with task-specific causal and physically based information. This work systematically investigates the influence of this additional information on the efficiency of design adaptations using a factorial test design.

KEYWORDS: design automation, artificial intelligence (AI), reinforcement learning, graph neural networks, information enrichment

1. Introduction

The deployment of artificial intelligence (AI) can significantly increase the productivity of industrial companies by automating complex tasks (Czarnitzki et al., 2023). Thus, in product development, tasks based on complex physical principles and whose solution is partly built on trial and error, such as design adaptations, can be automated using Reinforcement Learning (RL) (Ng et al., 2021; Panzer & Bender, 2022). However, these design adaptations can have various objectives, which are defined in the associated Design for X (DfX). Examples include Design for Manufacture and Design for Strength. So, the information relevant to decision-making is defined by the specific DfX goal. (Bauer & Meerkamm 2007; Holt & Barnes, 2010) Integrating this specific information into the RL state can improve the efficiency of design adaptations by reducing the number of simulations required while maintaining or improving design performance. To quantify the impact of structured adaptation-specific information in the RL algorithm on the efficiency of the design adaptation, ablation studies are conducted. In these ablation studies, additional information is systematically added to an initially sparse state (initial value) in order to analyse the influence of the individual state enrichments on the result separately. To this end, task-relevant additional information is first collected and classified, and then converted into states with varying degrees of enrichment. These states are thereafter tested for their associated sample efficiency and policy quality which characterise the efficiency of design adaptations. Conclusions about the benefit of the state enrichments are drawn from direct comparisons. By holding the RL algorithm, reward, and action space constant, this approach isolates the influence of the individual additional information on the learning dynamics.

2. State of the art

RL is a subfield of machine learning (ML). Its functionality can be seen in Figure 1. An agent maps each state of an environment to an action and receives a reward as feedback, which assesses the achievement of specified goals in the new state. A state is the agent's observation of the environment. Following the prior action, a new action is selected and executed for the new state. The aim is to determine a policy that maximises the long-term reward. (Sutton & Barto, 2014) This can be achieved using deep Q-learning, a technique which optimises sequential decision-making processes, such as design adaptations. This is

https://doi.org/10.1017/pds.2026.10445 Published online by Cambridge University Press

accomplished by approximating the expected discounted cumulative future reward (Q-values) using Deep Q-Networks (DQNs). (Mnih et al., 2015; Watkins & Dayan, 1992)
In deep Q-learning, the Q-values are calculated based on the state (Mnih et al., 2015). However, setting up the state requires considering the Markov property. This means that calculating the next state and expected reward depends only on the current state and action. (Sutton & Barto, 2014)

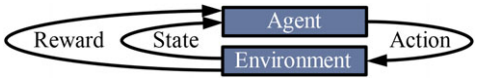


Figure 1. Schematic functionality of RL in accordance to Sutton and Barto (2014)

In product development, RL states are modelled differently depending on the application. In Dworschak et al. (2022), a state vector is used to optimise bicycle components. This vector represents parameters that can be directly adapted by the RL, such as sheet thicknesses or hole diameters. A similar approach is used in Brown et al. (2022), where the elements of a 2D discretised topology to be adapted are represented directly within a state grid. As the locations of the supports and forces are marked on the grid alongside the topology, this state representation provides a more precise description of the physical mechanisms. However, the explicit encoding of causal and physics-based mechanisms in the state and the analysis of their influence on the result are rarely applied in the field of RL for structural design adaptations. Nevertheless, the learning time of RL can be improved by using causal and physically meaningful state representations, for example by integrating Lagrange multipliers into the state in order to solve RL problems with multiple conditions (Banerjee et al., 2025; Calvo-Fullana et al., 2023). However, the state modelling requires a trade-off to be found, because states with insufficient information lead to suboptimal policies, while states with redundant information result in increased complexity and a greater number of samples being required for the training of the network (B. Huang et al., 2022).
Furthermore, these current approaches for structural design adaptations operate on vector- or grid-based states that have a fixed size. In contrast, Hayashi & Ohsaki (2020) and Hayashi & Ohsaki (2021) optimise trusses by representing the states as graphs. Here, the joints are represented as nodes and the beams as edges, whereby adjacent joints are directly connected. Graph neural networks (GNNs) are used to process this state. Compared to grid- and vector-based states and their processing with multi-layer perceptrons (MLPs), this setup offers advantages in terms of the size and structure of the inputs that can be processed, as these are not limited to fixed grid sizes. In addition, causal and physical relations can be explicitly encoded in graphs.
In general, GNNs have different characteristics than MLPs, which have to be considered when adapting designs using GNN-based RL. GNNs directly utilise the graph structure and learn from the interactions between graph neighbours through message passing between neighbouring nodes, whereas classical MLPs treat all inputs equally and cannot explicitly model relationships (Gilmer et al., 2017; Wu et al., 2022). Working on the basis of graph structure leads to permutation invariance because graphs have no explicit node order. This also enables GNNs to be used for graphs of various sizes originating from the same domain. (Zhou et al., 2020) By contrast, MLPs cannot be used directly for different state sizes. For new state sizes, MLPs would have to be adapted to the new input size and retrained from scratch. Furthermore, with sufficient connectivity, the generalisation capability to new use cases is higher for GNNs than for MLPs due to its signal learning (W. Huang et al., 2025). However, it should be noted that excessive edge density can lead to oversmoothing during the processing of the graph in the GNN, which impairs the network's learning ability. In graphs with too many edges, the nodes lose their distinctiveness during message passing and become increasingly homogeneous. (Wu et al., 2022) Furthermore, low sample efficiency is a potential shortcoming of graph-based approaches (Awasthi et al., 2022).

3. Research gap

Although some research has already used RL to automate design adaptations, whose states are partly enriched with additional information, the influence of various types of additional information on the efficiency of design adaptations has not yet been systematically investigated. This means, it is unclear what additional benefit each type of information provides in terms of the efficiency of RL-based design adaptations.
Furthermore, the states of the structural mechanical optimisation are usually represented in a vector- or grid-based manner. So, in parametric design adaptation, each element that can be adapted is represented by an input neuron (Dworschak et al., 2022). However, these approaches can only be applied to

https://doi.org/10.1017/pds.2026.10445 Published online by Cambridge University Press

components with the same structural design and the same number of parameters to be adapted. For example, when optimising the ply thicknesses of fibre-reinforced composites with eight plies, the network can only be used for composites with eight plies without changing the network architecture and retraining, which severely limits its practical usability in product development. The utilization of GNN-based RL helps to overcome these limitations, since they can process different numbers of components represented by nodes. Another advantage of these GNN-based approaches is that additional information can be represented not only in the graph's features, but also via edges indicating structural relationships between components (e.g., to indicate symmetry or load paths). This can enhance the causal and physical description of the state.

However, it remains to be investigated how enriching the GNN-based state with structural information improves the efficiency of RL-automated design adaptations without facing issues such as oversmoothing. This raises the following research question: **How do different levels of additional task-relevant structured information embedded in the GNN-based RL-state influence the efficiency of automated design adaptations?**

To answer the research question and assess the efficiency of automated design adaptations, ablation studies are conducted. Herein, the effect of additional information on the sample efficiency and policy quality, which characterise the efficiency of design adaptations, is analysed. The chosen domain example is the optimisation of the thicknesses and fibre orientations of the plies of a fibre-reinforced composite component, which has strong mutual relationships between the individual plies.

4. Methodical approach

The methodical approach of analysing the states enriched with different levels of task-relevant information in terms of efficiency of automated design adaptations is shown in Figure 2. First, it is necessary to determine which engineering information can be used in the state. This information is identified and classified into categories in Section 4.1. In Section 4.2, these categories are converted into graphs with varying degrees of information to form different state variants. Subsequently, in Section 4.3, GNN-based RL algorithms are trained with these various variants as respective states. During training, ablation studies are carried out at regular intervals to evaluate the effect of individual state changes on the sample efficiency and policy quality.

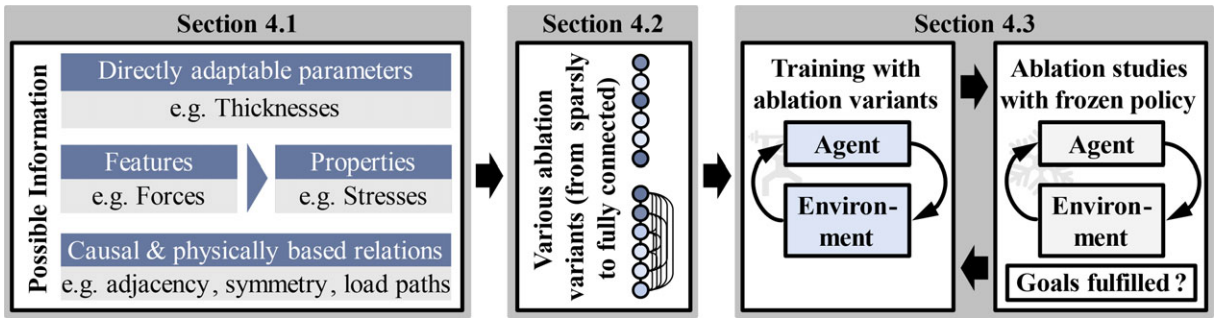


Figure 2. Overview of the methodical approach

4.1. Definition of possible information enrichment within the state

In addition to the *parameters that can be adapted directly* by the agent, the state may contain other influencing parameters for a design adaptation. To determine these influencing parameters, the goal of design adaptation according to DfX must first be clarified. From this goal, the influencing parameters can be derived. These parameters are physically based and used by engineers in design adaptation tasks. In order to test the amount of additional information required to increase the efficiency of design adaptations, these influencing parameters must first be classified into categories. For this purpose, a distinction is made between *task-relevant features* (such as the location of forces) and the resulting *properties derived from them* (such as stress values).

However, additional information can be expressed not only by parameters such as features and properties, but also by *causal, physically based relations* between design elements, for example, regarding adjacencies, force paths and symmetries. Therefore, various types of relationship between the design elements are distinguished in order to assess their impact on sample efficiency and policy quality.

https://doi.org/10.1017/pds.2026.10445 Published online by Cambridge University Press

4.2. Transfer to various information-enriched graphs

In physical systems such as mechanically stressed designs, the components and parameters are interrelated. To model this in the proposed graph-based approach, components are represented by nodes and their relations by edges. This allows information such as *directly adaptable parameters*, *task-relevant features* and *derived properties* to be described in the node features of the individual components, while *causal, physically based relations* can be represented by edges. To avoid unnecessary effort and oversmoothing, the effects of different amounts of information described in the respective states are tested within ablation studies. In order to isolate the effects of various changes in the state representations, two factors are defined:

Factor A — Node-feature content:

- A1: node features contain only the directly adaptable parameters
- A2: A1 plus additional task-relevant features (e.g., location/magnitude of forces)
- A3: A2 plus additional task-relevant properties derived from respective features (e.g., stresses)

Factor B — Edge set/edge density:

- B1: adjacency edges only (i.e., edges between neighbouring components; minimum number of edges in a graph)
- B2: B1 plus a set of semantically meaningful, task-relevant edges (e.g., symmetry edges)
- B3-Bx: various edge-sets which progressively densify with further task-relevant edges beyond B2 up to a fully connected graph (fully connected = maximal number of edges in a graph)

This results in the following sets of states, in which only one information class changes between variants, enabling the direct identification of the effects of the changes:

- A1 x B1: sparse state with the directly adaptable parameters and basic adjacent edges to represent the baseline
- A1 x B2: previous combination plus task-relevant edges to allow individual assessment of the influence of extra task-relevant edges on sample efficiency and policy quality
- A2 x B2: extension of the graph features from the previous combination with additional task-relevant features to assess their influence on the result
- A2 x B3-Bx: previous combination plus additional semantically relevant edges to find the trade-off between information density and edge density
- A3 x B2 and A3 x B3-Bx: extension of the node features from the two previous combinations with additional properties derived from the task-relevant features in order to assess the effect of these properties

Across all conditions, the other constraints such as the RL algorithm, the reward function, and the action space are constant to enable causal attribution to the state design.

4.3. Assessing sample efficiency and policy quality through ablation studies

A GNN is used to process the different state sets and calculate action-specific Q-values. Herein, node embeddings that contain both local and global information are built through edge-conditioned convolutions and message passing. These node embeddings are then processed by an MLP to calculate the Q-values. Only one MLP is required for all node embeddings, as one of three possible actions can be performed for each component: increase, decrease or keep the same. To reduce the action space, however, only one component can be adapted per step.

To validate the efficiency of the automated design adaptation process, the associated RL is evaluated every 100 steps during training. This is achieved by using known and randomly generated component configurations with different seeds, in order to robustly assess sample efficiency and policy quality. This frequency strikes a balance between monitoring the quality of the agent's policy and avoiding prolonging training unnecessarily. This is because, during these checks, the network processes new test cases whose stress properties must be simulated. During these checks, the policy is frozen, meaning it cannot learn from the test cases through backpropagation. This makes the results of the different variants comparable. Additionally, to prevent random actions, the agent acts continuously without exploring. This also promotes comparability, as no good random actions can be taken by chance during the checks.

During these evaluation checks, the percentage of runs that meet the adaptation goals of the respective test cases is logged (success rate). Sample efficiency is measured as the number of unique simulation

https://doi.org/10.1017/pds.2026.10445 Published online by Cambridge University Press

calls (unseen samples) required to first reach three consecutive episodes in which all adaptation goals are achieved. The criterion of three consecutive episodes with a success rate of 100% prevents a random success from distorting the result. So, this criterion can be used to assess when the RL understands how to solve the task. Consequently, it could be used to determine the end of the training. To assess the policy quality, the focus is on the period after this sample efficiency criterion has been reached, as few conclusions about the long-term policy can be drawn from the previous period, when correlations are being learned. So, to analyse the policy quality, the area under the success rate (Area under Curve; AUC) is calculated after reaching the sample efficiency criterion and related to the change in the step number. This calculates the average success rate of the policy at the evaluation checks. Figure 3 shows the schematic determination of the sample efficiency and the policy quality.

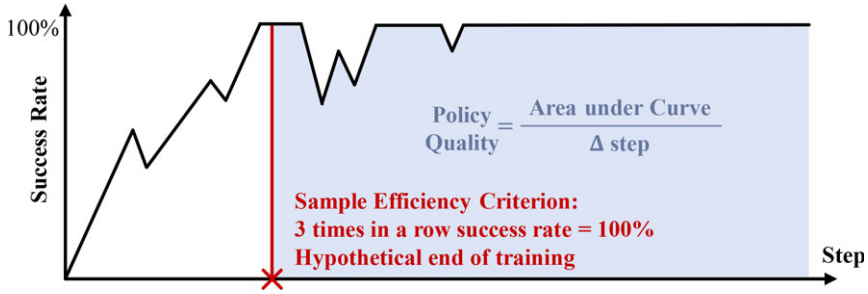


Figure 3. Schematic representation of the measurands to assess the sample efficiency and policy quality

5. Study - simultaneous adaptation of the ply thickness and fibre orientation in a fibre-reinforced composite design

This methodical approach is applied to the stress-based design adaptation of fibre-reinforced composites. The composite structure is therefore represented by a graph, with the plies represented by nodes. To avoid distortion, the ply stack-up is made symmetric, meaning that the fibre orientation and thickness of opposite plies are equal. The fibre orientation and ply thickness influence the stresses in the design. (Cardarelli, 2008) Thus, these parameters are adapted within the RL-based design adaptation to ensure compliance with the permitted stress.

In the case under consideration, represented in Figure 4, the agent can simultaneously adapt the thicknesses and fibre orientations of a composite built of eight plies. The ply thicknesses can be adapted in 0,1 mm increments within the range [0,2 mm, 0,4 mm], and the fibre orientations in 45° increments within the range [0°, 135°]. This results in 20.736 possible combinations.

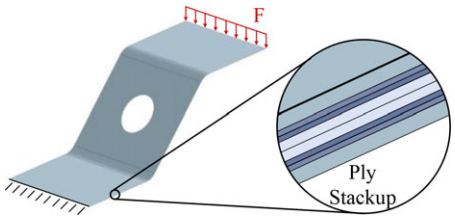


Figure 4. Representation of the load case in the application example

The strength of anisotropic materials such as the considered fibre-reinforced composites can be evaluated using the Tsai-Wu failure criteria (Tsai & Wu, 1972). To this end, the design adaptation is assessed using the Tsai-Wu value per ply. The adapted components are simulated using the Python interface of Ansys 2024 R2 to determine the Tsai-Wu distribution for each ply. To avoid reducing the Tsai-Wu value simply by increasing ply thickness, an additional weight-related condition is introduced.

5.1. Definition of the possible information contained within the different states

Firstly, the information that could be contained within the state are analysed. According to Section 4.1, the *directly adaptable parameters*, *additional task-relevant features* and *properties derived from them*, as well as *causal, physically based relations* between the plies, are investigated.

https://doi.org/10.1017/pds.2026.10445 Published online by Cambridge University Press

The *directly adaptable parameters* are represented by the fibre orientation and thickness of each ply of the laminate. Regarding *additional task-relevant features*, the distance of a ply to the middle of the laminate can be used to calculate the bending stiffness (Jones, 1999). This distance is thus task-relevant and can therefore be included in the state, making this information available to the agent. In addition, *properties derived from the features* can also be used to enrich the state. These are directly related to the adaptation goal. To provide the agent with this kind of information, the maximal Tsai-Wu value per ply can be included in the respective nodes. This enables the agent to observe the intermediate effects of its actions and learn which node to adapt to achieve the adaptation goal.
In terms of *causal, physically based relations*, three relation types can be encoded: adjacency between neighbouring plies (bonded interfaces), symmetry between opposite plies in symmetric laminates, and orientation-dependent coupling between ply pairs, which represents the mechanical correlations through fibre orientation difference $\Delta\theta$ (Memarzadeh et al., 2024). To define discrete information levels, $\Delta\theta$ is discretized into $\{0°, 45°, 90°\}$ with $\Delta\theta = 90°$ as the maximum fibre orientation difference. According to Section 4.2, Table 1 shows the possible information, which can be represented within the state.

Table 1. Contribution of the possible information in the state

Possible information (Factor A)			**Possible links (Factor B)**				
Directly adaptable parameters	***Task-relevant Features***	***Derived Properties***	***Causal and physically based relations***				
A1: Orientation, thickness	A2: A1 + distance ply - centre	A3: A2+ Tsai-Wu per ply	B1: Adjacent plies	B2: B1 + opponent plies	B3: B2 + $\Delta\theta = 0°$	B4: B3 + $\Delta\theta = 45°$	B5: B4 + $\Delta\theta = 90°$

5.2. Building graphs with several levels of information

Ten state variants are tested (see Table 2). This set is derived from the test plan outlined in Section 4.2 and the definition of possible information within the state, as detailed in Section 5.1.

Table 2. Set of various state variants

	Node-Features			**Edges**				
	Directly adaptable parameters	***Task-relevant Features***	***Derived Properties***	***Causal and physically based relations***				
	Orientation, thickness	**Distanceply - centre**	**Tsai-Wu per ply**	**Adjacency**	**Symmetry**	$\Delta\theta = 0°$	$\Delta\theta = 45°$	$\Delta\theta = 90°$
V1 (A1 x B1)	x			x				
V2 (A1 x B2)	x			x	x			
V3 (A2 x B2)	x	x		x	x			
V4 (A2 x B3)	x	x		x	x	x		
V5 (A2 x B4)	x	x		x	x	x	x	
V6 (A2 x B5)	x	x		x	x	x	x	x
V7 (A3 x B2)	x	x	x	x	x			
V8 (A3 x B3)	x	x	x	x	x	x		
V9 (A3 x B4)	x	x	x	x	x	x	x	
V10 (A3 x B5)	x	x	x	x	x	x	x	x

5.3. Investigating sample efficiency and policy quality through evaluation checks

Evaluation checks are conducted every 100 steps. These evaluation checks consist of test cases with the following initial thicknesses and fibre orientations, which the RL must adapt in order to achieve the strength-based goal:

https://doi.org/10.1017/pds.2026.10445 Published online by Cambridge University Press

1. Thickness of 0,2 mm and fibre orientation of 0° for all plies, respectively: tested three times with three different seeds
2. Thickness of 0,4 mm and fibre orientation of 90° for all plies, respectively: tested three times with three different seeds
3. Random thicknesses and fibre orientations within the borders for all plies, respectively: three different variants tested with different seeds

The first two variants, with initial ply thicknesses of 0,2 mm or 0,4 mm and initial fibre orientations of 0° or 90°, are chosen to cover the border area. Testing these variants with different seeds, as well as variants of random thicknesses and fibre orientations with different seeds, ensures robust results. To analyse the efficiency of the design adaptation, the step is investigated at which all checks for the nine combinations have been passed three times in a row with a 100% success rate. Then, the agent could solve a total of 27 different combinations (9 combinations x 3 checks) in a row, 9 of which were chosen completely at random. Afterwards, the relative success rate is calculated to assess the quality of the respective policy.

5.4. Results of the evaluation checks

The results of the evaluation checks within the training sessions of V1 - V10, each consisting of 500 episodes with 20 steps (10.000 steps in total), are depicted in Figure 5.

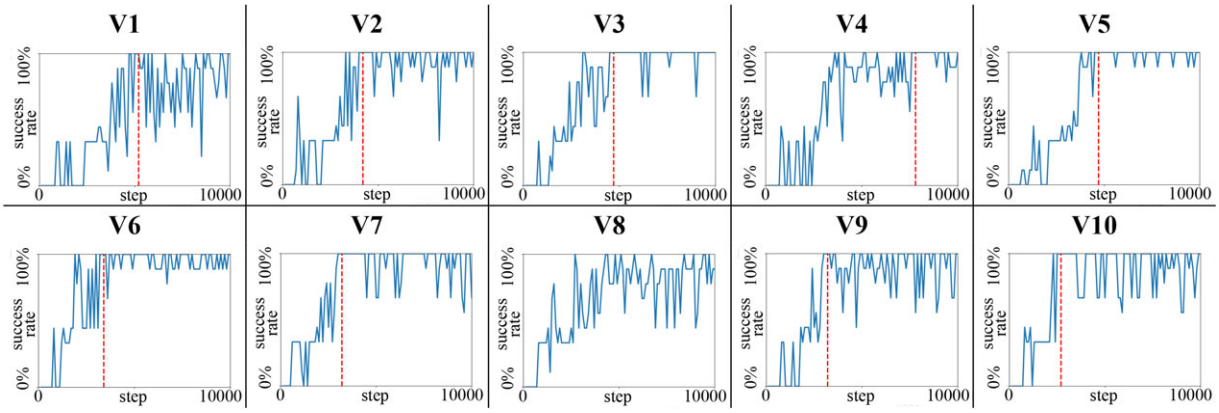


Figure 5. Resulting curves (blue) of the evaluation checks with indication of the sample efficiency criterion (red dotted)

In addition, Table 2 lists the step and the corresponding sample size when all evaluation checks were passed 100% three times in a row (sample efficiency criterion), as well as the average relative success rate after that point to indicate the long-time quality of the policy.

Table 3. Representation of the step number until reaching the sample efficiency criterion, the corresponding sample size and the average success rate

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
Steps until sample efficiency criterion	5200	4200	4700	7800	4700	3400	3200	–	3200	2700
Samples until sample efficiency criterion	2369	2073	2053	2250	2063	1872	1824	–	1820	1703
Final average success rate [%]	78,5	93,7	97,9	96,5	98,5	95,5	94,0	–	89,4	89,8

5.5. Analysis of the results of the ablation studies

The information classes of the graphs are analysed. This involves comparing the results of the sparse graph V1, the graph with additional causal and physically based edges V2, the graphs with additional task-relevant features (V3–V6) and the graphs with extra properties derived from them (V7–V10) to investigate the type of information that helps the agent choose the right actions. Furthermore, the influence of edge density on sample efficiency and policy quality is analysed.

Adding causal edges to the baseline graph: V1, the sparse baseline graph, requires 5.200 steps and 2.369 samples to achieve the adaptation goal of all test cases within the evaluation checks three times in a row.

https://doi.org/10.1017/pds.2026.10445 Published online by Cambridge University Press

The average success rate after this point is 78,5%. This means that after step 5.200, in 78,5% of test cases during the evaluation checks, the adaptation goal is achieved. Adding task-relevant edges (V2) improves the number of steps (4.200; -19,2%) and the sample size (2.073; -12,5%) until the sample efficiency criterion is reached, as well as the average success rate afterwards (and thus the quality of the policy). V2 has an average success rate of 93.7%, which is 19,4% higher than V1's. This can be attributed to the enhanced flow of information through the additional edges.

Adding task-relevant features: The final average success rate and thus the quality of the policy can be increased by providing extra task-relevant features. This can be seen from the final success rate of V3 (97,9%; +4,5%). This is because the graph of V3 more accurately describe the physical system by adding physically based decision-relevant parameters.

Adding causal and physically based edges (until the graph is fully connected): The edge density of the combinations V3 - V6 is now being analysed to investigate the impact of edge number on the results. The highest success rates are shown in graphs V3 and V5. V3's high success rate is due to the fact that its edges are constant and do not depend on the state. However, the edges of V4 and V5 depend on differences in fibre orientation and are thus dynamic. Graphs with dynamic edges can increase the complexity and decrease the efficiency of GNNs (Zheng et al., 2025). This increase in complexity stems from an increase in the size of the state space. While V5 can overcome this difficulty thanks to its greater amount of information and more precise description of the physical system, the efficiency of V4 is comparatively low due to its limited information input and dynamic edges. Like V3, V6 has constant edges, but in contrast to V3, V6 is a fully connected graph. However, due to the high edge density of V6 and the subsequent oversmoothing (Wu et al., 2022), its quality of the policy is lower.

The main observations also hold for the Tsai–Wu–enriched variants (V7–V10). In graphs with constant edges (V7 and V10), the quality of the policy is lower for fully connected graphs. In contrast, in graphs with dynamic edges (V8 and V9), adding edges and hence more information improves the efficiency of design adaptations.

Adding properties derived from the task-relevant features: Enriching the state with properties such as Tsai-Wu values decreases the final average success rate of the policies. This is made evident by comparing the final average success rate of the respective graphs, once with and once without additional properties derived from task-relevant features (V7: 94,0%, −4,0% compared to V3; V8: −; V9: 89,4%, −9,2% compared to V5; V10: 89,8%, -6,0% compared to V6). This is due to the fact that adaptations regarding ply thickness and fibre orientation can cause significant changes in the Tsai-Wu values of the plies. This is exacerbated by the fact that the state does not always contain the Tsai-Wu value of the same FE-element, but rather the maximum Tsai-Wu value across all FE-elements of a ply. Consequently, the FE-elements with the highest Tsai-Wu values change frequently. This amplifies the discontinuities, which are difficult for the agent to learn. On the other hand, the decrease in policy quality is due to the high correlation between the Tsai-Wu values in the state and the reward. Consequently, the agent quickly learns rigid, fragile approximations, resulting in steeper learning curves at the beginning of training. However, this can also cause issues when patterns change slightly or inaccuracies occur, resulting in a loss of policy quality. This faster learning process at the beginning of training is evident in the present results because if all adaptation goals are achieved three times in a row, both the associated step and sample size are generally lower than in the respective graphs V3-V6. However, this less robust policy, which includes a number of outliers, results in V8 not even reaching the sample efficiency criterion in the 10.000 training steps. This suggests that these graphs, which have additional properties derived from task-relevant features (regarding stress/RL-reward), reach the point faster at which they can automatically adapt these designs. However, the associated policy is less robust afterwards than for graphs with additional task-relevant features but without these extra properties.

6. Discussion

Ablation studies were conducted to investigate the influence of state enrichment with additional information on the efficiency of design adaptations, with direct comparisons of various graph variants. State variants that differ in only one aspect, such as the addition of extra stress properties, were investigated to identify the cause of different results. The factorial ablation design makes it possible to evaluate the contribution of node features (factor A) and edge sets (factor B) to the result separately.

In order to control random influences and thus assign the changes directly to the different state representations, different seeds are used. However, to ensure comparability between the individual state variants, identical seed tuples are used for all state variants (seed-matched comparison).

https://doi.org/10.1017/pds.2026.10445 Published online by Cambridge University Press

The comparability between different state representations is also ensured by applying identical framework conditions for all state variants (same network, reward design, etc.). To obtain robust, reliable results, nine evaluation checks were performed every 100 steps, which is a good balance between coverage and effort, as these tests covered both edge cases and random combinations across different seeds.

The results suggest that enhancing sample efficiency while maintaining a high policy quality by enriching the state with additional information is advantageous when providing task-relevant features alongside the directly adaptable parameters. However, resulting properties such as stress-based Tsai-Wu values with a direct reference to the reward function can be included in the state if rapid automation results are required and lower policy quality is sufficient. Nevertheless, this promotes reward shortcuts and hampers the learning of real structural relations. Regarding edge density, fully connected graphs and graphs with dynamic edges and low information impact can decrease the policy's quality.

With regard to the measurands and the results of the evaluation checks, the sample efficiency criterion is particularly vulnerable to interference as it provides a snapshot of a stochastic application. To mitigate this problem and reduce the influence of randomness, this measurand could have been selected with a certain degree of tolerance, for example by setting the average success rate in three consecutive evaluation checks to at least 95%. Nevertheless, the same general key findings would have emerged even with such definitions of the sample efficiency criterion.

In addition to the ability to directly assign edges and thus leverage relations to learn from structural patterns, graph-based machine learning applications offer advantages compared to vector- and grid-based networks in terms of transferability to other test cases with different sizes of the mechanical design (Ruiz et al., 2023). To avoid exceeding the scope of the work, only one fibre-reinforced composite with eight plies was tested. However, it is recommended to investigate the transferability of this approach to test cases with other ply sizes in future work.

7. Conclusion and outlook

This paper presents an analysis of how enriching the state of RL with additional task-relevant information affects the efficiency of automated design adaptations. The approach used graphs and GNNs to learn from additional information and structural patterns. The amount of information and its integration within the graph were tested. Therefore, various graph variants with different amounts of node features and edges were analysed in ablation studies. This approach was used in the context of strength-based design adaptations of a fibre-reinforced composite by simultaneously adapting thicknesses and orientations of the plies. However, the approach can also be applied to further design adaptations, which take place in the embodiment and final design stages, e.g., to ensure manufacturability or assemblability. To this end, additional parameters in the state, such as task-relevant features, and causal relations must be selected depending on the goals of the design adaptation.

The key finding is that enriching the state with additional information can make automated design adaptations more efficient. Even adding a few task-relevant relations between components in the state enhanced the efficiency in the tests. Further gains in efficiency were achieved by adding additional task-relevant features that are also considered by engineers when adapting designs. Adding properties derived from these features reduced the number of simulations until the RL learned the relevant relationships to adapt the design. However, the quality of the design adaptations decreased after this point compared to the variants with extra task-relevant features, but without additional properties derived from them.

Building on this work, the efficiency of automated design adaptations will be investigated by integrating not only the information used by product developers, but also their procedures into RL. This can be achieved by integrating design rules within the reward function of the RL.

Acknowledgement

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 522180880.

References

Awasthi, P., Das, A., & Gollapudi, S. (2022). Beyond GNNs: An Efficient Architecture for Graph Problems. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(6), 6019–6027. https://doi.org/10.1609/aaai.v36i6.20548

Banerjee, C., Nguyen, K., Fookes, C., & Raissi, M. (2025). A Survey on Physics Informed Reinforcement Learning: Review and Open Problems. https://doi.org/10.2139/ssrn.4597487

https://doi.org/10.1017/pds.2026.10445 Published online by Cambridge University Press

Bauer, S., & Meerkamm, H. (2007). Decision making with interdependent objectives in Design for X. In DS 42: Proceedings of ICED 2007. The Design Society.
Brown, N. K., Garland, A. P., Fadel, G. M., & Li, G. (2022). Deep reinforcement learning for engineering design through topology optimization of elementally discretized design domains. *Materials & Design*, 218, 110672. https://doi.org/10.1016/j.matdes.2022.110672
Calvo-Fullana, M., Paternain, S., Chamon, L. F., & Ribeiro, A. (2023). State augmented constrained reinforcement learning: Overcoming the limitations of learning with rewards. *IEEE Transactions on Automatic Control*, 69(7), 4275–4290. https://doi.org/10.1109/tac.2023.3319070
Cardarelli, F. (2008). *Materials Handbook*. Springer London. https://doi.org/10.1007/978-1-84628-669-8
Czarnitzki, D., Fernández, G. P., & Rammer, C. (2023). Artificial intelligence and firm-level productivity. *Journal of Economic Behavior & Organization*, 211, 188–205. https://doi.org/10.1016/j.jebo.2023.05.008
Dworschak, F., Dietze, S., Wittmann, M., Schleich, B., & Wartzack, S. (2022). Reinforcement Learning for Engineering Design Automation. *Advanced Engineering Informatics*, 52, 1–13. https://doi.org/10.1016/j.aei.2022.101612
Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., & Dahl, G. E. (2017). Neural message passing for quantum chemistry. International conference on machine learning (1263-1272). PMLR.
Hayashi, K., & Ohsaki, M. (2020). Reinforcement Learning and Graph Embedding for Binary Truss Topology Optimization Under Stress and Displacement Constraints. *Frontiers in Built Environment*, 6, 1–15. https://doi.org/10.3389/fbuil.2020.00059
Hayashi, K., & Ohsaki, M. (2021). Reinforcement learning for optimum design of a plane frame under static loads. *Engineering with Computers*, 37(3), 1999–2011. https://doi.org/10.1007/s00366-019-00926-7
Holt, R., & Barnes, C. (2010). Towards an integrated approach to "Design for X": an agenda for decision-based DFX research. *Research in engineering design*, 21(2), 123–136. https://doi.org/10.1007/s00163-009-0081-6
Huang, B., Lu, C., Leqi, L., Hernández-Lobato, J. M., Glymour, C., Schölkopf, B., & Zhang, K. (2022). Action-sufficient state representation learning for control with structural constraints. International Conference on Machine Learning (9260-9279). PMLR.
Huang, W., Cao, Y., Wang, H., Cao, X., & Suzuki, T. (2025). Quantifying the Optimization and Generalization Advantages of Graph Neural Networks Over Multilayer Perceptrons. The 28th International Conference on Artificial Intelligence and Statistics, 28.
Jones, R. M. (1999). *Mechanics of composite materials (Second edition)*. Taylor & Francis Group.
Memarzadeh, A., Onyibo, E. C., Asmael, M., Safaei, B. (2024). Dynamic Effect of Ply Angle and Fiber Orientation on Composite Plates. *Spectrum of Mechanical Engineering and Operational Research*, 1(1), 90–110.
Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529–533. https://doi.org/10.1038/nature14236
Ng, K. K., Chen, C. H., Lee, C. K., Jiao, J. R., & Yang, Z. X. (2021). A systematic literature review on intelligent automation: Aligning concepts from theory, practice, and future perspectives. *Advanced Engineering Informatics*, 47. https://doi.org/10.1016/j.aei.2021.101246
Panzer, M., & Bender, B. (2022). Deep reinforcement learning in production systems: A systematic literature review. *International Journal of Production Research*, 60(13), 4316–4341. https://doi.org/10.1080/00207543.2021.1973138
Ruiz, L., Chamon, L. F., & Ribeiro, A. (2023). Transferability properties of graph neural networks. *IEEE Transactions on Signal Processing*, 71, 3474–3489. https://doi.org/10.1109/TSP.2023.3297848
Sutton, R. S., & Barto, A. G. (2014). *Reinforcement Learning: An Introduction (Second Edition)*. MIT Press.
Tsai, S. W., & Wu, E. M. (1972). *A General Theory of Strength for Anisotropic Materials*. Air Force Materials Laboratory.
Watkins, C., & Dayan, P. (1992). *Q-Learning*. Kluwer Academic Publisher.
Wu, L., Cui, P., Pei, J., & Zhao, L. (2022). *Graph Neural Networks: Foundations, Frontiers, and Applications*. Springer Nature Singapore. https://doi.org/10.1007/978-981-16-6054-2
Zheng, Y., Yi, L., & Wei, Z. (2025). A survey of dynamic graph neural networks. *Frontiers of Computer Science*, 19(6), 196323. https://doi.org/10.1007/s11704-024-3853-2
Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2020). Graph neural networks: A review of methods and applications. *AI Open*, 1, 57–81. https://doi.org/10.1016/j.aiopen.2021.01.001

https://doi.org/10.1017/pds.2026.10445 Published online by Cambridge University Press