Proceedings of the 2022 Winter Simulation Conference
B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C.G. Corlu, L.H. Lee, E.P. Chew, T. Roeder, and P. Lendermann, eds.

TOWARDS DEADLOCK HANDLING WITH MACHINE LEARNING IN A SIMULATION-BASED LEARNING ENVIRONMENT

Marcel Müller
Tobias Reggelin
Iegor Kutsenko
Hartmut Zadek

Institute of Logistics and Material Handling Systems
Otto von Guericke University Magdeburg
Universitätsplatz 2
Magdeburg, 39106, GERMANY

Lorena S. Reyes-Rubiano

School of Economic and
Administrative Science
Universidad de La Sabana
Chia, 250001, COLOMBIA

ABSTRACT

The planning of complex logistic systems must ensure collision- and deadlock-free operation of the logistic system. Problem-specific rule-based algorithms used so far are inflexible with respect to infrastructure changes and scale poorly with systems that grow larger. This paper shows a first approach to handle logistic deadlocks with machine learning. We present a conceptual approach on how to handle logistic deadlocks with artificial neural networks. The paper also provides a technical implementation with a single agent approach based on reinforcement learning with deep Q-networks. A discrete event simulation of an automated guided vehicle system is used as the learning environment. The first results show that artificial neural networks can learn to handle deadlock capable logistic systems with low complexity.

1 PROBLEM AND MOTIVATION

Due to the increased use of automated systems in production and logistics, the planning and control of the respective systems are becoming more complex. As a result, the use of resources within the system becomes opaquer and more dynamic. This makes it more difficult to detect circular dependencies in demand on resources in planning. The missing detection of circular dependencies can lead to more occurrences of deadlocks in the automated material flow systems, which can bring the entire logistics system to a standstill through chain reactions. Deadlocks occur when at least two processes wait for each other in the form of a circular reference and cannot perform the next process step due to infeasible conditions (Coffman et al. 1971; Tanenbaum 2015). There are several types of deadlocks. This paper will only consider so-called resource deadlocks (Tanenbaum 2015) and will be simplistically referred to as deadlock in the following. Deadlocks can also occur when not all resources in a system are used (Coffman et al. 1971).

Simulation models can help detect deadlocks and develop appropriate solution strategies. Due to the numerous random experiments, even very improbable unusual cases can be discovered, which are not apparent from the outset or analytically determinable in a complex logistics system. In logistics, deadlocks are often handled with problem-specific control rules (Mayer and Furmans 2010; Müller, Schmidt, and Reggelin 2019) or with preventive, deterministic measures through planning and reservation of resources (Kim et al. 2006; Lienert and Fottner 2017). There needs to be sufficient information about the current situation in a logistic system for the reservation systems to ensure a correct scheduling of the reservation slots.

Both approaches require a central control approach and scale poorly with larger scaled systems. For large systems, it is difficult to achieve complete coverage of all deadlock situations with problem-specific control rules, since there may be unusual cases that occur only very rarely and are therefore not detected in advance of planning. Reservation systems, on the other hand, with their deterministic approach, move further away from reality for larger systems, as special disturbance events are more likely to occur. Creating optimal reservation schedules can lead to high computation times when many processes and resources result in numerous possible combinations of resource reservations. Furthermore, the changing availability of processes and resources due to the dynamics of the logistics scenario provides new optimal solutions even more frequently, so that recalculations of the reservation plan become necessary even more often.

Decentralized control approaches in logistics that promise such scalability have so far also relied on problem-specific avoidance rules through optimization algorithms for deadlocks (Schönung et al. 2011; Seibold 2016) and agent-based approaches (Forget et al. 2009; Mors 2010; Yalcin 2017; Lu et al. 2019).

Another problematic aspect of dealing with deadlocks is that even small changes to logistic parameters can change the best strategy approach between prevention, avoidance, and detection & resolution (Müller et al. 2020). The poor scalability and flexibility with respect to structural changes of previous deadlock solutions will therefore be addressed with solutions from the field of machine learning. The hypothesis is that trained AI agents that take on decision problems relevant to the emergence of deadlock will provide better flexibility and scalability. It is also possible that such an approach would render the categorization of the strategy approaches for deadlock handling prevention, avoidance, and detection & resolution obsolete.

2 LITERATURE

2.1 Conventional Deadlock Handling

There are many problem-specific approaches in literature for deadlock handling. Most of these approaches do not come from logistics but from computer science, where this problem has been recognized and discussed since the late 1960s and early 1970s (Havender 1968; Coffman et al. 1971; Holt 1972). Here, deadlocks occurred primarily in parallel/multi-programming, and general strategy approaches were developed to break the conditions for deadlocks to occur. These four conditions have always been considered as starting points for dealing with deadlocks (Coffman et al. 1971) and are also used in current problems in production and logistics (Palmer et al. 2018; Bashir et al. 2018; Zheng et al. 2020). In addition to focusing on breaking one of these deadlock conditions, a distinction is made between three basic, strategic approaches to handle deadlocks (Coffman et al. 1971):

- Prevention
- Avoidance
- Detection & Resolution

Both early considerations by Coffman et al. (1971) and later views by, for example, Tanenbaum (2015), from the perspective of informatics, saw in the strategic approach to prevention primarily changes in the way resources are claimed. They referred to proposals by Havender (1968). This view was not always followed in production and logistics. Either prevention was not clearly distinguished from avoidance to the point of being used synonymously (Kim et al. 2006; Mayer and Furmans 2010), or, like Lehmann (2006), deadlock prevention was understood to mean the design of a logistics system, process, and infrastructure in such a way that deadlock could no longer occur. In designing the system, reference was often made only to the rules for resource utilization (Kim and Kim 1997; Lienert and Fottner 2017), instead of seeing the system structure and dimensioning of the system's capacities as an integral part of deadlock prevention. The mutual exclusion condition of Coffman et al. (1971) in particular could be solved by clever infrastructure planning, which has not been discussed in detail in the literature so far. Deadlock detection first relies on a description of the system state. Here, graph-theoretic approaches are mainly used in the literature. The types of graphs are primarily resource allocation graphs (Holt 1972) and Petri nets (Petri 1962). Various applications of Petri nets for deadlock handling can be found in literature (Liu 2016; Sun et al. 2018; Luo et al. 2019).

The type of application or the actual design of the deadlock situation plays an important role. In contrast to the problems in production and logistics, in information technology, for example, a processor can quickly release and reallocate its resources to resolve a deadlock situation. In logistics processes, this is often not possible or only possible under certain circumstances and thus increases the effort required to resolve a deadlock.

2.2 Deadlock Handling with Machine Learning

Machine learning has regained importance in recent years due to significant advances in research. In the field of image recognition, efficiently implemented convolutional neural networks have now been able to match human image recognition performance with the help of powerful graphics processors (Cireşan et al. 2012). Extremely large neural networks such as the GPT-3, with 175 billion parameters, have been shown to be highly versatile in handling a variety of complex text tasks (Brown et al. 2020). In the area of reinforcement learning, games such as Go or StarCraft, which were said to require human intuition due to numerous courses of action, also managed to outperform top human performers (Mnih et al. 2015; Silver et al. 2016; Vinyals et al. 2019), in some cases without ever having played against a human (Silver et al. 2017). These advances have also led to increasing research into what control tasks and decisions can be performed by artificial neural networks in production and logistics. Example use cases include scheduling in production systems (Waschneck et al. 2018; Zhou et al. 2021) or resource management in logistics networks (Li et al. 2019; Yu et al. 2021).

In contrast, the occurrence of deadlocks in the context of machine learning has only been briefly addressed (Bouderba and Moussa 2019; Bouton et al. 2019; Kujirai and Yokota 2019; Reijnen et al. 2020). A more intensive investigation of how occurring deadlocks affect the learning behavior of machine learning algorithms and whether AI agents actually learn to deal with deadlocks has not yet been comprehensively elucidated. Statements on this have only been made by Sørensen et al. (2020). The authors implement a reinforcement learning approach with a dueling deep Q-Network architecture for an airport baggage system augmented with a deadlock avoidance algorithm. The AI agent reduced the number of deadlocks but did not avoid them entirely. In particular, as the volume of baggage increased, deadlocks occurred more frequently and caused the episode to restart in the learning environment.

2.3 Discrete Event Simulation as Learning Environment

Recently, reinforcement learning studies have integrated simulation techniques as learning environments. Simulation techniques allow reinforcement learning models to improve performance and to save computational time and effort (Rabe and Dross 2015). The most recent reinforcement learning and simulation studies define optimal simulation environments to accelerate the learning process (Kim, Jang, and Kim 2021).

Reinforcement learning and simulation integration is mainly used for automated industrial environments. Below are some applications of reinforcement learning and simulation:

- **Interaction between robots and humans**: Reinforcement learning teaches robots to interact with humans physically. In this field, simulation techniques allow emulating different social patterns of behaviors (Akalin and Loutfi 2020; Liu et al. 2021).
- **Service robots**: Reinforcement learning applies to home robots to automatically learn to provide a service in a home. For example: cleaning tasks (Zhang et al. 2021), assisting people with disabilities in a variety of activities of daily living, such as dressing (Zhang et al. 2021; Farhan et al. 2020).
- **Autonomous robots**: reinforcement learning enables the robots or autonomous vehicle to make driving decisions with minimal risk (Zhao et al. 2022; Wang et al. 2021).

Studies have shown that integrating reinforcement learning and discrete event simulation works very well for dynamic problems involving multi-agent. Discrete event simulation has been widely used to

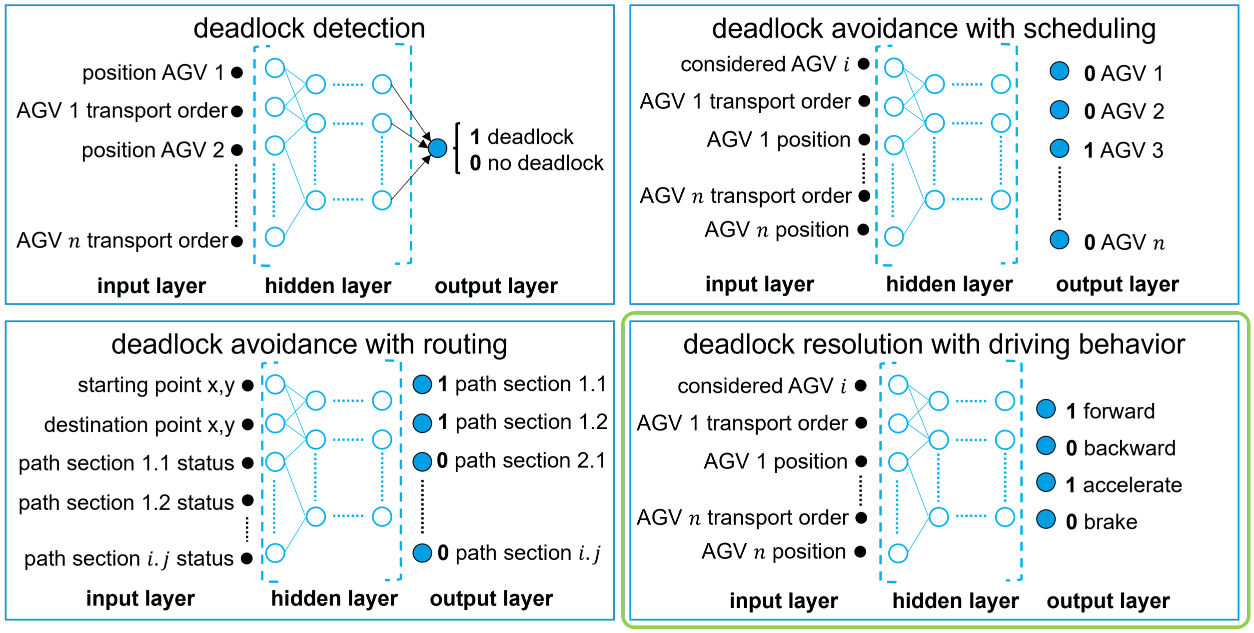

Figure 1: Example applications of artificial neural networks to deal with deadlocks in an automated guided vehicle system.

determine the matrix of states and actions for multiple agents. The matrix of states and actions allows calculating the expected reward for each state and action. The set of actions with highest reward determines the target state for reinforcement learning (Capocchi and Santucci 2022; Canese et al. 2021).

3 METHODOLOGY

3.1 Machine Learning

There are various approaches to use machine learning for deadlock handling and this also raises a whole series of questions. The first decision to be made is whether a supervised, reinforcement learning or unsupervised learning approach is most appropriate. We first propose a reinforcement learning approach, as we consider the creation of labeled data to teach the agent to be difficult. In very deterministic systems with very good information, a decision could be clearly evaluated whether it would lead to a deadlock or not. The dynamics of the system's behavior also argue for a reinforcement learning approach. We did not initially consider an unsupervised learning approach.

Another aspect is which decision should be taken by the agent that is relevant for the occurrence of a deadlock in a logistic system. An artificial neural network could take over various decisions, which would determine the inputs and outputs of the artificial neural network. The reinforcement learning algorithm uses a matrix with structure (state, action). The reinforcement learning algorithm stores the best action performed given the current state. The information (state, action) is updated after each event. Based on the information available, the neural network learns about the state-action pair and predicts the value of the reward. More specifically, the neural network seeks to learn an action policy that maximizes the total reward. Figure 1 gives an overview of possible applications of artificial neural network for deadlock handling in an automated guided vehicle system. For our approach we considered the usage of an artificial neural network to learn the driving behavior of the AGVs to ensure the avoidance of collisions and deadlocks.

We propose to use Markovian Decision-Making Process to determine the matrix of states and actions. The discrete event simulation is used as a modeling tool for the Markovian Decision-Making Process. Given the matrix, the state with highest reward is determined to be used as a target state for reinforcement

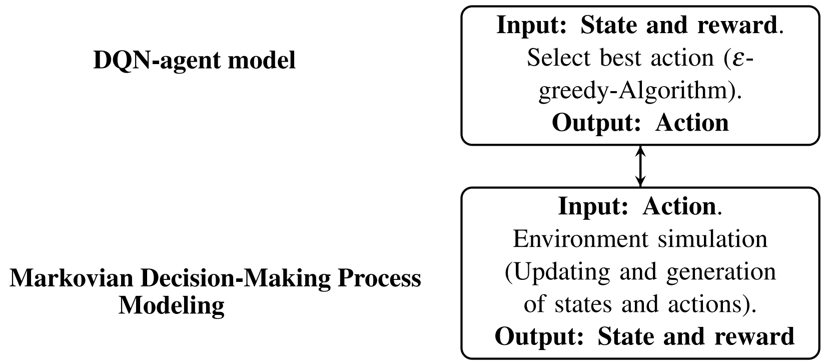

Figure 2: Discrete event simulation for reinforcement learning.

learning. The best state and expected reward are transferred to the DQN-agent model to determine the best action given the state and action policy (ε-greedy-Algorithm). Once the DQN-agent model determines the best action for the given state, the reward achieved is used to update the state-action matrix and the expected reward for each state. Figure 2 describes the integration of reinforcement learning and discrete event simulation.

3.2 Conceptual Model

In order to apply machine learning methods to deal with deadlocks, the considered logistic system must be deadlock capable. This means that deadlocks must be able to occur and must not be excluded from the outset by a prevention strategy. Since this is our first approach using machine learning to deal with deadlocks, the logistic system was deliberately kept simple and we avoided unnecessary complexity. This is reflected in a small number of processes and resources in the considered logistic system and the avoidance of deadlocks across different resource types.

Adapting Müller et al. (2020) we also consider a floor-bound automated guided vehicle system. The layout is fixed in this case. There is a network of bidirectional one-lane tracks connecting three manufacturing cells and a warehouse. The warehouse acts as a source for the logistic system. The AGVs start from the AGV pool, load items at the transfer point of the warehouse and transport the items to a transfer point of the manufacturing cells. After the processing by the manufacturing cells the items leave the system via the sinks. The AGVs can move forward, backwards and can stop. Because of the one-lane tracks, this driving behavior can lead to deadlocks. We consider two scenarios: one with three AGVs and one with four AGVs. Figure 3 shows the conceptual model of the logistic system.

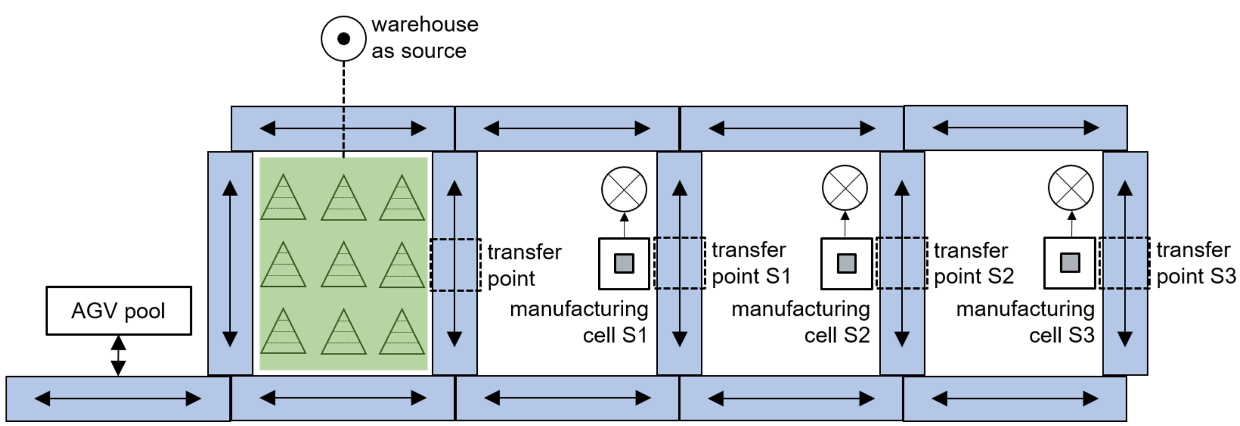

Figure 3: Conceptual model of the considered logistic system with manufacturing cells.

Since the considered scenario is dynamic and the moving AGVs can change possible shortest path to the destination of other AGVs, classical path planning algorithms such as Dijkstra (Dijkstra 1959) and A* (Hart et al. 1968) are not sufficient to guarantee deadlock-free driving behavior.

4 TECHNICAL IMPLEMENTATION

4.1 Simulation Model

The simulation model was created in Tecnomatix Plant Simulation 16.0. The written methods, which are located directly in Plant Simulation, use the integrated programming language 'SimTalk 2.0'. Figure 4 shows the material flow elements of the implemented simulation model in Tecnomatix Plant Simulation. In figure 4 the one-lane tracks seem to be unidirectional but the AGVs can use the tracks in two directions by driving forward and backward. The transfer points are implemented via sensors, which trigger the loading and unloading process.

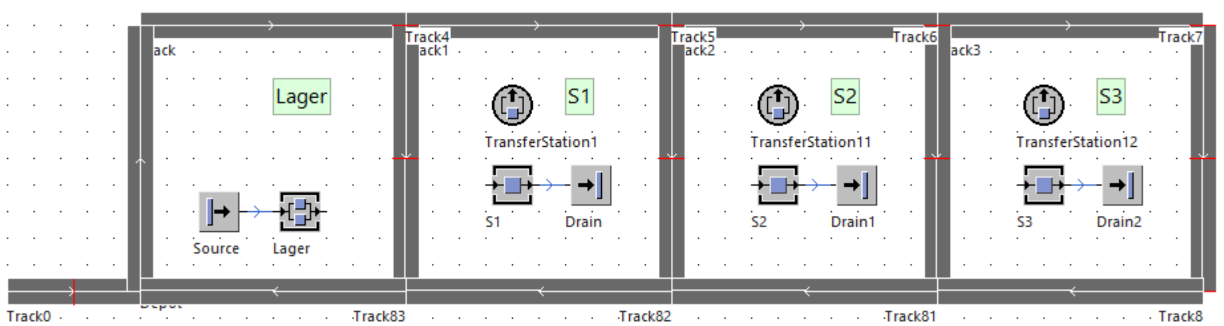

Figure 4: Material flow elements of the simulation model.

4.2 Deep Q-Network (DQN)

Our approach is based on Q-learning. We use a multilayer perceptron (MLP) as an artificial neural network to learn the decisions for the driving behavior of the AGVs. We refer to our approach as Deep Q-learning according to Mnih et al. (2015), although our artificial neural network with two hidden layers is not that deep. A target Q-network and an experience replay buffer are also used to stabilize and improve the learning curve of the agent. Table 1 shows the selected hyperparameters for both experiments.

Table 1: Selected hyperparameters for the DQN-algorithm.

Hyperparameter	3 AGVs	4 AGVs
replay buffer size	20,000	50,000
training starts after x steps	500	2,000
discount factor γ	0.95	0.95
ε_{min}	0.01	0.01
ε_{decay}	0.995	0.995
learning rate α	0.001	0.001
batch size	32	32

The state-space depends on the number of AGVs in the logistic system. Each vehicle needs six input neurons for the artificial neural network. The experiment with 3 AGVs needs 18 input neurons and with 4 AGVs there are 24 input neurons. The attributes sent to the artificial neural network for each vehicle are presented in table 2. The action space is represented by the output neurons and depend on the possible actions of each AGV. The actions that each AGV can perform are forward, backward and stop. Turning left or right is handled by methods in Plant Simulation and is therefore not part of the output layer. Since the agent in the system must make real-time decisions and use only one central agent, the number of output neurons is the combination of the three possible actions for each vehicle and the number of vehicles n. The action space has a size of 3^n, which is also the number of output neurons. The chosen action of the artificial neural network always includes an order for each vehicle.

Table 2: Input and outputs of the artificial neural network for each vehicle.

Input	Output
vehicle ID	forward
vehicle x-position	backward
vehicle y-position	stop
vehicle current speed	
vehicle destination x-position	
vehicle destination y-position	

In addition to the input layer and output layer of the artificial neural network, the network also consists of two hidden layers. The first hidden layer consists of 64 neurons and the second layer consists of 32 neurons. All neurons are connected to all neurons of the previous and subsequent layer.

The reward function focus on the transport of the goods. There is a high reward of 100 when a good reaches its goal. If an AGV hits an obstacle or another AGV, it will receive a penalty value of -25. To avoid unnecessary paths, AGVs receive a penalty value of -1 for each path segment they move on.

5 RESULTS

The episodes are limited to the number of agent actions. Each episode has 2,000 steps. One run-through of an episode lasted approximately 5.5 minutes. Two experiments with 1,000 episodes each were run as part of the investigation. The experiment lasted about five days in a configuration with an Intel Xeon Gold 5120. Figure 5 shows the cumulative rewards per episode for each experiment.

Initially, the number of collisions exceeded several thousand, and the AGVs were unable to navigate the path network. The first strategy chosen by the neural network was a standstill. At the same time, the number of collisions decreased significantly, but the AGVs were still unable to perform their assigned functions. After about 150 episodes, AGVs begin to navigate the path network and consider other AGVs. Collisions are much less frequent than initially, but there are still episodes where deadlocks occur. Such episodes are characterized by many collisions per episode.

In the second experiment with four AGVs, the amount of input and output data changed. This led to a change in the dimensions of the neural network. The number of neurons at the input changed to 24 and at the output to 81. With four AGVs, the number of collisions increased sharply. In the beginning, the AGVs could not move. After about 200 episodes, the transporters started to move and navigate in the path network. At the same time, the number of collisions per episode increased, but the transporters began to perform their main function - delivering cargo, which helped them in later episodes to achieve a positive reward for an episode and reduce the number of deadlocks. After 400 episodes, the number of collisions and deadlocks decreased rapidly. Deadlocks no longer occurred, there were only single collisions. In the end, the learning process slowed down considerably. This can have several reasons:

Hypothesis 1: Action selection is performed using the ε-greedy-Algorithm. The disadvantage of this algorithm is that it may not explore the space of options sufficiently and therefore does not estimate well enough which option is the best and gets stuck with a suboptimal option. Obviously, 1000 epochs are not enough to find the optimal strategy. It is necessary to perform training with a large number of epochs.

Hypothesis 2: The sigmoid activation function has an extremely small derivative at all points except for a small interval. This greatly complicates the process of weight improvement by gradient descent. In

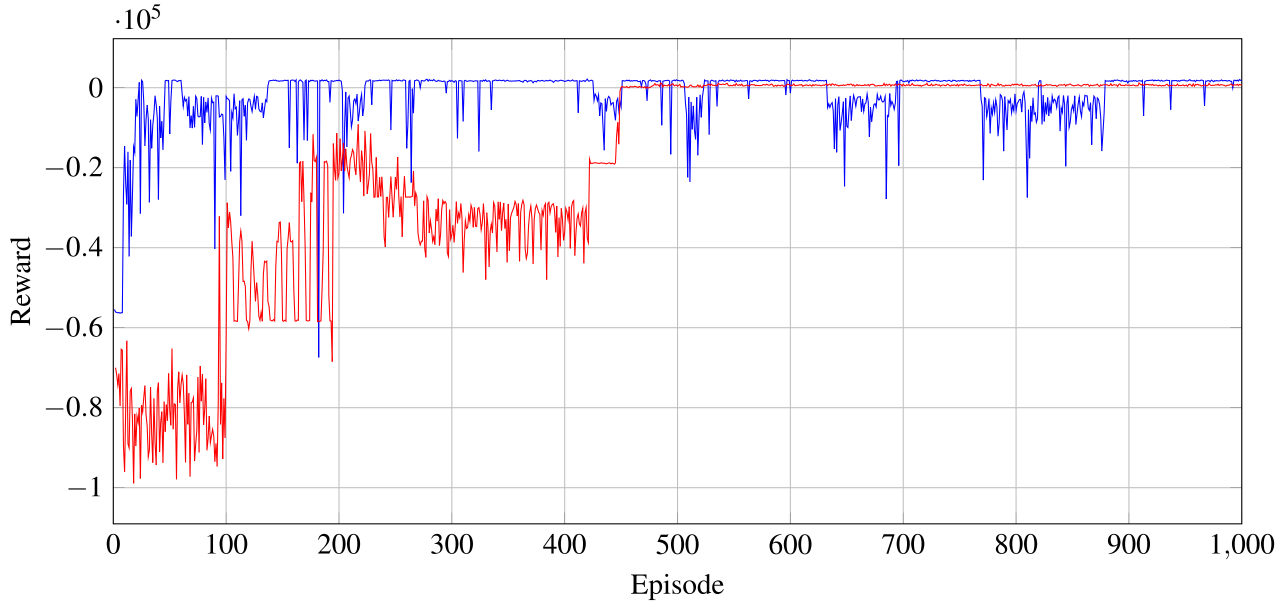

Figure 5: Cumulative rewards per episode for three (blue) and four (red) AGVs.

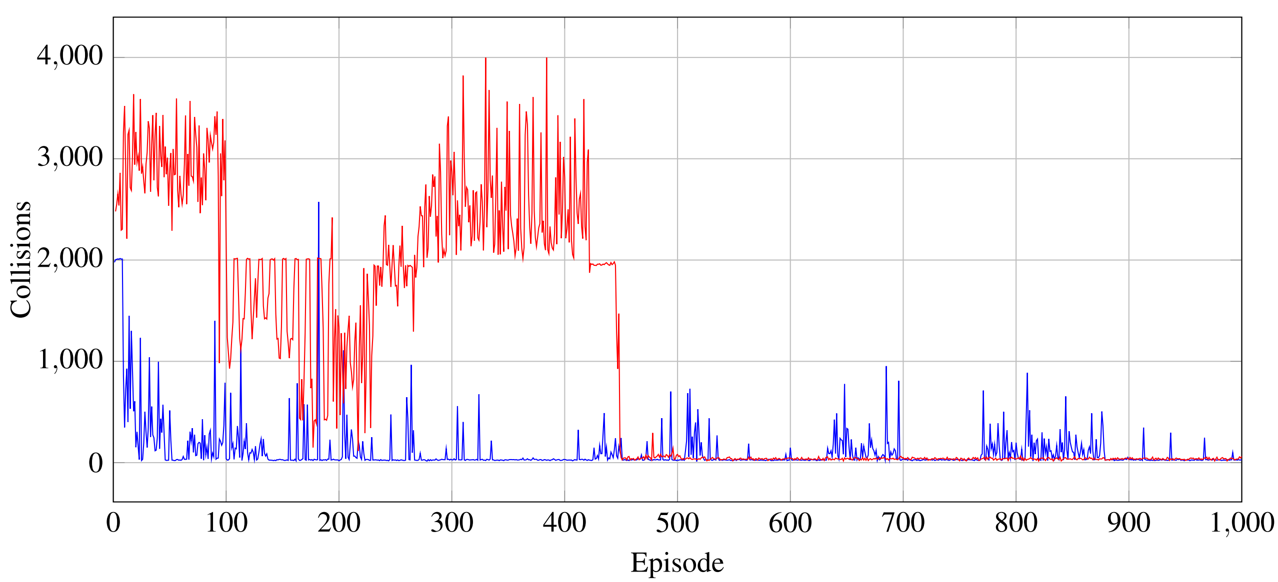

Figure 6: Collisions per episode for three (blue) and four (red) AGVs.

this case, the gradient value is small or disappears, because it cannot make any significant change due to its extremely small value. The neural network refuses to continue learning or does extremely slowly. There are several methods to solve this problem, for example, using a self-regulating neural network (SNN) (Klambauer et al. 2017).

Hypothesis 3: Since the Rectified Linear Unit (ReLU) part is a horizontal line for negative x values, the slope in this part is zero. Since the slope is zero, the weights are not adjusted during the descent. This means that neurons in this state do not respond to changes in error or input (simply because the gradient is zero, nothing changes). This phenomenon is called the dying ReLu problem (Lu 2020). Because of this problem, some neurons simply shut down and do not respond, making a significant portion of the neural network passive. There are variants of ReLu that can solve this problem, such as Leaky ReLu.

6 CONCLUSION AND OUTLOOK

Based on the obtained results, it can be concluded that machine learning, in particular deep learning with artificial neural networks, is able to make decisions independently, can be applied for production and logistics and can learn to handle logistic deadlocks. The amount of AGVs seems to have a high impact on the learning curve. Our approach in its form is very resource and time consuming. Therefore, it makes sense to modernize the approach to increase its efficiency. The transition to a multi-agent system will allow better and faster training of agents. This transition requires significant changes in the neural network architecture and the structure of the agent and learning environment.

Artificial neural networks have a significant advantage over classical rules for handling collisions and deadlocks. They can adapt autonomously to changes in the environment. Artificial neural networks that have been trained to operate in a specific environment can be easily relearned to operate under conditions of minor variations in environmental parameters. The higher the adaptability of the system, the more stable its work will be in a dynamic environment.

The application of machine learning and neural networks to deal with deadlocks in logistics appears to be an up-and-coming area of research and will be deepened with further investigations to enable a contribution to the creation of intelligent automated systems.

REFERENCES

Akalin, N., and A. Loutfi. 2020. "Reinforcement Learning Approaches in Social Robotics". *Sensors* 21(4):21–33.

Bashir, M., D. Liu, M. Uzam, N. Wu, A. Al-Ahmari, and Z. Li. 2018. "Optimal Enforcement of Liveness to Flexible Manufacturing Systems Modeled with Petri Nets via Transition-based Controllers". *Advances in Mechanical Engineering* 10(1):1–3.

Bouderba, S. I., and N. Moussa. 2019. "Reinforcement Learning (Q-LEARNING) Traffic Light Controller within Intersection Traffic System". In *Proceedings of the 4th International Conference on Big Data and Internet of Things*, edited by M. Lazaar, C. Duvallet, M. Al Achhab, O. Mahboub, and H. Silkan, 1–6. New York City, New York: Association for Computing Machinery.

Bouton, M., A. Nakhaei, K. Fujimura, and M. J. Kochenderfer. 2019. "Cooperation-Aware Reinforcement Learning for Merging in Dense Traffic". In *Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference*, 3441–3447. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Brown, T., B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei. 2020. "Language Models are Few-Shot Learners". In *Advances in Neural Information Processing Systems*, edited by H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, Volume 33, 1877–1901. Red Hook, New York: Curran Associates, Inc.

Canese, L., G. C. Cardarilli, L. Di Nunzio, R. Fazzolari, D. Giardino, M. Re, and S. Spanò. 2021. "Multi-Agent Reinforcement Learning: A Review of Challenges and Applications". *Applied Sciences* 11(11).

Capocchi, L., and J.-F. Santucci. 2022. "Discrete Event Modeling and Simulation for Reinforcement Learning System Design". *Information* 13(3):121.

Cireşan, D., U. Meier, and J. Schmidhuber. 2012. "Multi-Column Deep Neural Networks for Image Classification". In *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition*, 3642–3649. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Coffman, E. G., M. Elphick, and A. Shoshani. 1971. “System Deadlocks”. *ACM Computing Surveys* 3(2):67–78.

Dijkstra, E. W. 1959. “A Note on Two Problems in Connexion with Graphs”. *Numerische mathematik* 1(1):269–271.

Farhan, M., B. Göhre, and E. Junprung. 2020. “Reinforcement Learning in Anylogic Simulation Models: a Guiding Example using Pathmind”. In *Proceedings of the 2020 Winter Simulation Conference*, edited by K.-H. G. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, 3212–3223. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Forget, P., S. D'Amours, J.-M. Frayret, and J. Gaudreault. 2009. “Study of the Performance of Multi-behaviour Agents for Supply Chain Planning”. *Computers in Industry* 60(9):698–708.

Hart, P. E., N. J. Nilsson, and B. Raphael. 1968. “A Formal Basis for the Heuristic Determination of Minimum Cost Paths”. *IEEE transactions on Systems Science and Cybernetics* 4(2):100–107.

Havender, J. W. 1968. “Avoiding Deadlock in Multitasking Systems”. *IBM Systems Journal* 7(2):74–84.

Holt, R. C. 1972. “Some Deadlock Properties of Computer Systems”. *ACM Computing Surveys* 4(3):179–196.

Kim, C.-O., and S. S. Kim. 1997. “An Efficient Real-Time Deadlock-Free Control Algorithm for Automated Manufacturing Systems”. *International Journal of Production Research* 35(6):1545–1560.

Kim, K. H., S. M. Jeon, and K. R. Ryu. 2006. “Deadlock Prevention for Automated Guided Vehicles in Automated Container Terminals”. *OR Spectrum* 28(4):659–679.

Kim, T., M. Jang, and J. Kim. 2021. “A Survey on Simulation Environments for Reinforcement Learning”. In *Proceedings of 2021 18th International Conference on Ubiquitous Robots*, 63–67. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Klambauer, G., T. Unterthiner, A. Mayr, and S. Hochreiter. 2017. “Self-Normalizing Neural Networks”. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 972–981. Red Hook, New York: Curran Associates, Inc.

Kujirai, T., and T. Yokota. 2019. “Breaking Deadlocks in Multi-agent Reinforcement Learning with Sparse Interaction”. In *Proceedings of the 16th Pacific Rim International Conference on Artificial Intelligence*, edited by A. C. Nayak and A. Sharma, Volume 11670 of *Lecture Notes in Computer Science*, 746–759. Cham, Switzerland: Springer Nature Switzerland AG.

Lehmann, M. 2006. *Einsatzplanung von Fahrerlosen Transportsystemen in Seehafen-Containerterminals*. Dissertation, Technische Universität Berlin, Berlin, Germany.

Li, X., J. Zhang, J. Bian, Y. Tong, and T.-Y. Liu. 2019. “A Cooperative Multi-Agent Reinforcement Learning Framework for Resource Balancing in Complex Logistics Network”. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, 980–988. Richland, South Carolina: International Foundation for Autonomous Agents and Multiagent Systems.

Lienert, T., and J. Fottner. 2017. “No More Deadlocks – Applying the Time Window Routing Method to Shuttle Systems”. In *Proceedings of the 31st European Conference on Modelling and Simulation*, edited by Z. Z. Paprika, P. Horák, K. Váradi, P. T. Zwierczyk, Á. Vidovics-Dancs, and J. P. Rádics, 169–175. Budapest, Hungary: European Council for Modelling and Simulation.

Liu, G. 2016. “Complexity of the deadlock problem for Petri nets modeling resource allocation systems”. *Information Sciences* 363:190–197.

Liu, L., N. Gurney, K. McCullough, and V. Ustun. 2021. “Graph Neural Network Based Behavior Prediction to Support Multi-agent Reinforcement Learning in Military Training Simulations”. In *Proceedings of the 2021 Winter Simulation Conference*, edited by B. F. Kim, S. M. K. Smith, C. S. Z. Zheng, and M. Loper, 1–12. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Lu, L. 2020. “Dying ReLU and Initialization: Theory and Numerical Examples”. *Communications in Computational Physics* 28(5):1671–1706.

Lu, Z., Z. Zhuang, Z. Huang, and W. Qin. 2019. “A Framework of Multi-Agent Based Intelligent Production Logistics System”. *Procedia CIRP* 83:557–562.

Luo, J., Z. Liu, M. Zhou, K. Xing, X. Wang, X. Li, and H. Liu. 2019. “Robust Deadlock Control of Automated manufacturing systems with multiple unreliable resources”. *Information Sciences* 479:401–415.

Mayer, S., and K. Furmans. 2010. “Deadlock Prevention in a Completely Decentralized Controlled Materials Flow Systems”. *Logistics Research* 2(3-4):147–158.

Mnih, V., K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis. 2015. “Human-level Control through Deep Reinforcement Learning”. *Nature* 518(7540):529–533.

Mors, A. W. t. 2010. *The World According to MARP: Multi-Agent Route Planning*. Dissertation, Delft, Netherlands.

Müller, M., S. Schmidt, and T. Reggelin. 2019. “Deadlock and Collision Handling for Automated Rail-Based Storage and Retrieval Units”. In *Proceedings of the 2019 Winter Simulation Conference*, edited by N. Mustafee, K.-H. G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, 1591–1601. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Müller, M., J. H. Ulrich, L. S. Reyes-Rubiano, T. Reggelin, and S. Lang. 2020. "Comparison of Deadlock Handling Strategies for different Warehouse Layouts with an AGVS". In *Proceedings of the 2020 Winter Simulation Conference*, edited by K.-H. G. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, 1300–1311. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Palmer, G. I., P. R. Harper, and V. A. Knight. 2018. "Modelling deadlock in open restricted queueing networks". *European Journal of Operational Research* 266(2):609–621.

Petri, C. A. 1962. *Kommunikation mit Automaten*. Dissertation, Technische Hochschule Darmstadt, Bonn, Germany.

Rabe, M., and F. Dross. 2015. "A Reinforcement Learning Approach for a Decision Support System for Logistics Networks". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 2020–2032. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Reijnen, R., Y. Zhang, W. Nuijten, C. Senaras, and M. Goldak-Altgassen. 2020. "Combining Deep Reinforcement Learning with Search Heuristics for Solving Multi-Agent Path Finding in Segment-based Layouts". In *Proceedings of the 2020 IEEE Symposium Series on Computational Intelligence*, 2647–2654. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Schönung, F., S. Mayer, and D. Berbig. 2011. "Dezentrale, deadlockfreie Steuerung eines Stetigförderers". *at - Automatisierungstechnik* 59(4):73.

Seibold, Z. 2016. *Logical Time for Decentralized Control of Material Handling Systems*. Dissertation, Karlsruher Institut für Technologie, Karlsruhe, Germany.

Silver, D., A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis. 2016. "Mastering the Game of Go with Deep Neural Networks and Tree Search". *Nature* 529(7587):484–489.

Silver, D., J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis. 2017. "Mastering the Game of Go without Human Knowledge". *Nature* 550(7676):354–359.

Sørensen, R. A., M. Nielsen, and H. Karstoft. 2020. "Routing in Congested Baggage Handling Systems using Deep Reinforcement Learning". *Integrated Computer-Aided Engineering* 27(2):139–152.

Sun, S., C. Gu, Q. Wan, H. Huang, and X. Jia. 2018. "CROTPN Based Collision-Free and Deadlock-Free Path Planning of AGVs in Logistic Center". In *Proceedings of the 15th International Conference on Control, Automation, Robotics and Vision*, 1685–1691. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Tanenbaum, A. S. 2015. *Modern Operating Systems*. Fourth edition ed. Upper Saddle River, New Jersey: Pearson Education, Inc.

Vinyals, O., I. Babuschkin, W. M. Czarnecki, M. Mathieu, A. Dudzik, J. Chung, D. H. Choi, R. Powell, T. Ewalds, P. Georgiev, J. Oh, D. Horgan, M. Kroiss, I. Danihelka, A. Huang, L. Sifre, T. Cai, J. P. Agapiou, M. Jaderberg, A. S. Vezhnevets, R. Leblond, T. Pohlen, V. Dalibard, D. Budden, Y. Sulsky, J. Molloy, T. L. Paine, C. Gulcehre, Z. Wang, T. Pfaff, Y. Wu, R. Ring, D. Yogatama, D. Wünsch, K. McKinney, O. Smith, T. Schaul, T. Lillicrap, K. Kavukcuoglu, D. Hassabis, C. Apps, and D. Silver. 2019. "Grandmaster Level in StarCraft II using Multi-Agent Reinforcement Learning". *Nature* 575(7782):350–354.

Wang, X., L. Zhang, Y. Laili, K. Xie, H. Lu, and C. Zhao. 2021. "MADES: A Unified Framework for Integrating Agent-Based Simulation with Multi-Agent Reinforcement Learning". In *Proceedings of the 2021 Annual Modeling and Simulation Conference*, 1–12. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Waschneck, B., A. Reichstaller, L. Belzner, T. Altenmuller, T. Bauernhansl, A. Knapp, and A. Kyek. 2018. "Deep Reinforcement Learning for Semiconductor Production Scheduling". In *Proceedings of the 29th Annual SEMI Advanced Semiconductor Manufacturing Conference*, 301–306. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Yalcin, A. 2017. *Multi-Agent Route Planning in Grid-Based Storage Systems*. Dissertation, Europa-Universität Viadrina Frankfurt, Frankfurt am Main, Germany.

Yu, L., C. Zhang, J. Jiang, H. Yang, and H. Shang. 2021. "Reinforcement Learning Approach for Resource Allocation in Humanitarian Logistics". *Expert Systems with Applications* 173:1–14.

Zhang, M., G. Tian, Y. Zhang, and P. Duan. 2021. "Service Skill Improvement for Home Robots: Autonomous Generation of action Sequence Based on Reinforcement Learning". *Knowledge-Based Systems* 212:106605.

Zhao, Z., Z. Wang, K. Han, P. Tiwari, G. Wu, and M. Barth. 2022. "Personalized Car Following for Autonomous Driving with Inverse Reinforcement Learning". In *Proceedings of the 2022 IEEE International Conference on Robotics and Automation*. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Zheng, B., C.-W. Lin, S. Shiraishi, and Q. Zhu. 2020. "Design and Analysis of Delay-Tolerant Intelligent Intersection Management". *ACM Transactions on Cyber-Physical Systems* 4(1):1–27.

Zhou, T., D. Tang, H. Zhu, and L. Wang. 2021. "Reinforcement Learning With Composite Rewards for Production Scheduling in a Smart Factory". In *IEEE Access*, Volume 9, 752–766. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

AUTHOR BIOGRAPHIES

MARCEL MÜLLER is a research fellow at the Otto von Guericke University Magdeburg. He earned his master degree in Industrial Engineering for Logistics at the Otto von Guericke University Magdeburg. His research interests include modeling and simulation of logistics systems and handling of deadlocks. His email address is marcel1.mueller@ovgu.de. His website is https://www.ilm.ovgu.de/mueller.

LORENA S. REYES-RUBIANO is a full-time lecturer at the Universidad de La Sabana, Chia-Colombia. She works also as a researcher at the Otto von Guericke University Magdeburg in Germany. She has a Ph.D. degree in mathematics and statistics from the Public University of Navarre (2019). Her research interests are urban logistics, humanitarian logistics, and multi-objective algorithms. Her email address is lorena.reyes1@unisabana.edu.co.

IEGOR KUTSENKO is an IT Business Analyst at Deutscher Auto Dienst GmbH. He earned his master degrees in Applied Mechanics at National Aerospace University "Kharkiv Aviation Institute" in Ukraine and in Industrial Engineering for Logistics at Otto von Guericke University Magdeburg in Germany. His research interests include automation, digital twins and artificial intelligence. His email address is egor.kutsenko@gmail.com.

TOBIAS REGGELIN is a project manager, researcher and lecturer at the Otto von Guericke University Magdeburg and the Fraunhofer Institute for Factory Operation and Automation IFF Magdeburg. His main research and work interests include modeling and simulation of production and logistics systems and developing and applying logistics management games. Tobias Reggelin received a doctoral degree in engineering from the Otto von Guericke University Magdeburg. Furthermore, he holds a master's degree in Engineering Management from Rose-Hulman Institute of Technology in Terre Haute, IN and a diploma degree in Industrial Engineering in Logistics from the Otto von Guericke University Magdeburg. His email address is tobias.reggelin@ovgu.de.

HARTMUT ZADEK received the call to Otto von Guericke University Magdeburg and has been head of the Chair of Logistics in the Institute of Logistics and Material Flow Engineering in the Faculty of Mechanical Engineering since January 1, 2008. He is also a lecturer at the University of Dortmund and a lecturer at the Hamburg School of Logistics. He is a member of the Association of German Industrial Engineers and the German Logistics Association (BVL). He has headed the Sustainable Production Logistics working group for BVL since June 2009. He is the editor of various specialist books, author of numerous specialist publications, sought-after speaker at specialist conferences in Germany and abroad, moderator of specialist congresses and seminars and leader of strategy workshops. In January 2010 he was appointed to the Saxony-Anhalt Logistics Advisory Board by the Minister for Regional Development and Transport. His email address is zadek@ovgu.de.