Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 00 (2019) 000–000

www.elsevier.com/locate/procedia

CIRPe 2020 – 8th CIRP Global Web Conference – Flexible Mass Customisation

# Decentralized Multi-Agent Production Control through Economic Model Bidding for Matrix Production Systems

Marvin Carl May[a,*], Lars Kiefer[a], Andreas Kuhnle[a], Nicole Stricker[a], Gisela Lanza[a]

[a]*wbk Institute of Production Science, Karlsruhe Institute of Technology (KIT), Kaiserstr. 12, 76131 Karlsruhe, Germany*

**Abstract**

Due to increasing demand for unique products, large variety in product portfolios and the associated rise in individualization, the efficient use of resources in traditional line production dwindles. One answer to these new challenges is the application of matrix-shaped layouts with multiple production cells, called Matrix Production Systems. The cycle time independence and redundancy of production cell capabilities within a Matrix Production System enable individual production paths per job for Flexible Mass Customisation. However, the increased degrees of freedom strengthen the need for reliable production control systems compared to traditional production systems such as line production. Beyond reliability a need for intelligent production within a smart factory in order to ensure goal-oriented production control under ever-changing manufacturing conditions can be ascertained. Learning-based methods can leverage condition-based reactions for goal-oriented production control.
While centralized control performs well in single-objective situations, it is hard to achieve contradictory targets for individual products or resources. Hence, in order to master these challenges, a production control concept based on a decentralized multi-agent bidding system is presented. In this price-based model, individual production agents – jobs, production cells and transport system – interact based on an economic model and attempt to maximize monetary revenues. Evaluating the application of learning and priority-based control policies shows that decentralized multi-agent production control can outperform traditional approaches for certain control objectives. The introduction of decentralized multi-agent reinforcement learning systems offers a starting point for further research in this area of intelligent production control within smart manufacturing.



*Keywords:* Smart Factory; Industry 4.0; Matrix Production; Production Planning and Control; Mass Customisation

## 1. Introduction

Due to an ever-increasing demand for individualized mass produced products, a trend towards flexible mass customization emerges [5] which requires a flexible and adaptive production system. The automotive industry, in particular, is frequently referred to as a main driver for this development, owing to the line takt time being determined by the station requiring the longest time to perform its operation [7]. However, to achieve high degrees of flexibility and adaptivity, a takt time independent production system is necessary. A Matrix Production System fulfills these requirements by enabling individual material flow. However, the more complex material flow and station redundancy necessitates more complex production control [6, 19, 25]. This triggers a trend towards intelligent adaptive Production Planning and Control. Due to the scalability and the autonomy of individual components, such as the transport unit, decentralized control is particularly suitable for matrix production Greschke [12]. To solve the challenges described above, this paper introduces an economic bidding model for production control combining reinforcement learning and decentralized control.
The paper is structured as follows. The introduction of Matrix Production Systems and relevant production control in Section 2 is followed by the presentation of the economic model bidding for production control in Section 3. An implementation and case study results are presented in Section 4. This paper concludes with a discussion and outlook in Sections 5 and **??**.

* Corresponding author. Tel.: +49-1523-950-2624 ; fax: +49-721-608-45005.

*E-mail address:* marvin.may@kit.edu (Marvin Carl May).

## 2. State-of-the-Art

This section discusses Matrix Production Systems (MPS) and suitable Production Planning and Control (PPC). A literature review on Multi-Agent-Systems (MAS) concludes this state-of-the-science discussion.

### 2.1. Matrix Production System

In line production, an enabler for Mass Customisation, the takt time restricts the throughput of a maximum production system. Hence, whenever a station requires less than the takt time, precious potentially value-adding time is lost. Thus, breaking down the rigid line structure into a matrix-shaped layout of individual working cells, also denoted stations, serves as an MPS basis. Due to the sub-autonomous cells and the flexible transport system, the control logic is responsible for supplying each work cell with sufficient material to achieve a highly utilized and fast measurement flow production [10]. Furthermore, the individual product becomes less important in matrix production and worse transport routes can be accepted as long as the production capacity utilization does not suffer [28]. The basis for takt time independence is a flexible transport system, which enables high route flexibility in material flow. One way to enable a flexible routing strategy is to use Automated Guided Vehicles (AGV). AGVs themselves are decentralized and autonomous vehicles that are not tied to a fixed route network and can, therefore, move freely. However, this presents new challenges for the control system, since the elimination of rigid logistics chains increases complexity. Additional static and dynamic influences must be considered, e.g. the increasing risk of deadlocks, [2].

Uncoupled modular stations enable the MPSs flexibility [8], which in this case is twofold: (1) routing flexibility due to the layout and (2) product flexibility due to modular and partially redundant stations. Popp [23] adds versatility and adaptability to conceivable advantages:

- *Flexibility* is the ability of a system to respond to changes in input and output variables and the conditions of production within defined limits without loss of stability and effectiveness [22].
- *Adaptability* describes the ability to change a system configuration across the flexibility limits with as little time and capital expenditure as possible [2].
- *Versatility* describes the ability to react to the unpredictable and, thus, supplements the concept of flexibility towards unpredictable influences [2].

### 2.2. Matrix suitable Production Planning and Control

The complexity of control and planning production depends mainly on the production system, in particular comparing MPS an line-production [12]. In line-production, the control effort is quite low, due to the lack of material flow flexibility, yet reacting to disturbances is hard. In MPS reacting to malfunctions is possible, as there is no fixed relationship between stations and each part, since the latter can have differing routes [17]. Hence, production control plays a much more crucial role in MPS than in line-production [6, 25]. Furthermore, the described control task is an NP-hard problem, so that no optimal result can be found in polynomial time [35].

Production control concepts are separated into decentralized and centralized ones. According to Greschke [12], decentralized control is more suitable for MPS as AGVs are designed to act autonomously. Each agent resembling an AGV decides based on the current MPS state and surrounding conditions, focused on which station to approached next [12]. The major pros and cons are discussed in Table 1.

Table 1. Production Control in MPS Pros and Cons

| Advantages | Disadvantages |
|---|---|
| Control task distribution leaves more but less complex sub-problems | Individual solution may not resemble global solution |
| Agent redundancy increases stability against the unexpected and control system failures | High risk of deadlocks and coordination problems |
| Learning control policies can easily adapt to structural changes | Learning and learning speed is hindered through coordination |

In the literature mainly two decentralized production control approaches for MPS exist: (1) priority rules [11, 17] and (2) hybrid solutions linking machine learning with priority rules [6]. The former are known for quick execution, requiring iterative development and needing fine tuning for each particular problem and producing ”acceptable but not necessarily optimal solutions” [21]. Hence, the lacking scalability is an additional reason to favor learning PPC, such as Reinforcement Learning [30].

#### 2.2.1. Reinforcement Learning

Reinforcement Learning (RL) provides an alternative to priority rules because of its behavior-based learning. Furthermore, the use of neural networks can reduce the initialization effort and decrease the problem complexity [27]. A RL model consists of one agent placed in an environment, where the agent perceives the environments state $s_t \in S$ and selects an action $a_t \in A$ to manipulate the next state $s_{t+1}$ [31] as shown in Figure 1. The agents goal is to learn to optimize a reward signal $r_t$, which can guide the agent to a meaningful control policy. The underlying concept is called the Markov Decision Process (MDP) with $MDP = (S, A, P, R)$ and includes the Markov property insofar as future states solely depend on the previous state and selected action: $P[S\,t + 1|S_t] = P[S_{t+1}|S_1 ... S_t]$.


Fig. 1. Reinforcement Learning cycle

Table 3. Comparison of relevant Multi-Agent System PPC literature that include learning components

| | Production matter | | Strategy task | | | | | | | Task | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Approach by** | MPS | Shop Floor | ML | RL | Global Reward | Priority rules | Bidding | Competitive | Cooperative | Routing | Scheduling | **Main topic** |
| Dittrich and Fohlmeister [4] | ○ | ◐ | ● | ● | ● | ○ | ○ | ○ | ● | ○ | ● | Scheduling with RL |
| Giordani et al. [9] | ◐ | ● | ○ | ○ | ○ | ○ | ● | ● | ○ | ● | ● | Two step dispatching of mobile robots |
| Heger et al. [14] | ○ | ◐ | ● | ○ | ○ | ● | ○ | ○ | ○ | ○ | ● | Switching priority rules at machines |
| Kaihara [15] | ○ | ○ | ○ | ○ | ○ | ○ | ● | ● | ○ | ○ | ○ | Supply Chain Management bidding |
| Malus et al. [18] | ◐ | ● | ● | ● | ○ | ○ | ● | ○ | ◐ | ● | ○ | Bidding RL control for Routing |
| Scholz-Reiter et al. [26] | ● | ● | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ● | ○ | Comparing different autonomy levels |
| Scholz-Reiter and Hamann [27] | ◐ | ● | ● | ◐ | ○ | ○ | ○ | ○ | ◐ | ● | ○ | Reducing inventory through ML |
| Tampuu et al. [32] | ○ | ○ | ● | ● | ○ | ○ | ● | ● | ● | ○ | ○ | Competition vs. cooperation in games |

Legend : ● regarded ◐ rudimentary regarded ○ not regarded

### 2.2.2. *Multi-Agent System PPC*

Due to the decentralized structure of MPS, a simulation based on a Multi-Agent System (MAS) is implemented. In MAS several agents interact with one and the same environment and perform similar or different tasks [31]. In contrast to single agent control tasks, each agent in a MAS has to consider the impact of its own decision and that of it's peers to adapt its behavior and, thus, may not necessarily have equal access to complete information impeding finding an optimal behavior [31]. Decomposing the control problem into smaller sub-problems, however, creates flexibility and robustness in the number of agents and faster calculations for each policy. The famous RL "credit assignment problem" is expanded, as besides timing the corresponding responsible agent is hard to determine [33]. Hence, it is difficult to distribute reward across agents [33], in particular in high dynamic environments [31].

The application of MAS to production relevant control problems has long fascinated researchers, in particular in the domain of Supply Chain Management [3, 15]. Very early production control MAS research uses real prices to coordinate multiple agents, such as Gu et al. [13], and only slowly start to include Machine Learning (ML) [27], yet their results, given the algorithms at that time, look promising. A further plentitude of papers focuses on MAS frameworks built on top of priority rules for machine control [14], dispatching [9] or self/organization [34]. The latter describes a detailed, but still hierarchical, MAS negotiation mechanism. Wang et al. [34] distinguish among agents for products, machines, transportation units, supplementary purposes and conflict resolving coordinator. However, the approach lacks the ability to integrate learning agents and relies on a central coordinator.

Applying the categorization of agents developed by Monostori et al. [20], apparently most MAS include purposeful and interacting behavior as well as autonomy. So far only the most recent publications by Dittrich and Fohlmeister [4], Malus et al. [18] include intelligence and learning behavior. Yet, their approach lack scalability, as reward signals are computed centrally or distributed to a hierarchical supervisory agent, while the highest level of autonomy [26] has not been realized. Only autonomous agents, without centralized control or rewards, show the high level of scalability required to control a production system with many different numbers of learning agents. Hence, this paper addresses Production Control through an Economic Model for MAS supported by bidding processes.

## 3. Economic Bidding Model for PPC

In order to enable high scalability, a "flat hierarchic coalition" [1], where several agents work in a coalition for short periods to increase their performance, is implemented. The economic model is a price-based multi-agent bidding system where all communication and decision making is based on the price offered. Within the system each potential decision making instance is characterized by an agent: (1) part agents which are akin to partial jobs or orders flowing through the production system, (2) station agents controlling the selection of parts at each station or machine and (3) AGV agents that respresent AGVs transporting parts within the system. The underlying economic model concept is that all parts, stations and AGVs want to generate maximum profit independently of one another. In order to do so, the part serves as a bidder and the AGV or station act as an auctioneer auctioning their capacities. For each process and transport to be completed, the part must make a bid with the goal of spending as little money as possible. The station and AGV, on the other hand, want to earn as possible, so they compare all parts' bids and select the most suitable, as shown in Figure 2. The bidding and evaluation of the offers should be done in a way that optimizes the Utilization Efficiency (UE) as introduced by Kang et al. [16]. This model is intended to reduce the many parameters and decision variables to the two variables price and process or transport id and thus enable the decision making process to scale easily. The earliest predecessor Kaihara [15] showed that an economic bidding model can lead to efficient resource allocation.

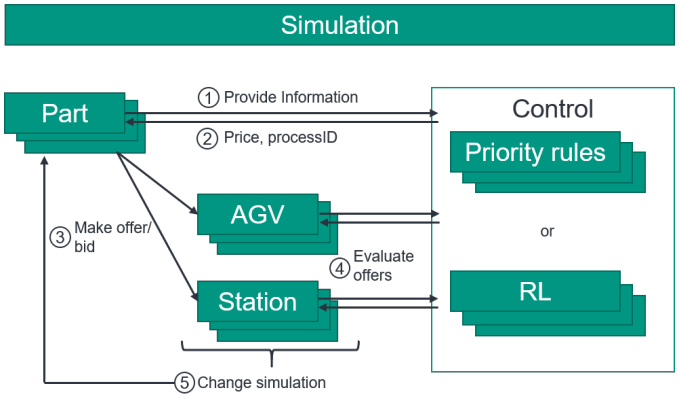


Fig. 2. Structure of the economic model

### 3.1. Part Agent

Part Agents select the next process to be performed or the next station to be approached and, thus, select the path of the individual order part they are tied to. Their goal is to finalize their part into a final product by sticking to technical and temporal requirements. Hence, they receive a reward for their individual performance, which they optimize. Depending on the current condition of the part and the system, the part sends a quotation that includes price and desired process. The algorithm determining the quotation depends on the logic, i.e. RL, chosen. As soon as a part is available for transport, it submits its quotation to a central market place. If a part arrives in a station buffer, it waits until the station has finished processing and starts a new bidding round. Both quotations are of similar shape and simultaneous quotations are possible.

### 3.2. Station Agent

As described in the previous section, a station requests a quote from all parts in its respective buffer as soon it can start a new process. Once the quoting process is complete, the station uses an algorithm to select the best quote and start processing. Pricing can be partially open or closed, depending on whether they receive information about previous quotations.

### 3.3. AGV Agent

In contrast to the station, AGV agents do not perform individual bidding rounds but select among the parts' quotations on a central market. Due to the time offset of individual orders becoming amenable to transport and a centralized market approach increases efficiency. Among these quotations the AGV agent selects the most suitable one, for loading, unloading and transportation operations. This enables seamless integration of AGVs that transport multiple unique parts on unique routes.

### 3.4. Learning and priority rule integration

The Economic Model is based on reducing the wide range of decision possibilities of decentralized production control to few but sufficient variables. Most crucial in the above presented coordination mechanism is the selection of individual agents' algorithms. One advantage is that this framework can mimic traditional heuristics behavior, implement different priority rules but also include learning. Priority rules, e.g. Longest Waiting Time (LWT) or Earliest Due Date (EDD), are known to perform sufficiently well in production control problems, in particular the ones presented in Table 3.4. Their dispatching application can be disaggregated into Scheduling, selecting the order of part processing at stations, and Routing, selecting the origin, destination and respective part to be transported next. The priority is assigned, according to the ranking defined in Table 3.4.

Table 5. Overview of relevant implemented priority rules

| Priority Rule | Scheduling | Routing |
|---|---|---|
| First-In-First-Out (FIFO) | LWT | LWT |
| Shortest Distance (SD) | LWT | $\frac{waitingtime}{distance}$ |
| Shortest Distance with fill level (SDF) | LWT | $\frac{waitingtime}{distance} \cdot filllevel$ |
| SDF and random ($r$) routing (SDFR) | LWT | $\frac{waitingtime}{distance} \cdot filllevel + r$ |

Moreover, RL agents and priority rules can be used interchangeably as they act based on information about the current production circumstances and submit quotations. The former achieves guided learning and remains at the core of this study. The goal of each agent is to maximize his money, which in the case of the RL agent leads to a reward maximization. One novel policy-based agent which is known for its robustness and quick convergence is Proximal Policy Optimization (PPO) introduced by Schulman et al. [29].

The direct link between individual quotes, i.e. actions, and state information is known for the following known advantages: (1) Trajectories remain stable, (2) fulfilling the Markov property and (3) high quotes increase individual production speed per part, as it is known, that a well-designed economic model can lead to an increase in the overall efficiency of production [33].

### 3.5. Reinforcement Learning state vector definition

For RL it is crucial to develop a comprehensive, yet economical, state representation [31]. From each agents perspective, the following elements are important to consider:

- **Market information**, containing others and/or previous quotes for *parts* and all current quotes for *stations & AGVs*
- **Individual agent information**: in particular necessary changeover times or failure information for *stations*, due dates and potential next processes for *parts* and positional data for *AGVs*
- **General production information**, allowing agents to perceive the current production system circumstances

While priority rules can easily access the above information and do solely regard parts, the RL state needs to be well defined and remain structurally unchanged. Concrete RL state implementation is realized through information aggregation in tabular form

which is handled by a RL policy representing Convolutional Neural Network (CNN), as provided by Schaarschmidt et al. [24]. Doing so combines PPO advantages on continual control problems [29] with the ability of CNNs to generalize and extract features [31]. The above presented method is computationally heavy, if every part, station and AGV is represented as a unique policy approximating network. In a similar vein to Dittrich and Fohlmeister [4], all entities can be represented by one policy keeping track of different trajectories. Thus, the suggested method is computationally manageable and advantageous insofar as the one policy observes many more samples compared to many different agents.

### 3.6. Reward design

The reward signal shall guide RL agents to desired control objectives [31] and, thus, is of particular importance. While the inclusion of global rewards can lead to successful production control [4, 30], the credit assignment problem remains. Based on detailed reward engineering for part agents, most promising results were obtained when combining a global reward consisting of UE after part completion with a local component depending on the monetary sum of accepted quotes. Continuous reward is based on non-value adding time and the number of consecutive unsuccessful quotations. AGV and station agents obtain a reward signal depending on the chosen quotation, value-adding time since the last decision and/or distance traveled.

To accelerate the learning process action masking, the selectable action vector restriction excluding impossible actions [24], is implemented.

## 4. Results

In the following, the different heuristics and RL approaches are compared using two MPS configurations according to 7, each featuring the production of 2000 parts. Being a MPS no balancing is applied, but the ability to perform certain operations as well as the layout and distance is similar in both scenarios. The performance is evaluated against the Utilization Efficiency (UE) and Actual Order Excution Time (AOET) introduced by Kang et al. [16].

| No. | System configuration |
|---|---|
| **1** | 15 AGVs (buffer size 5), 10 Stations (buffer size 5, standard) |
| **2** | 15 AGVs (buffer size 1), 10 Stations (buffer size 5, standard) |

Table 7. MPS configurations used

Figure 3 shows the results for a multitude of compared PPC algorithms. The scenario *Part RL* uses an RL only for the part, while station and AGV act based on FIFO. In *RL Station* only stations are controlled by RL, whereas *RL Part with SDFR* replaces the AGV and station heuristic through SDFR. Additionally, a *Random* action is selection is included.

A comparison of the results shows that in addition to FIFO, SDFR delivers good results and SDFR outperforms FIFO due to the enhanced routing strategy ensuring improved part distribution across stations. Out of all control methods that include learning, the *RL Station* approach is most promising, as its performance is on par or above the heuristics in both scenarios. In particular compared to FIFO it increases production speed by more than 10% in configuration 2, owing to a constant and high buffer fill level in the regarded MPS. Furthermore, the performance is stable across both configurations, in contrast to heuristics such as FIFO. While SDFR performs slightly better in both scenarios, it lacks the ability to learn and cannot be improved in consecutive studies.

Furthermore, *RL Part* control, with or without SDFR, performs significantly better in reducing AOET, as each part has the incentive to quickly finish production as shown in Figure 3, yet, utilization is far lower. Ultimately, the front of pareto solutions is expanded through the inclusion of RL. Varying the state and reward signal, even by small margins, can influence the production control performance. Thus, the above presented framework can be used to derive better and more specialized MPS PPC, based on learning agents.

## 5. Discussion & Outlook

This economic bidding model for MAS production control is characterized by its scalability and adaptability. Thus, the control concept, which communicates mainly via two variables, price and process ID, can be used for a wide variety of MPS configurations. Furthermore, new heuristics and RL strategies can be easily implemented and optimized. The results show that the routing of parts plays a decisive role in overall performance. For selecting the next station a combination of distance and fill level (SDFR) has proven to be suitable. Also, it could be shown that the use of RL agents controlling stations can outperform many heuristics due to a superior local scheduling strategy and that a combination of SDFR and Part RL leads to a significant reduction in part production time. However, no optimal solution could be found when using solely RL agents, due to the many short-sighted decisions.

Thus, a combination of priority rules and learning components is favorable and future research shall focus on improving such control policies. Furthermore, in order to reduce the problem of short sighted decisions and based on observations for MAS control in games [32], the model could be adapted in such a way that instead of the strong competition, cooperation between the components is facilitated.

## Acknowledgements

This research work was undertaken in the context of DIGIMAN4.0 project (“DIGItal MANufacturing Technologies for Zero-defect Industry 4.0 Production”, http://www.digiman4-0.mek.dtu.dk/). DIGIMAN4.0 is a European Training Network supported by Horizon 2020, the EU Framework Programme for Research and Innovation (Project ID: 814225).

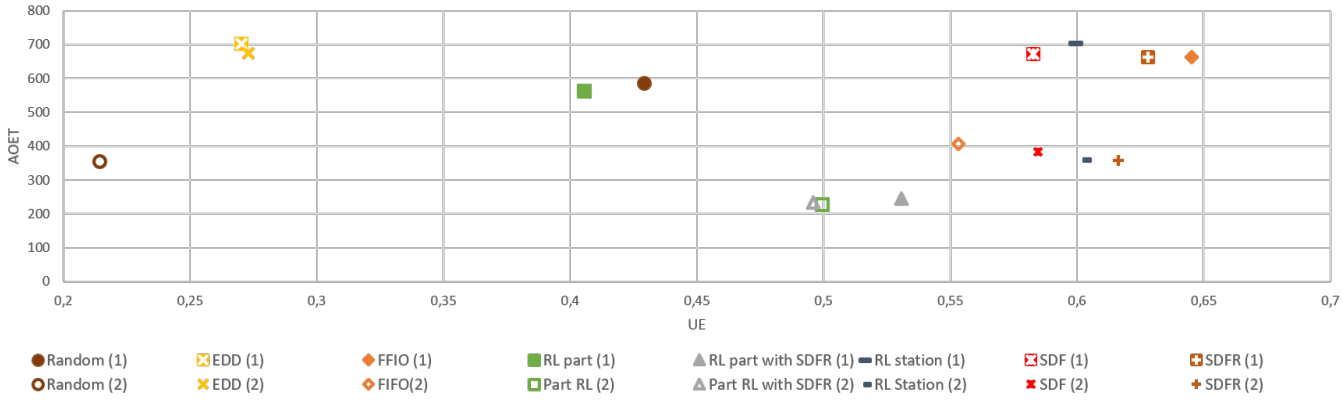


Fig. 3. Comparison of RL and Heuristic for both MPS configurations

## References


[1] Balaji, P., Srinivasan, D., 2010. An introduction to multi-agent systems, in: Innovations in multi-agent systems and applications-1. Springer, pp. 1–27.

[2] Bochmann, L.S., 2018. Entwicklung und Bewertung eines flexiblen und dezentral gesteuerten Fertigungssystems für variantenreiche Produkte. Ph.D. thesis. ETH Zurich.

[3] Davidsson, P., Wernstedt, F., 2002. A multi-agent system architecture for coordination of just-in-time production and distribution. The Knowledge Engineering Review 17, 317–329.

[4] Dittrich, M.A., Fohlmeister, S., 2020. Cooperative multi-agent system for production control using reinforcement learning. CIRP Annals .

[5] Duray, R., Ward, P.T., Milligan, G.W., Berry, W.L., 2000. Approaches to mass customization: configurations and empirical validation. Journal of operations management 18, 605–625.

[6] Echsler Minguillon, F., Lanza, G., 2017. Maschinelles lernen in der pps. Wt Werkstatttechnik Online 107, 630–634.

[7] ElMaraghy, H., Schuh, G., ElMaraghy, W., Piller, F., Schönsleben, P., Tseng, M., Bernard, A., 2013. Product variety management. Cirp Annals 62, 629–652.

[8] Filz, M.A., Gerberding, J., Herrmann, C., Thiede, S., 2019. Analyzing different material supply strategies in matrix-structured manufacturing systems. Procedia CIRP 81, 1004–1009.

[9] Giordani, S., Lujak, M., Martinelli, F., 2013. A distributed multi-agent production planning and scheduling framework for mobile robots. Computers & Industrial Engineering 64, 19–30.

[10] Greschke, P., Herrmann, C., 2014. Das humanpotenzial einer taktunabhängigen montage. ZWF Zeitschrift für wirtschaftlichen Fabrikbetrieb 109, 687–690.

[11] Greschke, P., Schönemann, M., Thiede, S., Herrmann, C., 2014. Matrix structures for high volumes and flexibility in production systems. Procedia CIRP 17, 160–165.

[12] Greschke, P.I., 2016. Matrix-Produktion als Konzept einer taktunabhängigen Fließfertigung. Vulkan Verlag.

[13] Gu, P., Balasubramanian, S., Norrie, D., 1997. Bidding-based process planning and scheduling in a multi-agent system. Computers & Industrial Engineering 32, 477–496.

[14] Heger, J., Branke, J., Hildebrandt, T., Scholz-Reiter, B., 2016. Dynamic adjustment of dispatching rule parameters in flow shops with sequence-dependent set-up times. International Journal of Production Research 54, 6812–6824.

[15] Kaihara, T., 2003. Multi-agent based supply chain modelling with dynamic environment. International Journal of Production Economics 85, 263–269.

[16] Kang, N., Zhao, C., Li, J., Horst, J.A., 2016. A hierarchical structure of key performance indicators for operation management and continuous improvement in production systems. International Journal of Production Research 54, 6333–6350.

[17] Kern, W., Rusitschka, F., Kopytynski, W., Keckl, S., Bauernhansl, T., 2015. Alternatives to assembly line production in the automotive industry, in: Proceedings of the 23rd International Conference on Production Research (IFPR).

[18] Malus, A., Kozjek, D., et al., 2020. Real-time order dispatching for a fleet of autonomous mobile robots using multi-agent reinforcement learning. CIRP Annals .

[19] May, M.C., Kuhnle, A., Lanza, G., 2020. Digitale produktion und intelligente steuerung. Wt Werkstatttechnik Online 110, 655–660. doi:https://doi.org/10.5445/IR/1000119555.

[20] Monostori, L., Váncza, J., Kumara, S.R., 2006. Agent-based systems for manufacturing. CIRP annals 55, 697–720.

[21] Naso, D., Turchiano, B., 2005. Multicriteria meta-heuristics for agv dispatching control based on computational intelligence. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 35, 208–226.

[22] Pleschak, F., 1988. Flexible Automatisierung: wirtschaftliche Gestaltung und Einsatzvorbereitung. Verlag Industrielle Organisation.

[23] Popp, J., 2018. Neuartige logistikkonzepte für eine flexible automobilproduktion ohne band .

[24] Schaarschmidt, M., Kuhnle, A., Fricke, K., 2017. Tensorforce: A tensorflow library for applied reinforcement learning. Web page .

[25] Schmitt, R., Göppert, A., Hüttemann, G., Lettmann, P., Rook-Weiler, K., Schönstein, D., Schreiber, A., Serbest, E., Steffens, M., Tomys-Brummerloh, A., 2017. Frei verkettete wandlungsfähige montage, in: Internet of Production für agile Unternehmen. C. Brecher, F. Klocke, R. Schmitt, G. Schuh, Eds. AWK Aachener Werkzeugmaschinen-Kolloquium.

[26] Scholz-Reiter, B., Görges, M., Philipp, T., 2009. Autonomously controlled production systems—influence of autonomous control level on logistic performance. CIRP annals 58, 395–398.

[27] Scholz-Reiter, B., Hamann, T., 2008. The behaviour of learning production control. CIRP annals 57, 459–462.

[28] Schönemann, M., Herrmann, C., Greschke, P., Thiede, S., 2015. Simulation of matrix-structured manufacturing systems. Journal of Manufacturing Systems 37, 104–112.

[29] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O., 2017. Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347 .

[30] Stricker, N., Kuhnle, A., Sturm, R., Friess, S., 2018. Reinforcement learning for adaptive order dispatching in the semiconductor industry. CIRP Annals 67, 511–514. doi:10.1016/j.cirp.2018.04.041.

[31] Sutton, R.S., Barto, A.G., 2018. Reinforcement learning: An introduction. MIT press.

[32] Tampuu, A., Matiisen, T., Kodelja, D., Kuzovkin, I., Korjus, K., Aru, J., Aru, J., Vicente, R., 2017. Multiagent cooperation and competition with deep reinforcement learning. PloS one 12.

[33] Tuyls, K., Nowe, A., Guessoum, Z., Kudenko, D., 2008. Adaptive Agents and Multi-Agent Systems III. Adaptation and Multi-Agent Learning: 5th, 6th, and 7th European Symposium, ALAMAS 2005-2007 on Adaptive

and Learning Agents and Multi-Agent Systems, Revised Selected Papers. Springer.

[34] Wang, S., Wan, J., Zhang, D., Li, D., Zhang, C., 2016. Towards smart factory for industry 4.0: a self-organized multi-agent system with big data based feedback and coordination. Computer Networks 101, 158–168.

[35] Williamson, D.P., Hall, L.A., Hoogeveen, J.A., Hurkens, C.A., Lenstra, J.K., Sevast'janov, S.V., Shmoys, D.B., 1997. Short shop schedules. Operations Research 45, 288–294.