REINFORCEMENT LEARNING: A CONTROL APPROACH FOR REDUCING COMPONENT DAMAGE IN MOBILE MACHINES

Lars Brinkschulte*, Marina Graf, Marcus Geimer

Institute Mobile Machines, Karlsruhe Institute of Technology, Rintheimer Querallee, 76131 Karlsruhe

*Corresponding author: Tel.: +49 721 608 45382; E-mail address: lars.brinkschulte@kit.edu

ABSTRACT

This paper presents an active component damage reducing control approach for driving manoeuvres of a wheel loader. For this purpose, the front and rear axle loads will be manipulated by force pulses induced into the machine chassis via the lifting cylinders of the function drive. The associated control approach is based on the principles of Reinforcement Learning. The essential advantage of such methods against linear control approaches is that no descriptive system properties are required, but the algorithm automatically determines the system behaviour. Due to the high number of necessary training runs, the algorithm is designed and taught using a validated wheel loader simulation model. After over 850 training runs, an optimal strategy for damping the axle loads could not yet be determined. In spite of the unprecedented convergence, initial improvements of the damage values have already been achieved on tracks that deviate from the training track. Some of these results show a 4.9 % lower component damage compared to a machine setting with no damping system. The results and limits of this strategy are discussed due to a comparison with other scientific active vibration damping approaches. Currently, a linear control method (P-PI-controller) has a higher damage reduction potential, but it is expected that further training runs and another learning algorithm could make the reinforcement learning approach even more effective. Coupling the linear control method with the self-learning approach shows the highest potential for the axle damage reduction.

Keywords: Reinforcement Learning, Active Vibration Damping, Damage Reduction, Wheel Loader, Holistic Wheel Loader Simulation

1. INTRODUCTION

Wheel loaders are subjected to constantly changing motion sequences and load situations as the machine operators execute various tasks in the working process such as digging, loading and transporting a wide variety of bulk material. In addition to the working task and the operator, the operating environment of a machine significantly influences the loads acting on the machine. Uneven road conditions, obstacles, such as stones, and the ground surface on construction sites are the determining factors during a driving manoeuvre.

The wheel loader system is thereby an oscillatory system. Vibrations in the working kinematics and the vehicle chassis are therefore caused by any movement of the machine. The consequences are reduced productivity and driving comfort as well as a reduction in the lifetime of structural machine components, such as the machines axles and parts of the working kinematics. Today most of the wheel loaders are not equipped with suspensions at the wheel axles. The pneumatic tyres act as vibration damping elements and are often combined with passive vibration damping (PVD) systems. A standard PVD consists of a hydraulic accumulator and valves, which are connected to the cylinders of the working kinematics. The vibrations are reduced by dissipating kinetic energy, but are optimized for a specific frequency range and can therefore only dampen axle load vibrations to a certain degree [1].

In addition to the PVD methods, there are systems that actively counterbalance the vibrations. In the case of a wheel loader, this is achieved by controlling the function drive (FD). The hydraulic cylinder forces resulting from the working kinematic movements are induced into the front end of the machine, which in turn counteracts the vibrations themselves. If the response and actuation time of the control system exceeds the frequency of the vibrations, active vibration damping (AVD) can be achieved for a broad speed and load range.

The damping can be achieved by valve-controlled [2–4] or displacement-controlled [5, 6] hydraulic systems. Depending on the driving manoeuvre, Madau shows in [4] a 45 % reduction in cabin acceleration vibration with a valve-controlled approach. The reduction is determined as the integral of the absolute cabin acceleration over the time. Williamson reduces the cabin vibration in [6] by up to 34 % using a displacement-controlled approach.

All control approaches, though, have in common that the control parameters must be determined and defined by technical expertise, often by carrying out test runs in simulation and reality. The analysis and interpretation of the results require a deep understanding of the interrelation-ships in the system. In contrast, artificial intelligence control systems can learn from their own experiences just as living creatures. These systems discover the optimum damping strategy by using for example the basic principles of reinforcement learning (RL).

The publication presents such a RL based AVD approach for driving processes of a wheel loader. The primary objective is to reduce the loads and therefore the damage of the machine axles. The application machine and the driving scenarios examined in the publication are presented in Section 2 of this paper. Section 3 deals with the description of RL approach. The training and testing of the control architecture takes place in a validated machine simulation, which is partly introduced in section 4. The presentation of the results from the training and testing as well as a comparison with other control approaches is given in section 5. The paper concludes with an outlook on possible improvements of the RL-AVD approach.

2. APPLICATION CASE: WHEEL LOADER

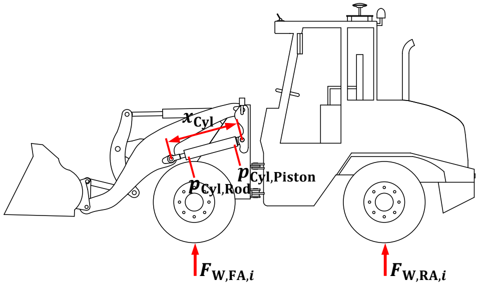
For wheel loaders of small power classes, the machine axles and parts of the working kinematics show the highest quantitative density of structural damage to components of these machines. This is the result of a scientific investigation in [7] that is based on maintenance and repair records. Since the vibrations in the axles lead to an increased component load and a loss of comfort for the machine operator, they constitute the component focus of this paper.

The considered machine is a wheel loader, which is mainly used for loading bulk material between two piles. In this publication, the passing over of obstacles on solid ground is investigated. Digging processes and their effect on the vibrations induction are not considered.

2.1. Application Machine L509 Speeder

The application machine is a wheel loader with a steering system combining articulated and rear wheel steering, an operating weight of 6.5 tonnes and a maximum payload of 1.8 tonnes. A hydrostatic drive with two speed levels is used for the traction drive. The working function is designed as a Z-kinematic system driven by a hydraulic pilot-controlled open-centre constant-flow system. To develop the RL-AVD approach a holistic machine simulation model has been developed for the L509 Speeder.

In order to validate the machine simulation, appropriate sensors were installed in a reference machine of the institute in order to record the relevant quantities to develop AVD approaches, see **Figure 1**.



Measured quantity	Variable
Wheel-load-force	$F_{\mathrm{W},i}$ in N
Pressure in lifting cylinder	$p_{\mathrm{Cyl},i}$ in bar
Extension stroke of lifting cylinder	x_{Cyl} in %

Figure 1: Reference machine and measured quantities

The calculation method for the machine axle damage is based on knowledge of the wheel and axle loads. To measure the wheel-load-forces, a strain gauge full bridge was applied to the machine axle on each wheel side. These determine the material strain due to shear stresses in the neutral phase of the axles and can be converted into wheel-load-forces by a suitable calibration. The setup is based on the approaches in [7–9]. Two wheel-load-forces are measured per axle, the sum provides the total axle load. Neglecting friction, the pressure in the rod and piston side of the lifting cylinders $p_{\mathrm{Cyl},i}$ are used to calculate the acting cylinder force F_{Cyl}. The cylinder extension stroke x_{Cyl} is measured by a laser sensor.

2.2. Driving Manoeuvres

Simple tracks were designed for developing, testing and validating the RL approach. They are characterized by a straight track with interchangeable obstacles. **Figure 2** shows the simplified setup of the tracks and configurations for the training and validation runs.

The training of the algorithm always takes place under constant conditions. The obstacles have the shape of a trapeze. The up and down gradients are equivalent, the length of the obstacle is smaller than the wheelbase. The lifting and tilting cylinders extension strokes are initially set in such a position that the lowest point of the bucket is 200 mm above the ground. For the validation process the number of obstacles and their positions are varied. For all test scenarios the mass of the bulk material in the bucket is $m_{\mathrm{PL}} = 1{,}500$ kg.

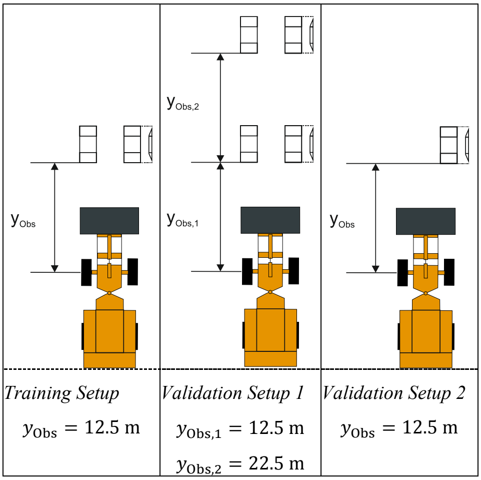

y_Obs
y_Obs,2
y_Obs,1
y_Obs
Training Setup
$y_{\mathrm{Obs}} = 12.5$ m
Validation Setup 1
$y_{\mathrm{Obs,1}} = 12.5$ m
$y_{\mathrm{Obs,2}} = 22.5$ m
Validation Setup 2
$y_{\mathrm{Obs}} = 12.5$ m

Figure 2: Training and validation setups

3. REINFORCEMENT LEARNING

3.1. General control approach

Reinforcement Learning is a machine learning method that learns through interaction with the environment. Trial and error are the basis on which the control-system (CS) learns an optimal behaviour for a given task; in case of the AVD for damping the axle load vibrations. The control-system consists of two main components:

- the environment, representing the evaluation part,
- and the agent, representing the learning part.

The system runs the same training track over and over again. Every run the agent can test different strategies for pressure pulses introduced into the lifting cylinder. At the end of every run, the vibration damping effectiveness is evaluated by the environment.

The whole problem is formalized as a Markov decision process. **Figure 3** shows the interaction between the agent and the environment, constituting the basic principle of reinforcement learning.

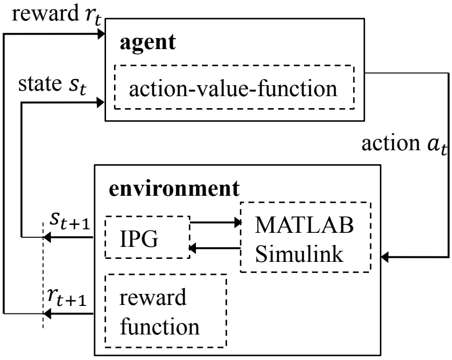


Figure 3: Reinforcement learning approach

For every time step $t = 0, 1, 2, \dots, \mathrm{T}$ the system including the agent is in a state s_t out of a finite set of possible states $S (s_t \in S)$. According to the current state s_t the agent selects an action a_t. The action space $A(s_t)$ is also finite and it applies $a_t \in A(s_t)$. For the action a_t the agent receives a reward r_{t+1} from the environment at the next time $t + 1$. At time $t + 1$ the system has the state s_{t+1} as a consequence of the chosen action a_t.

The environment is composed of the simulation framework and the reward function. It is not practicable to evaluate every action, because it is not known, how good a single action is. Therefore a delayed feedback is implemented after a completed training run.

The agent strives to maximize the sum of all received rewards. So the feedback of the environment leads the agent to actions, which are expected to be valued as positive. [10, 11]

3.2. State Space

The problem is described by a continuous state space. To simplify the matter, the space has been discretized. For the AVD-RL-CS, the state space is spanned of three different variables (s_1, s_2, s_3):

- s_1: the pilot pressure cylinder $p_{\mathrm{Cyl,Ctrl}}$ to control the lifting cylinder
- s_2: the force of the lifting cylinder F_{Cyl}
- s_3: the lifting cylinder extension x_{Cyl}

The objective of the learning process is to help the agent to learn how the control pressure can be used to actuate the lift cylinder so that the axle loads can be reduced. Therefore, the state s_1 provides partial information about the reaction of a chosen action. Of interest are the axle load vibrations. Typically, strain gages are not installed in series wheel loaders for measuring axle loads. However, the axle load vibrations behave similarly to the oscillations of the force at the lifting cylinder. This is due to the fact that the working kinematics and the vehicle chassis are connected without spring-damper elements. So the vibrations from the working hydraulics also affects the undamped attached axles. To describe the deviation of the cylinder force from its mean value $F_{\mathrm{Cyl,fil}}$, F_{Cyl} is filtered by a PT$_1$ element (high pass filter).

The bucket should be prevented from touching the ground. This information is contained in the lifting cylinder extension.

The chosen discretisation is listed in **Table 1**. Nearest-neighbour interpolation was implemented for the assignment of the states s_1 and s_2 , for state s_3 piecewise constant interpolation.

Table 1: Implemented state space

State	Discretization
s_1: $p_{\mathrm{Cyl,Ctrl}}$	[-57, -40, -27, 0, 27, 40, 57] %
s_2: $F_{\mathrm{Cyl,fil}}$	[-30 : 10 : 40] kN
s_3: x_{Cyl}	[0.03, 0.08, 0.15] mm

3.3. Action Space

The action space $A(s_t)$ is discretized like the state space. The delay between control signal and valve movement is modelled by a PT$_1$T$_t$ element. These delays lead to the choice of an action space with five values $A(s_t)$ = [-40, -33, 0, 40 60] % of the pilot pressure for the lifting cylinder.

An action is selected with a defined frequency $f = 400$ Hz that is significantly higher than the frequency of the axle load vibrations, see **Figure 6**.

At the beginning of the training, the agent does not know which action leads to a big reward. Therefore, he has to do some trial and error.

For the AVD-RL-CS, the decreasing-ε-strategy is implemented. With this strategy the learning process starts with a high exploration rate, which is reduced over time. The value ε determines whether the agent chooses an explored action through choosing a random action or a profitable action he already knows. For ε = 0 the agents is greedy and takes profitable actions, for ε = 1 a random action $a_t \in A(s_t)$ is taken and so the agent explores the environment. [11]

As start value ε = 0.9 is set, the final value after 850 runs is close to 0. At this point, the agent only uses his knowledge and can maximize the sum of rewards. The knowledge will converge to an optimal strategy.

3.4. Reward

The reward indicates how positively or negatively the environment evaluates the chosen action by the agent.

For the AVD, the variant of a delayed feedback is implemented. The reward is defined by the results of the linear axle-damage-accumulation according to Miner elementary. The results are compared with those of a non-damped system. Four different cases are implemented:

- The agent will receive the maximum punishment, if a termination criterion is offended. These criteria are defined so that the agent does not leave the required bucket height.
- The second highest penalty is assigned for boosting the front and the rear axle load. If the agent only improves one axle load, whether

```
Initialize $Q(s,a)$, for all $s \in S$, $a \in A(s)$, arbitrarily, and $Q$ (terminal-state, $\cdot$) = 0
Repeat (for each episode):
    Initialize $S$
    Repeat (for each step of episode)
        Choose $a$ from $s$ using policy derived from $Q$ (ε-greedy)
        Take action $a_t$, observe $r_t, s_t$
        $Q(s,a) \leftarrow Q(s_t,a_t) + \alpha\,[R + \gamma \max_{a} Q\,(s_{t+1},a_t) - Q(s_t,a_t)]$
        $s_t \leftarrow s_{t+1}$
    until $S$ is terminal
```

Figure 4: Pseudocode Q-Learning

for the front or the rear axle, he receives a smaller penalty.

- The only positive reward the agent receives is when he improves both damage values for the axle load vibrations.
- Additionally to these huge rewards, a small positive one has been implemented for each action choice that does not infringe a termination criterion.

The rewards are saved using an action-value-function $Q(s,a)$.

3.5. Algorithm and Action-Value-Function

The agent is trained by a Q-Learning algorithm. This is an off-policy algorithm, which is based on Temporal-Difference-Learning. It is a model free algorithm. [10]

Figure 4 shows the pseudo code according to [11]. The agent observes his state s_t and uses it to select his next action a_t, which, according to the action-value-function $Q(s_t,a_t)$, is the most promising. The received reward r_t and the following state s_{t+1} is used to update the action-value-function $Q(s,a)$. For the update, two parameters are implemented:

- α: This parameter specifies how strongly the existing action-value-function is corrected by the new reward values.
- γ: The discount rate defines the actual value of the future rewards. For $\gamma = 0$ the agent is only interested in maximizing the immediate following reward. He is “myopic”. For a higher value of γ, the agent is more farsighted and considers future rewards more strongly in the current action choice. If $\gamma < 1$, the sum of the return R has a finite value for non-episodic tasks. [11]

The agent with its learning algorithm is implemented as a Stateflow-diagram in MATLAB-Simulink [12].

4. HOLISTIC MACHINE MODEL

A holistic machine simulation model of the described wheel loader is set up to develop the AVD-RL-CS. The model considers the hydrostatic traction drive (HTD), the function and steering drives (FD and SD), the multi-body dynamics and the interaction with a 3D-environment. The hydraulics of the traction, function and steering drives are modelled in MATLAB-Simscape-Fluids, the multi-body simulation (MBS) models of the working and steering kinematics systems in MATLAB-Simscape-Multibody. The individual models were parameterized by manufacturer specifications and by measurements using the reference machine. **Figure 5** shows a schematic of the machine model and the coupled variables between the subsystems.

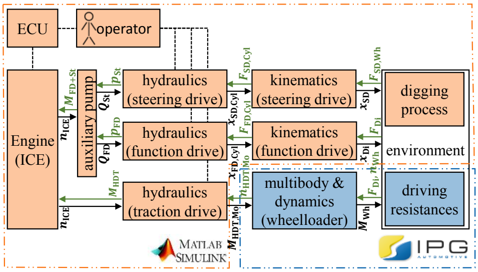


Figure 5: Holistic Machine Model

The IPG-TruckMaker 3D-environment is integrated into the coupling to consider the elastic and damping properties of the entire machine, which are largely defined by the tyres. The tyre-ground-contact is considered by the data interpolation based contact model IPG-Tire [13]. The model uses a single-point-contact between tyre and ground, but offers the advantage of short computing times and numerical stability. The forces transmitted from the lifting mechanism to the front carriage of the machine (connections of the boom and the hydraulic cylinders) result in changing axle loads and tyre deformations. The resulting 3D-position-change of the kinematic linkages serves as an input to the MBS of the working kinematic. The virtual 3D-environment calculates the resulting driving dynamic variables, such as wheel speeds and steering forces, and transfers them to the simulation models of the HTD and the SD in Matlab-Simscape. This complex coupling of simulation models allows the testing of driving and working scenarios like they occur in reality as well as scenarios under load conditions that would otherwise be difficult to reproduce. A detailed description of the validation of axle load vibrations is provided hereafter. A validation of the traction and function drives are part of another contribution within the scope of this conference [14].

4.1. Validation of Axle Loads

For the validation of the holistic machine model, the front axle loads are analysed when passing an obstacle on the training course with $m_{\mathrm{PL}} = 1{,}510$ kg. The lifting cylinder extension is $x_{\mathrm{Cyl}} = 10$ %, the tilting cylinder is extended to its maximum stroke. This means that in a standstill position the lowest point of the bucket is approximately 240 mm above the ground in a fully tilted position and thus represents a realistic load case. The model has been compared and validated with respect to the real machine behaviour.

Figure 6 shows the exemplary results of the front-axle-forces in direction of gravity for crossing the obstacle. The upper graph shows the front axle (FA) loads, the middle graph the rear axle (RA) loads and the two lower graphs the results of discrete fast Fourier transformations of the FA and RA loads. The blue lines correspond to the experimental data and the dashed-dotted red lines to the simulation results.

In the measurement, the front axle load rises slightly as the machine drives onto the obstacle ($t = 12.2$ s). The force on the rear axle behaves in the opposite direction, as it is decreasing.

In simulation, the vibration initiation is much stronger than in reality. This is with high probability due to the simplified single-point-contact-model of the tyre. As a consequence, the resistance of the ground only has an influence when it is below the centre of the tyre. This leads to an abrupt change in the position of the tyre-ground-contact-point and thus to a sudden load build-up, as can be seen in the simulation in the period between $t = 12.5$ s and $t = 12.7$ s. The maximum deviation during this period, defined as

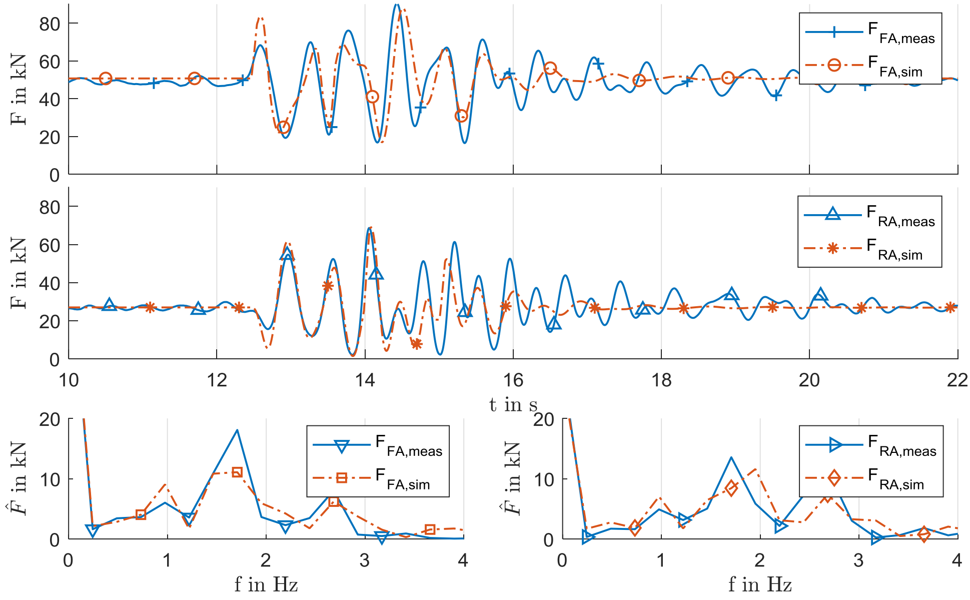


Figure 6: Axle loads for threshold crossing with all wheels

the difference between simulation and measurement divided by the measurement value, is 22 % for the front axle and 64 % for the rear axle.

In the following the mass shifts to the rear axle, whereupon the front axle forces decrease to a minimum and the rear axle force to a maximum ($t = 12.9$ s). When the front wheels have passed the obstacle ($t = 14$ s), the axle load of the rear axle rises to its maximum. The front axle reaches its global maximum at $t = 14.4$ s.

When the rear wheels have passed the obstacle completely, the axle load of the rear axle rises to its second highest amplitude ($t = 15.2$ s). With a slight deceleration, the front axle experiences its global minimum with a following maximum. After three seconds, the axle loads have settled in simulation. In the measurement, the decay of the vibration lasts five seconds.

The frequency of the occurring vibrations are similar in simulation and measurement. This can also be seen in the transformations of the vibrations into the frequency domain by a standardized, discrete fast Fourier transformation. The main frequencies in the measurement and simulation are between 0.7 and 2.7 Hz.

The deviations in the amplitudes are mainly caused by the simplified single-point-contact-model of the tyres. The parameters of this model have been determined on the basis of literature values and a similar scenario of an obstacle crossing with $m_{\mathrm{PL}} = 970$ kg.

However, the sequence of maxima and minima of the described events corresponds to the results of the measurement. In general, it can be concluded that the simulation represents the essential axles force vibration parameters such as amplitudes, frequency and decay time well.

4.2. Machine Axle Damage Model

The machine model is extended by damage models of different wheel loader components. To relate the wheel and axle forces to a damage value the bending beam theory is used. The axle is simplified as a construction of square tubes with different external and internal dimensions. The force is applied at the wheel mounting points. The centre of the axle is mounted to the machine frame, which counteracts with the force F, see **Figure 7**.

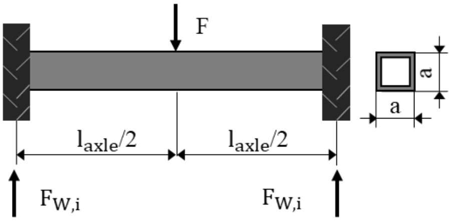


Figure 7: Damage model of the machine axles

For this load case, the maximum bending moment occurs at the fixed clamping to the machine frame and can be calculated as follows:

$$M_{\mathrm{B}} = F_{\mathrm{W},i} \cdot \frac{l_{\mathrm{Axle}}}{2} \qquad (1)$$

The material stress due to the bending can be calculated by determining the resistance moment W_{Axle} from the second moment of inertia I_{Axle} of the square tubes:

$$\sigma_{\mathrm{B}} = \frac{M_{\mathrm{B}}}{W_{\mathrm{Axle}}} = \frac{F_{\mathrm{W},i} \cdot \frac{l_{\mathrm{Axle}}}{2}}{\frac{2 \cdot I_{\mathrm{Axle}}}{h}} = \frac{F_{\mathrm{W},i} \cdot l_{\mathrm{Axle}} \cdot h}{4 \cdot I_{\mathrm{Axle}}} \qquad (2)$$

The loads occurring during a manoeuvre are separated into individual vibrations by the MATLAB rainflow counts according to the ASTM E 1049 standard [15]. The partial damage is calculated using the elementary form of the Miner rule [16]. The Woehler exponent k, which is directly related to the damage, is assumed to be $k = 5$ on the basis of the FKM guideline [17]. The sum of the partial damages results in the total damage of the respective axles D_i.

With this approach geometries and notch effects are simplified. Further the bending load case represents the main load case, but there are occurring others which are not taken into account. So the determined damage values represent the damage effects just to a limited extend and can therefore only be regarded as approximate values.

5. RESULTS

5.1. Training

In **Table 2** the results of the training are shown. The system was trained through 851 training runs (TR). In 19 TRs the training was stopped, because the actions of the agent lead to an abort criterion. 255 TRs have led to an increase in the damage values for the axle loads, in some TRs up to 160 %. In 392 TRs only one axle load has been optimized, see case 2. The training results show that often one axle load could be significantly

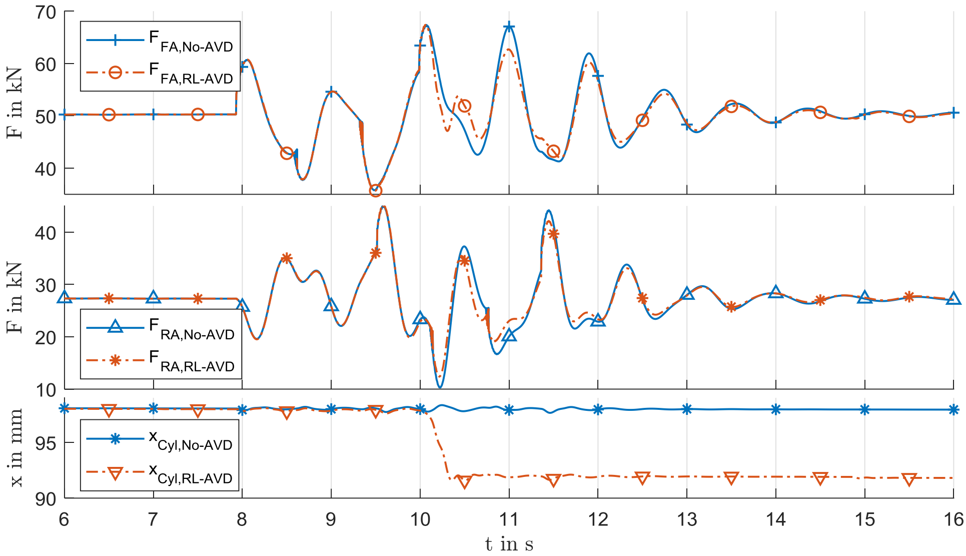


Figure 8: AVD using RL: Training Setup

reduced, while the second one was not necessarily worsened. The agent was able to reduce both axle loads in 185 TRs.

Table 2: Results TS

Case	Reward	Proportion of TRs [%]
1	Both axle loads optimized	21.7
2	One axle load optimized	46.1
3	No axle load optimized	30.0
4	Abortion	2.2

One exemplary training result is shown in **Figure 8**. The upper graph shows the axle loads on the front axle, the middle graph the rear axle loads and the lower graph the lifting cylinder extension. The blue lines correspond to the results of a non-damped system, the dotted red lines to the results of the damped system by the RL agent. For $t = 0\ s$ the wheel loader starts driving. The vehicle moves up the obstacle at $t = 8\,\mathrm{s}$. At $t = 12.5\ s$ the wheel loader has passed the obstacle.

The strategy, the agent is pursuing, is one of lowering the bucket. Therefore, the cylinder extension is reduced about $\Delta x_{\mathrm{Cyl}} = 6\,\mathrm{mm}$. This continuous lowering is leading to reduced force maximums of the front axle. The highest reduction is about 4.3 kN at $t = 11\,\mathrm{s}$. In conclusion the damage value for the front axle is about 15 % smaller compared with a non-damped system. Also the rear axle load could be reduced in several maximums. For the global maxima at $t = 11.5$ s the reduction is about 2 kN. In sum, the damage value for the rear axle is reduced by 24 %.

The action selection by the agent is leading to additional pressure oscillations in the lifting cylinder chambers and in the pilot pressure for the lifting cylinder. For the pilot pressure the values vary between -35 % (lowering) and 30 % (lifting) of the maximum control pressure. The frequency with which the agent chooses the several actions is reflected here. For lower frequencies, the agility of the system decreases too much.

5.2. Validation

For the validation three different setups (One TS and two VSs) were used (cf. **Figure 2**).

The results are compared to a further AVD approach. Research well known approaches are AVD-CS using techniques from linear control engineering. An essential element of these approaches is the feedback of the current signal to the controller, which continuously counteracts any deviation from the setpoint [2–4]. In this publication, these approaches are transferred to the axle load vibrations in a proportional and proportional-integral (P-PI controller) form. The controller uses the wheel-load-forces of the front axle F_{FA} and lifting cylinder extension as input variables. F_{FA} is cleared from the mean value by a high pass filter and thus provides the control deviation. The working kinematic system is actuated to counter-excitations, e.g. when F_{FA} increases by lowering the mechanics. In order to keep to the nominal extension of the lifting cylinder, the current lifting level is compared with the setpoint desired by the operator. The proportional and integral of the PI controller part is used to keep the piston position along the setpoint value over time. The output of the controller ($u(t)$) is the sum of the control-components and is therefore a combination of damping and position keeping. Mathematically, this can be considered as the sum of the proportional ($K_{\mathrm{P}}, K_{\mathrm{S}}$) and integral parts ($K_{\mathrm{I}}$), cf. formula 3.

$$u(t) = K_{\mathrm{P}} \cdot F_{\mathrm{W}}(t) + K_{\mathrm{S}} \cdot x_{\mathrm{Cyl}}(t) + K_{\mathrm{I}} \int_0^t x_{\mathrm{Cyl}}(\tau) d\tau \qquad (3)$$

The parameters K_{P}, K_{S} and K_{I} were determined by full factorial parameterization while passing the training course several times. Instead of the wheel-load-forces $F_{\mathrm{W},i}$, it would also be conceivable to use the cylinder forces F_{Cyl} as the controller input. However, using F_{Cyl} the authors' results for vibration damping and thus for damage reduction were significantly lower.

Another approach is the cumulative combination of the two approaches (P-PI and RL approach). The optimally working P-PI controller is used to reduce the fundamental vibration, the RL controller is used for fine adjustment and consideration of individual valve characteristics.
The results of all approaches for the driving manoeuvers listed in **Figure 2** are shown in **Table 3**. Compared to the undamped system, positive values describe a reduction of the damage, negative values describe an increase of the damage.

For the TS the RL agent can reduce the damage value about 4.9 % for the front axle. The damage of the rear axle remains unchanged. The agent uses a lowering strategy for the bucket. For the VS 1 (overrun two obstacles) and VS 2 (overrun one obstacle with only the left tyres) the damage values are reduced by a maximum of 1.9 %.
The linear control approach reduces the damage to individual axles by up to 77.7 % during the same manoeuvres. However, this factor is 15 times higher than the achieved results of the RL approach. The reason is the known fact in the controller design that the vibrations in the axles are compensated by counter-excitations. The RL approach must learn this knowledge on its own. The P-PI approach maintains the original lifting height, this behaviour is not known to the RF approach either and has to be learned as well.

The RL-P-PI-combination is able to reduce the damage to individual axles by up to 70.6 % (TS 1 Rear Axle) during the same manoeuvres. This is a further improvement of 31.2 % compared to the P-PI-approach. However, it also needs to be mentioned that compared to the P-PI approach, in the mentioned manoeuvre the reduced damage of the rear axle is accompanied by a small increase in front axle damage (-7.0 %).

It could be shown that a self-learning system is able to learn a damage-reducing behaviour. However, the chosen learning algorithm of Q-Learning reaches its limits due to the chosen number of training runs, the training manoeuvres itself and the discretization of the state space. An expansion of the state space and the use of advanced learning algorithms, such as SARSA, DQN and DDPG, could lead to further damage reduction of the machine axles. Nevertheless, it will be challenging to achieve the reduction potential of classical linear control techniques with self-learning systems. The coupling of the individual approaches shows the most promising results for the damage-reducing-application.

Table 3: Damage Reduction Results

	RL [%]	P-PI [%]	RL-P-PI [%]
	TS		
Front Axle	4.9	65.6	58.6
Rear Axle	0.2	39.4	70.6
	VS 1		
Front Axle	-20.5	77.7	55.9
Rear Axle	0.0	38.7	33.4
	VS 2		
Front Left Axle	-0.8	2.7	-3.4
Front Right Axle	1.9	0.1	-45.5
Rear Left Axle	0.0	37.0	36.7
Rear Right Axle	0.0	-12.4	23.0

6. CONCLUSION

This contribution has presented a new and innovative reinforcement learning (RL) approach for an active vibration damping and damage reduction of wheel loader axles. It could be shown, that self-learning approaches are capable of learning a control behaviour that leads the machine to reduced-damage situations during operation. This approach has been developed using a holistic machine simulation model. A validation of the axle load vibrations using measurement data from a reference machine shows a good agreement between measurement and simulation.

Using reproducible training and validation scenarios, the self-learning system was trained, tested and validated. A maximum axle damage reduction of 4.9 % was achieved for the considered training and validation runs. The comparison between a linear control approach (P-PI-Controller) and the Reinforcement Learning system shows the potential of vibration damping to be achieved. Coupling of these two approaches shows the best damage reducing results, while the P-PI controller serves as the basic controller and the RL approach includes the properties of the valve characteristic behaviour.

In addition to the execution of further training runs, current work focuses on the implementation of the following optimization approaches.

An improvement of the algorithm could be achieved, if at the beginning of the learning process the RL approach would have information about an effective damping behaviour. For this the control signals and resulting axle loads from a linear control approach (P-PI-Controller) could be used.

In addition or instead of considering absolute cylinder forces, it could be more effective to consider the force gradients. Thus, the RL approach would not have to learn the relationship between these parameters independently, but would receive them directly as an input.

The transfer of the self-learning system from simulation to a real machine is still pending.

NOMENCLATURE

A	Action Space
AVD	Active Vibration Damping
CS	Control System
Cyl	Cylinder
D_i	Total Damage of Axle i
Di	Digging Process
F_{Cyl}	Lifting Cylinder Force
$F_{\mathrm{Cyl,fil}}$	Deviation of Lifting Cylinder Force from Mean Value
$F_{\mathrm{W},i}$	Wheel-Load-Force of Wheel i
FA	Front Axle
FD	Function Drive
HTD	Hydrostatic Traction Drive
I_{Axle}	Axle Second Moment of Inertia
ICE	Internal Combustion Engine
K_i	Factor i of P-PI-Controller
M	Torque
M_{B}	Bending Moment
MBS	Multi Body Simulation
P	Proportional
PI	Proportional-Integral
PVD	Passive Vibration Damping
Q	Action Value Function
RA	Rear Axle
RL	Reinforcement Learning
S	Set of Possible States
SD	Steering Drive
TS	Training Setup
TR	Training Run
VS	Validation Setup
W_{Axle}	Resistance Moment of Machine Axle
Wh	Wheel
a_t	Action at Time t
f	Frequency
fil	Filtered
k	Woehler Exponent
l_{Axle}	Machine Axle Width
m_{PL}	Payload in Bucket
$meas$	Measurement
n	Speed
$p_{\mathrm{Cyl},i}$	Pressure in Lifting Cylinder Chamber i
$p_{\mathrm{Cyl,Ctrl}}$	Pilot Pressure of Lifting Cylinder
r_t	Reward at Time t
s_t	State at Time t
sim	Simulation
t	Time
$u(t)$	Controller Output
x_{Cyl}	Lifting Cylinder Extension
$y_{\mathrm{Obs},i}$	Distance to Obstacle i
α	Learning Rate
ε	Greed Factor
γ	Discount Rate

ACKNOWLEDGEMENTS

The authors want to express special thanks to the IPG Automotive GmbH for providing the license for the IPG-TruckMaker program.

REFERENCES

[1] Latour C, Biener R (2003) Schwingungstilgung in Radladern - Vergleich von aktiven und passiven Systemen. Ölhydraulik und Pneumatik 47(3): 171–174

[2] Alexander A, Vacca A, Cristofori D (2017) Active Vibration Damping in Hydraulic Construction Machinery. Procedia Engineering 176: 514–528. doi: 10.1016/j.proeng.2017.02.351

[3] Bianchi R, Alexander A, Vacca A (2016) Active Vibration Damping for Construction Machines Based on Frequency Identification. In: SAE 2016 Commercial Vehicle Engineering Congress, SAE Technical Paper 2016-01-8121. SAE International400 Commonwealth Drive, Warrendale, PA, United States

[4] Madau R, Vacca A (2019) Active Ride Control for Construction Machines Based on Pressure Feedback. In: ASME/BATH 2019 Symposium on Fluid Power and Motion Control. American Society of Mechanical Engineers

[5] Rahmfeld R, Ivantysynova M, Eggers B (2004) Active Vibration Damping for Off-Road Vehicles Using Valveless Linear Actuators. In: SAE International400 Commonwealth Drive, Warrendale, PA, United States

[6] Williamson C, Lee S, Ivantysynova M (2009) Active Vibration Damping for an Off-Road Vehicle with Displacement Controlled Actuators. International Journal of Fluid Power 10(3): 5–16. doi: 10.1080/14399776.2009.10780984

[7] Bös M (2015) Untersuchung und Optimierung der Fahrkomfort- und Fahrdynamikeigenschaften von Radladern unter Berücksichtigung der prozessspezifischen Randbedingungen. KIT Scientific Publishing

[8] Küppers RT (2000) Untersuchungen zur Kippstabilität von Radladern. Dissertation, RWTH Aachen

[9] Hoffmann K (1987) Titel: Eine Einführung in die Technik des Messens mit Dehnungsmeßstreifen, 8th edn. Hottlinger Baldwin Messtechnik, Darmstadt

[10] Russell SJ, Norvig P (2016) Artificial intelligence: A modern approach, Third edition, Global edition. Always learning. Pearson, Boston, Columbus, Indianapolis

[11] Sutton RS, Barto A (2018) Reinforcement learning: An introduction, Second edition. Adaptive computation and machine learning. The MIT Press, Cambridge, MA, London

[12] MathWorks Maker Team (2018) Reinforcement learning with Self-balancing motorcycle. https://de.mathworks.com/matlabcentral/fileexchange/68396-reinforcement-learning-with-self-balancing-motorcycle. Accessed 13 Nov 2019

[13] Schieschke R, Wurster U (1988) IPG-TIRE Ein flexibles, umfassendes, Reifenmodell fuer den Einsatz in Simulationsumgebungen. Automobil-Industrie 33(5): 495–500

[14] Brinkschulte L, Pult F, Geimer M (2020) The Use of a Holistic Machine Simulation for the Development of Hydraulic Hybrid Modules to Reduce Transient Engine-Out Emissions (submitted, not yet published). In: Proceedings of 12th International Fluid Power Conference

[15] E08 Committee E1049 − 85: Practices for Cycle Counting in Fatigue Analysis

[16] Haibach E (2006) Betriebsfestigkeit: Verfahren und Daten zur Bauteilberechnung, 3., korrigierte und erg. Aufl. VDI-Buch. Springer, Berlin

[17] Rennert R, Kullig E, Vormwald M et al. (2012) Rechnerischer Festigkeitsnachweis für Maschinenbauteile aus Stahl, Eisenguss- und Aluminiumwerkstoffen, 6., überarb. Ausg. FKM-Richtlinie. VDMA-Verl., Frankfurt am Main