Niklas Kiefl
Frederik Wulle
Clemens Ackermann
Daniel Holder *Editors*

Advances in Automotive Production Technology – Towards Software-Defined Manufacturing and Resilient Supply Chains

Stuttgart Conference on Automotive Production (SCAP2022)

ARENA2036

Springer

ARENA2036

Series Editor

ARENA2036 e.V., *Stuttgart, Germany*

Die Buchreihe dokumentiert die Ergebnisse eines ambitionierten Forschungsprojektes im Automobilbau. Ziel des Projekts ist die Entwicklung einer nachhaltigen Industrie 4.0 und die Realisierung eines Technologiewandels, der individuelle Mobilität mit niedrigem Energieverbrauch basierend auf neuartigen Produktionskonzepten realisiert. Den Schlüssel liefern wandlungsfähige Produktionsformen für den intelligenten, funktionsintegrierten, multimaterialen Leichtbau. Nachhaltigkeit, Sicherheit, Komfort, Individualität und Innovation werden als Einheit gedacht.

Wissenschaftler verschiedener Disziplinen arbeiten mit Experten und Entscheidungsträgern aus der Wirtschaft auf Augenhöhe zusammen. Gemeinsam arbeiten sie unter einem Dach und entwickeln das Automobil der Zukunft in der Industrie 4.0.

The book series presents the results of an ambitious research project in automotive production. The goal of the project is the development of a sustainable Industry 4.0 and the realization of a technology shift that will realize the mobility of the future with low energy consumption based on innovative production concepts. The key is provided by intelligent and flexible forms of production. Sustainability, safety, comfort, individuality and innovation are conceived as a unity. Scientists from various disciplines work together with experts and decision-makers from industry on an equal footing. Together, they work under one roof and develop the automobile of the future in Industry 4.0.

Niklas Kiefl · Frederik Wulle ·
Clemens Ackermann · Daniel Holder
Editors

Advances in Automotive Production Technology – Towards Software-Defined Manufacturing and Resilient Supply Chains

Stuttgart Conference on Automotive Production (SCAP2022)

Springer

Editors
Niklas Kiefl
ARENA2036 e.V.
Stuttgart, Germany

Frederik Wulle
ARENA2036 e.V.
Stuttgart, Germany

Clemens Ackermann
ARENA2036 e.V.
Stuttgart, Germany

Daniel Holder
ARENA2036 e.V.
Stuttgart, Germany

Editorial

Stuttgart Conference on Automotive Production: *Towards Software-Defined Manufacturing and Resilient Supply Chains*

A growing variety of products combined with shorter production times represents a major challenge to the flexibility of today's machines and production systems. Numerous necessary modifications to the system must be considered on the hardware side, which can lead to expensive modifications of the hardware as well as the software. However, this situation stands in stark contrast to the relevance of being able to produce economically and ecologically despite highly volatile markets and under dynamic conditions. Software-defined manufacturing (SDM) revolutionizes the traditional production process at this position, as the import of new functions is possible independently of the hardware and purely by software. SDM is a concept in which manufacturing processes are controlled by software and which uses computer programs to control the entire manufacturing infrastructure. This allows greater flexibility and efficiency in production and allows companies to respond quickly to changes in demand and the market.

The 2nd edition of the Stuttgart Conference on Automotive Production in 2022 was organized by ARENA2036 in close collaboration with the Institute for Control Engineering of Machine Tools and Manufacturing Units (ISW) of the University of Stuttgart. The ISW is one of the leading research centres in the field of control engineering. ARENA2036 is the innovation platform for future technologies, where science and industry work together on the development of new and potentially disruptive future technologies.

The aim of the conference was to address advanced technologies in robotics and automation, logistics, and manufacturing innovations for future automotive production. The theme of this year's conference was: *Towards Software-Defined Manufacturing and Resilient Supply Chains.*

The contributions in this book are arranged thematically in three parts, allowing the readers to choose their fields of interest from a broad range of topics. Part A focuses on *Software-defined Manufacturing*, Part B concentrates on *Data-driven Technologies*, and Part C discusses *Advanced Manufacturing and Sustainability*. Every single contribution[1] was peer-reviewed and evaluated by the members of the scientific committee; an international group of 23 experts with great expertise in their respective research areas.

The success of the SCAP2022 was based on the national and international scientists who presented their latest research results, exchanged ideas, and promoted the exchange of information between industry and science. Without the commitment and scientific passion of the authors, this success could have never been achieved.

---

[1] This book includes contributions submitted directly by the respective authors. The editors cannot assume responsibility for any inaccuracies, comments, and opinions.

In addition, the session chairs and the members of the scientific committee should be thanked for their great support in organizing and supporting the conference. Furthermore, we would like to thank the thought leaders, who delivered a keynote during our conference! Thank you for sharing your insights and for donating your time.

There are—as always—too many individuals to be named one by one. We want to thank everyone who contributed to making SCAP2022 a success story.

In closing, we kindly invite you to stay in touch with ARENA2036, to stay tuned for the next *Stuttgart Conference on Automotive Production*, and to enjoy the following papers.

December 2022   Niklas Kiefl
Frederik Wulle
Clemens Ackermann
Daniel Holder

Contents

x   Contents

Part A: Software-Defined Manufacturing

Real-Time Capable Architecture
for Software-Defined Manufacturing

Stefan Oechsle[✉], Moritz Walker, Marc Fischer, Florian Frick, Armin Lechler,
and Alexander Verl

Institute for Control Engineering of Machine Tools and Manufacturing Units,
University of Stuttgart, Seidenstr. 36, 70174 Stuttgart, Germany
`stefan.oechsle@isw.uni-stuttgart.de`

Abstract. Production systems are characterized by static configurations and slow adaption to changing requirements. They no longer meet current trends in mutability and dynamic adaptation. Software-defined Manufacturing (SDM) like other software-defined approaches leverages abstraction of hardware to achieve higher flexibility. Based on abstracted hardware, software defines desired functionalities. Requirements from the Operational Technology (OT), especially determinism, must be combined with the flexibility and interoperability of Information Technology (IT). This paper proposes a stack that enables the implementation of SDM based on a requirements analysis. It covers the main phases of the life cycle of automation applications and additional requirements from SDM. We derive the necessary components while resorting to existing approaches whenever possible. Means for applications engineering, configuration, deployment, and orchestration, as well as execution at run time, are developed.

Keywords: real time · software defined manufacturing · container · virtualization

1 Introduction

Digitalization is the most important driver of innovation across all industries and is strategically important for providing the flexibility and adaptability needed to succeed in times of highly volatile markets and rapidly changing requirements. Even though increasingly flexible and adaptable systems have been under development for a long time, production systems remain relatively rigid. The IT world, with its high innovative power and dynamics, which is characterized by adaptability, rapid innovation cycles, and a flexible response to customer requirements, is seen as a successful counter design.

1.1 Limitations of State of the Art Manufacturing Systems

Industrial systems combine application-specific hardware and software, compute platforms, and communication infrastructure. Since machines and systems are

typically tailored to specific products, the hardware and software are tightly coupled, not adaptable, and integrated as a proprietary system. Furthermore, these systems are characterized by heterogeneous technologies, manufacturer-specific ecosystems, and monolithic solutions. Compute platforms are typically highly specialized controllers implemented as dedicated devices, e.g., PLCs. While real-time performance is a key requirement, the reasons for dedicated, decentralized devices are historic and business-model-related. Virtual PLCs are a trend; however, the used platforms are typically still dedicated and specialized devices. Connectivity is a key enabler for industrial systems, connecting physical devices and compute platforms. In the context of digitalization, the connectivity towards IT has become increasingly important. Today, various field buses are used, which are optimized but lack interoperability and connectivity. Current trends like Time-sensitive Networking (TSN) and related technologies like 5G, DetNet, or OPC UA on the higher layers have the potential to replace proprietary systems with standardized and interoperable solutions.

1.2   Software-Defined Manufacturing as a New Paradigm

A significant step toward IT-like flexibility and adaptability in production environments requires a paradigm shift: besides technologies, also methods, processes, business models and, in particular, the mindset must evolve. This new paradigm is referred to as Software-defined Manufacturing (SDM). The implementation of SDM requires a rethink from the system architecture to the technical implementation. Following the example of IT and the divide-and-conquer approach prevailing there, a division of the system takes place: cooperating subsystems, which have clear tasks, are connected by open interfaces and are characterized by interoperability. Another decisive factor is consistent abstraction by means of a layer model, which abstracts and encapsulates complex technologies for applications. Based on a requirements analysis in Sect. 2, we propose an architecture that enables the implementation of SDM in Sect. 3. Following the divide-and-conquer approach, our architecture is responsible for real-time network management and orchestration. Section 4 provides concluding remarks.

2   Requirements Analysis Based on Related Work

We identify basic requirements from the automation application development life cycle (AADLC) and emerging ones from Software-defined Manufacturing (SDM). Note that we focus on software, computing and networking infrastructure. We exclude the problem of designing SDM-enabling mechanical systems.

2.1   Requirements from the Automation Application Life Cycle

The AADLC consists of four major phases. Requirements engineering and design address the specification and validation of (non-)functional requirements, which are then iteratively refined between customers and development teams [1].

One major non-functional requirement for automation software systems is adaptability [2]. The development phase focuses on implementing application logic using Programmable Logic Controllers (PLCs). During commissioning, engineering methods have to be used, which support a quick and error-free set-up of automation systems based on pre-engineered modules [3]. Requirements will evolve as automation plants have lifetimes of several years. Thus, life cycle management and automated maintenance of existing applications must be employed [4,5]. A significant problem in changes to automation software is the degradation of code quality caused by uncontrolled on-site changes [6]. There is an increasing need to extend functionalities and scale automation systems [7]. Proprietary automation software frameworks make adding functionalities hard. E.g., PLCs do not allow simple integration of third-party real-time software, as PLC run-times might not include required libraries. During operation, reliable vertical and horizontal communication is required [3]. Summing up, we derive the following requirements:

R1: Applications' structure must be flexible, modular and extensible.
R2: Virtualization and hardware-independent deployment are needed for easier adaption of automation systems.
R3: Automation system software architectures should not be dependent on vendor-specific frameworks and allow simple third-party software integration.
R4: Changes to applications should be conducted via well-defined and quality-preserving processes while keeping track of which software version is deployed on which hardware.

2.2 Requirements Based on Published Approaches to SDM

Based on published work regarding SDM, we extend the requirements. To support SDM, the application layer of manufacturing hardware should be fully adaptable [8,9]. This exceeds the capabilities of reprogramming within planned functions, such as replacing NC-Code. Based on a minimal platform adaption layer (PAL), functionalities and communication interfaces, i.e., the cyber part, can be defined within cyber-physical systems' physical constraints. Orchestration, deployment, and configuration of services for basic functionalities can realize such an approach [8]. A common control plane then abstracts these low-level functionalities, which provides generic interfaces to upper layers [10]. ICT (information and communications technology) infrastructure has to be reconfigurable as well to achieve the necessary flexibility [11]. Software-defined Control (SDC) [12] is a concept similar to SDM. SDC consolidates information from production and enterprise levels. A central controller conducts configuration decisions leveraging application-level reconfiguration. Software-Defined Cloud Manufacturing is described in [13]. At run-time, applications are dynamically composed to match requests from upper layers. Higher-level systems define the application layer of manufacturing resources [14]. SDCWorks [15] provides formal methods for SDC. It allows analysis and verification of, i.a., real-time requirements. Summing up, we derive the following additional requirements:

R5: Manufacturing systems must be configurable by defining the application software layer.
R6: Defined application layers must be integrable in higher-level systems based on generic interfaces.
R7: The ICT infrastructure must be configurable and provide QoS concerning networking and computing.
R8: Higher-level systems must be able to define the automation application software layer.

3   Architecture Proposal

3.1   Integration in a Conceptual SDM Framework

Figure 1 (left) shows our approach to SDM based on [9], for which we concretize the implementation of the tree layers of SDM. The conceptual framework is structured hierarchically. Initially, the steps needed for production are derived from a product description. Based on the manufacturing description and knowledge of available manufacturing hardware, necessary automation applications are defined by combining generic functionalities. These are encapsulated in reusable units of deployment, i.e., services. Now, it is known which high-level functionalities, e.g., milling a part based on G-Code, a composed application provides. Thus, northbound interfaces for integration in higher-level systems are defined. Standardized job interfaces, e.g., OPC UA-based ISA 95 job control [16], are used for this. Composed applications are then tested via virtual commissioning and formal methods, such as SDCWorks [15], which allow the verification of QoS and real-time requirements. Then, the composed application descriptions are annotated with said requirements. The annotated deployment descriptions now have to be deployed on-site on physical infrastructure. Deployment is done using the architecture below, combining network management and orchestration (see Fig. 1, right).

3.2   Virtualization and Orchestration

Services are encapsulated in containers or virtual machines. Linux patched with PREEMPT_RT is used as the operating system (OS), as real-time containers *and* VMs are only available for this OS. We opt for reservation-based hierarchical real-time scheduling based on the SCHED_DEADLINE policy. Corresponding and compatible scheduling mechanisms are available for containers [17] and virtual machines [18]. A Kubernetes-based real-time orchestrator such as REACT [19] is used to assign real-time services to compute nodes. Network and application configuration are further steps needed to fully bring composed applications up.
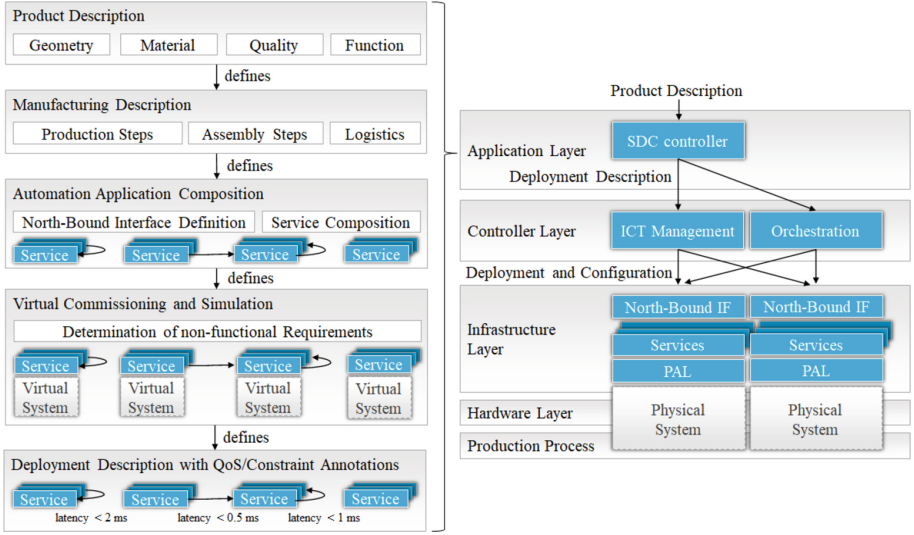
Fig. 1. Extended approach to SDM based on [9] on the left and Network management and orchestration as integral components of SDM on the right. Here, the SDC controller generates deployment descriptions as described above.

3.3   Network and Application Configuration

Description of the Physical Infrastructure. Knowledge of the underlying infrastructure is essential for network configuration and application deployment. In addition to general information about the type of connected devices, such as bridges and (bridged) endpoints, the available resources and capabilities of the individual devices are also particularly relevant. Examples of this are the OS, the installed RAM, the number of processor cores and their clock frequency. The capabilities of the devices can provide information about their real-time characteristics, both in terms of program execution and communication. In addition, the topology of the network must be known. The topology includes a description of which devices are connected in the network, the properties of the individual connections, such as wired or wireless, and the maximum transmission rate. Mechanisms like the Link Layer Discovery Protocol (LLDP) enable partially automatic network topology detection. This protocol allows information about the connection layout, type of connected devices, and other information such as MAC addresses to be read out. However, as things stand, additional mechanisms or manual additions are needed to retrieve device-specific information, such as resources. However, to enable the mutability required by SDM through dynamic configurations and deployments, such a holistic mapping is necessary so that deployability can be evaluated while maintaining real-time capability.

Description of Distributed Applications and Service Modularization.
For the requirements to be met by SDM and a flexible real-time network, the
description of the physical infrastructure alone is not sufficient: A description of
the distributed real-time application is also required to perform dynamic deploy-
ments. In addition to a list of the individual real-time applications involved
and their specific (hardware) requirements, the interaction or interrelationships
between the services are also particularly important. On the one hand, data
exchange must be considered here to answer the questions of which data (quan-
tities) should be exchanged and how often between which services. On the other
hand, it must be possible to define dependencies, for example, in the form of
fixed sequences and deadlines to be met with regard to communication between
the services. This holistic description of the logical flow, i.e., service descrip-
tion, of the overall application, in combination with the available (hardware)
resources of the physical infrastructure and the current workloads, enables the
deployability of the services on the nodes to be evaluated. In addition, opti-
mizations are possible, e.g., to consistently utilize the nodes or to defragment
unused resources left behind by terminated services. To achieve the goals of SDM
and thus the interaction of different components, converged communication is
required, as well as converged services. In addition to interchangeability, a high
possible combinability of multiple services also increases flexibility and the holis-
tic feasibility of the overall applications. One way to achieve this is a uniform
service specification. The specification defines a fixed scheme or structure for the
creation and implementation of services. This structure enables a uniform defi-
nition of inputs and outputs and additional meta information. When designing
the overall application, this provides an immediate check of the combinability
of services based on the required inputs and the outputs provided. Furthermore,
additional encapsulation of complexity can be achieved by abstracting real-time
communication. Here, abstracted send and receive methods are provided to the
service, generalizing the underlying communication mechanisms (e.g., access to
the network interface and time handling) depending on the technologies and
operating systems used. The goal of modularization or abstraction of services
is to simplify the creation of the actual application logic. The implementation
does not necessarily require knowledge of the underlying technologies (e.g., TSN,
5G, WiFi 6, OPC UA, or MQTT). The abstraction also enables separation of
responsibilities: Developers define individual applications without knowledge of
the infrastructure used, and the integrator, e.g., the SDC controller from Fig. 1,
defines the interaction and data exchange of the distributed real-time applica-
tion by linking multiple, individual applications without knowing the concrete
implementation details.

Deployment and Configuration Workflow. Figure 2 shows in the upper
part the deployment and orchestration workflow described below, while the lower
part details the sequence of the individual steps.
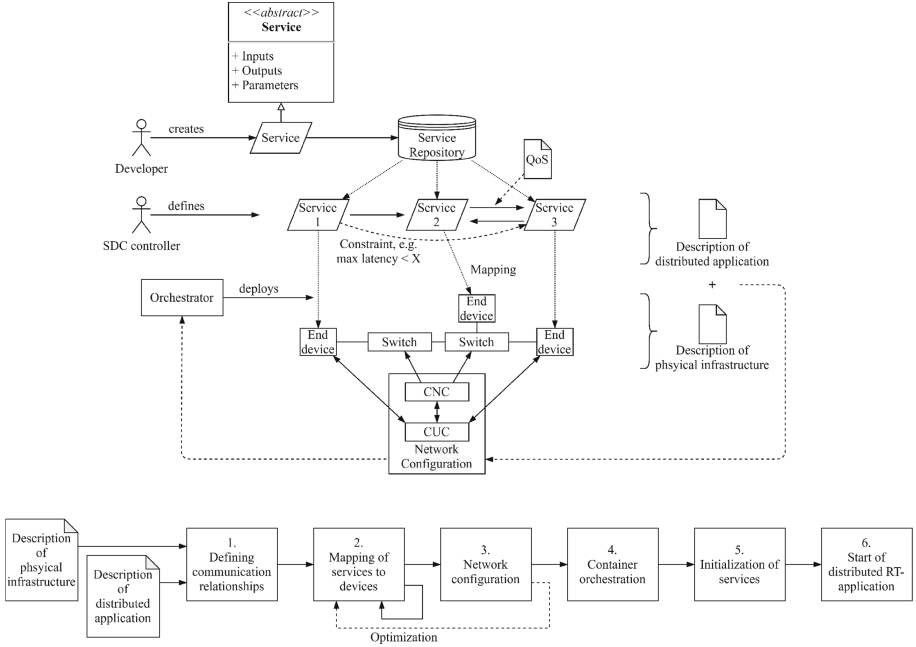
Fig. 2. Overview of the deployment and orchestration workflow.

The two descriptions, physical infrastructure and logical flow (service composition), are mapped in the deployment and configuration of the distributed real-time application. On the one hand, the communication relationships between the individual services are defined. As mentioned above, this is done by the SDC controller in Fig. 1 by linking the inputs and outputs of the respective services to define the logical data flow of the overall application. In the form of constraints, it is possible to specify this logical flow further using conditions to be met (e.g., specification of the maximum latency between two services) and thus limit the solution space for the subsequent configuration. The communication relationships are detailed by specifying various Quality of Service (QoS) parameters. These include the transmission interval, priority specifications, the required reliability, e.g., through redundant transmission, and the earliest and latest possible transmission times within an interval. Another component of the deployment is assigning individual services to physical end devices. This mapping represents an optimization problem that can be solved manually or automatically using suitable algorithms. The solution space depends on the specified communication relationships and the defined constraints, as well as on the provided resources of the end devices and the required resources of the services. A concrete example would be the compliance with the constraint of a minimum throughput of data between two services, which has to be evaluated depending on the available bandwidths given by the topology (e.g., speed of the cables, number of hops).

The result is used as input for the configuration. First, the configuration of the communication takes place. Using TSN, the standard provides three different approaches: centralized, decentralized, and hybrid configuration [20]. Only the central approach is considered here, with the assumption of the existence of a Central User Controller (CUC) and a Central Network Controller (CNC). The CNC represents a central network management instance, and the CUC can be understood as an intermediary between the user or application and the CNC. The CUC first collects the communication relationships previously defined in the deployment description. In this case, a communication relationship describes a TSN stream consisting of the talker, the listeners and specified QoS. From the CUC, all stream requests are then forwarded to the CNC for calculation. The CNC has an overview of streams already in the network and thus knows the available resources. Accordingly, a schedule, i.e., the specific transmission times, is calculated based on the streams already computed and the requested streams. If no solution can be found, services can be redistributed to other nodes. This process is repeated until an (optimized) solution is found. The computed configurations for the requested streams are then sent back to the CUC. These contain, among other things, the specific send offset relative to the interval start, the calculated accumulated latency, the VLAN ID, the PCP and the source and destination MAC. The existing information from deployment description and mapping is then enriched with the stream configurations and serves as input to the subsequent container orchestration step. Accordingly, the orchestrator can deploy the containers on the nodes assigned to them, along with the computed communication configuration. Alternatively, it is feasible to communicate this information to the containers only after deployment using a provided interface. In this way, subsequent changes regarding communication (e.g., additional listeners) could also be taken into account. However, this also requires an additional communication channel and effort. The deployment is followed by the initialization of the services based on the mapping and the communication configuration. First, the initialization must set up the real-time communication channels. Based on the configuration, the talkers and listeners are created here based on their MAC addresses. Using OPC UA, this can be implemented, e.g., through the publish-subscribe extension (Part 14) [21]. Likewise, the execution of the service must be coordinated, with the actual application logic of the service being encapsulated in a cyclically executed RT thread. The process within a cycle is comparable to the execution logic of a PLC: at the beginning, the inputs are read, i.e., data is received, based on which new data is subsequently calculated.

Table 1. Comparison of the identified requirements and the proposed solutions through the architecture

Requirement(s)	Proposed solution by the architecture
R1	Service-oriented real-time architecture
R2	Hardware independence through virtualization
R3	Abstraction layer for service development
R4	Deployment and orchestration of containers and VMs based on common resource interfaces (see Sect. 3.2)
R5, R8	Deployment descriptions as API for the SDM application layer (see Sect. 3.1)
R6	Integration through well know high-level interfaces of manufacturing units, e.g., job control via OPC UA and corresponding companion specifications
R7	Establishment of a convergent network based on TSN, automated network, and application configuration (see Sect. 3.3)

4   Conclusion and Future Work

A major goal of SDM is to increase the reconfigurability of automation systems by making the functionalities of manufacturing hardware definable through software. Based on a requirements analysis regarding the automation application life-cycle and SDM, we proposed a conceptual architecture that allows higher-level systems to define the application layer of manufacturing systems while meeting QoS requirements concerning networking and computation. Furthermore, we describe how such software-defined functionalities can be integrated into higher-level systems based on a job order-oriented northbound interface. Table 1 summarizes how the requirements identified in Sect. 2 are addressed by the proposed architecture. As part of future work, we plan to implement the described architecture.

References

1. Dubey, A.: Evaluating software engineering methods in the context of automation applications. In: 2011 9th IEEE International Conference on Industrial Informatics, pp. 585–590 (2011)
2. Fischer, J., Bougouffa, S., Schlie, A., Schaefer, I., Vogel-Heuser, B.: A qualitative study of variability management of control software for industrial automation systems. In: 2018 IEEE International Conference on Software Maintenance and Evolution, ICSME, Piscataway, NJ, pp. 615–624. IEEE (2018)

3. Dotoli, M., Fay, A., Miśkowicz, M., Seatzu, C.: An overview of current technologies and emerging trends in factory automation. Int. J. Prod. Res. **57**(15–16), 5047–5067 (2019)

4. Vyatkin, V.: Software engineering in industrial automation: state-of-the-art review. IEEE Trans. Ind. Inf. **9**(3), 1234–1249 (2013)

5. Froschauer, R., Dhungana, D., Grunbacher, P.: Managing the life-cycle of industrial automation systems with product line variability models. In: Proceedings of the 34th Euromicro Conference on Software Engineering and Advanced Applications, Los Alamitos, California, pp. 35–42. IEEE Computer Society (2008)

6. Vogel-Heuser, B., Fischer, J., Rosch, S., Feldmann, S., Ulewicz, S.: Challenges for maintenance of PLC-software and its related hardware for automated production systems: selected industrial case studies. In: Koschke, R., Krinke, J., Robillard, M. (eds.) 2015 IEEE International Conference on Software Maintenance and Evolution (ICSME 2015), Piscataway, NJ, pp. 362–371. IEEE (2015)

7. Moga, A., Sivanthi, T., Franke, C.: OS-level virtualization for industrial automation systems. In: Ossowski, S., (ed.) Proceedings of the 31st Annual ACM Symposium on Applied Computing, pp. 1838–1843. ACM, New York (2016)

8. Lechler, A., Verl, A.: Software defined manufacturing extends cloud-based control. In: Proceedings of the ASME 12th International Manufacturing Science and Engineering Conference - 2017. The American Society of Mechanical Engineers, New York (2017)

9. Lechler, A., Riedel, O., Coupek, D.: Virtual representation of physical objects for software defined manufacturing. DEStech Trans. Eng. Technol. Res. (2018). ICPR

10. Yang, C., Lan, S., Shen, W., Wang, L., Huang, G.Q.: Software-defined cloud manufacturing with edge computing for industry 4.0. In: Rachedi, A., Guizani, M. (eds.) IWCMC 2020, Piscataway, NJ, pp. 1618–1623. IEEE (2020)

11. Nayak, N.G., Durr, F., Rothermel, K.: Software-defined environment for reconfigurable manufacturing systems. In: 2015 5th International Conference on the Internet of Things (IOT), pp. 122–129. IEEE (2015)

12. Lopez, F., Shao, Y., Mao, Z.M., Moyne, J., Barton, K., Tilbury, D.: A software-defined framework for the integrated management of smart manufacturing systems. Manuf. Lett. **15**, 18–21 (2018)

13. Thames, L., Schaefer, D.: Software-defined cloud manufacturing for industry 4.0. Procedia CIRP **52**, 12–17 (2016)

14. Yang, C., Liao, F., Lan, S., Wang, L., Shen, W., Huang, G.Q.: Flexible resource scheduling for software-defined cloud manufacturing with edge computing. Engineering **22**, 60–70 (2023)

15. Potok, M., Chen, C.-Y., Mitra, S., Mohan, S.: SDCWorks: a formal framework for software defined control of smart manufacturing systems. In: 2018 ACM/IEEE 9th International Conference on Cyber-Physical Systems (ICCPS), pp. 88–97. IEEE (2018)

16. OPC Foundation: OPC 10031-4: OPC UA for ISA-95: Part 4: Job control, 09 May 2021

17. Abeni, L., Balsini, A., Cucinotta, T.: Container-based real-time scheduling in the Linux kernel. ACM SIGBED Rev. **16**(3), 33–38 (2019)

18. Abeni, L., Biondi, A., Bini, E.: Hierarchical scheduling of real-time tasks over Linux-based virtual machines. J. Syst. Softw. **149**, 234–249 (2019)

19. Struhar, V., Craciunas, S.S., Ashjaei, M., Behnam, M., Papadopoulos, A.V.: REACT: enabling real-time container orchestration. In: Proceedings of the 2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Piscataway, New Jersey, pp. 1–8. IEEE (2021)

20. LAN/MAN Standards Committee of the IEEE Computer Society: IEEE Std 802.1Qcc-2018. S.l. IEEE (2018)
21. Eckhardt, A., Muller, S., Leurs, L.: An evaluation of the applicability of OPC UA Publish Subscribe on factory automation use cases. In: 2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA), pp. 1071–1074. IEEE (2018)

Analysis of Real-Time Execution Models for Container-Based Control Applications

Moritz Walker[✉], Timur Tasci, Armin Lechler, and Alexander Verl

Institute for Control Engineering of Machine Tools and Manufacturing Units,
University of Stuttgart, 70174 Stuttgart, Germany
`moritz.walker@isw.uni-stuttgart.de`

Abstract. Software-defined Manufacturing (SDM) aims to enhance the flexibility of production systems. Classical automation systems are not a suitable technological basis for SDM. While their hierarchical, rigid structures are increasingly being dissolved. Container-based virtualization, and modular software architectures, gain traction in automation systems. However, today's PLCs are not a perfect fit for virtualization, as the control program still is a monolithic piece of software. We analyze cyclic and event-based real-time scheduling models for modular PLCs. Furthermore, techniques for reconfiguration at runtime are developed based on the selected execution models.

Keywords: Containers · Real-Time · Virtualization · Automation Systems · Software-defined Manufacturing

1 Introduction

Today's manufacturing systems only support manual reconfiguration at the application level. However, the control software, e.g., the NC kernel, is fixed and bound to hardware. Thus, adapting core functionality is impossible or requires high manual effort. For this reason, the fixed programming of physical machines via PLC or NC code must be replaced by an adaptable software layer to enable Software-defined Manufacturing [1]. Software development for Programmable Logic Controllers (PLCs) is typically done monolithically. As monolithic control applications age, they become increasingly difficult to maintain. Components, e.g., function blocks, have low reusability and scalability [2]. Modular software architectures, such as Service-oriented Architectures (SOAs), address these drawbacks. Monolithic architectures are modularized into services, forming cohesive applications by loosely coupled interaction. Software containers are increasingly used to deploy modular architectures. We extend previous work [3,4] towards a modular control platform, which implements a Microservices architecture. Specifically, we extend the container-based control system by event-based and cyclic execution, i.e. orchestration, models that meet the requirements of modular software architectures and the real-time requirements of control systems.

2   Related Work

Cucinotta et al. [5] present an SOA in which real-time communication bypasses the Simple Object Access Protocol (SOAP) stack and directly uses UDP/IP. Service invocations are scheduled as sporadic tasks using Earliest Deadline First (EDF). Dai et al. [6] present an industrial SOA based on IEC 61499, within which function blocks offer functionalities as services. The communication takes place via SOAP and TCP/IP. Tsai et al. [7] present RTSOA, an SOA extended by soft real-time capabilities. A concept for a container-based, real-time automation platform is defined in [8]. The scheduling algorithm is Fixed Priority Preemptive Scheduling (FPS). While the other publications do not provide temporal synchronization of tasks across a compute node's boundaries, Telschig's container-based architecture [9] does so by introducing globally valid time slots for message exchange and task execution. Furthermore, not all architectures support a guaranteed execution order of tasks and the deployment or update of components at run-time. EDF, commonly used on single-core and multicore systems, is also suitable for scheduling applications whose tasks have dependencies that can be modeled as DAGs. A DAG is transformed into deadlines and activation times of the subtasks for this purpose [10,11]. The methods known for single-processor and multiprocessor systems are extended by Rivas et al. [12] for distributed systems. Saifullah et al. [10] present a scheduling method for parallel real-time execution of multiple periodic DAGs on a multicore processor. An algorithm decomposes one or more DAGs into sequential tasks by assigning activation offsets and deadlines to individual tasks. Global EDF is used as the scheduling algorithm. Jiang et al. [11] present a similar decomposition algorithm for DAG tasks. Peng et al. [13] present methods for the FPS and EDF of DAGs, where no decomposition is necessary.

3   Analysis of Scheduling Methods
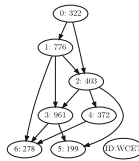
Fig. 1. An exemplary DAG task consists of subtasks with worst-case execution times (WCETs). Every subtask is deployed as a container.

Cyclic Scheduling of Dependent Tasks: Some use cases, such as sensor fusion, benefit from executing tasks in a specified order because this can reduce the end-to-end latency. Therefore, we compare two [10,11] principally applicable methods for cyclic scheduling of task sets modeled as DAGs. Both methods rely

(a) Flowchart before decomposition.

(b) Offset and deadline assignment [10] ($T = 8\ ms$).

(c) Offset and deadline assignment [11] ($T = 10\ ms$).

(d) Number of schedulable task sets on a m-core processor.
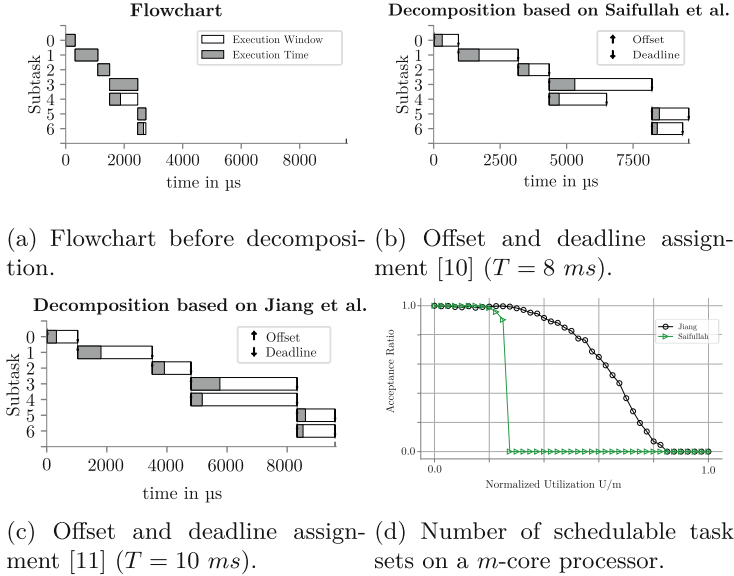
Fig. 2. Comparison of Jiang's [11] and Saifullah's [10] decomposition algorithms.

on the decomposition of DAG tasks, i.e., the assignment of offsets and deadlines to subtasks. Furthermore, both methods use EDF at runtime. Figure 1 depicts an exemplary task modeled as a DAG. In the following, a DAG is called a task and its components are called subtasks. Arrows symbolize precedence constraints. The methods considered in this work are the algorithms of Saifullah et al. [10] and Jiang et al. [11]. Figure 2 shows the decomposition of the DAG in Fig. 1 under Saifullah's (Fig. 2b) and Jiang's (Fig. 2c) methods. Vertical markers at the end of the abscissa symbolize specified deadlines. Both methods are compared as follows. Random task sets with varying parameters are generated, and the numbers of schedulable task sets are compared. The task set generation follows the following scheme based on [10,14]:

1. Determination of the parameters: The connection probability $p \in (0,1)$ specifies the probability of precedence constraints. n is the number of subtasks per task, U_{set} is the desired processor utilization of the entire task set. $\beta < 1$ is a factor used to influence individual task utilization.
2. The generation of n random task is done using the $G(n,p)$ method according to Ernős-Rényi.
3. The WCET $C_{i,j}$ of a subtask j in task i is randomly chosen from the interval [100 µs, 1000 µs]. The length of the critical path L_i, i.e., the sum over the WCETs of subtasks on the longest path in the DAG, and the workload W_i, i.e., the sum over the WCETs of all subtasks in the DAG, are calculated.
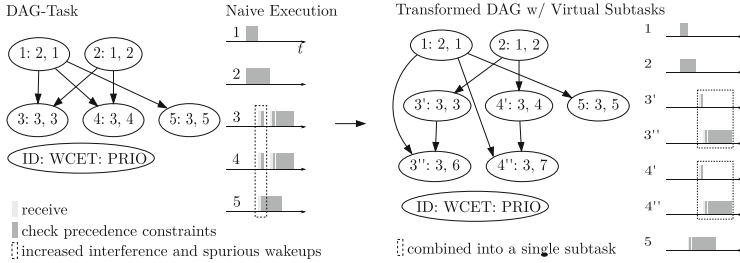
Fig. 3. DAG-transformation to minimize intra-task-interference under G-FPS.

4. The period of a task T_i is chosen randomly from the interval $[L_i, W_i/\beta]$. Thus, the processor utilization U_i lies in the interval $[\beta, W_i/L_i]$. The deadline is implicitly set to be identical to the period.
5. Step 2 to 4 is repeated until $\sum_{i=1}^{n} U_i > U_{set}$ holds, and the period of the last task is chosen so that exactly the desired total utilization is reached.

1000 random task sets per utilization were generated using the parameters $p \in [0.01; 0.2]$, $n \in [20; 100]$, $\beta = 0.1$, and $C_{i,j} \in [100 \text{ μs}; 1000 \text{ μs}]$. Figure 2d shows the acceptance rates for the randomly generated task sets with different utilizations. Since Jiang's decomposition strategy can decompose significantly more task sets in a schedulable manner, it is applied if the application requires cyclic scheduling considering the execution order.

Event-Based Scheduling of Dependent Tasks. For event-based scheduling of DAG tasks, global EDF (G-EDF) [15] and global FPS (G-FPS) [14] are suitable algorithms. Using global FPS, the subtasks within a task are assigned priorities according to their topological order. Different tasks (DAGs) are assigned priorities according to Deadline Monotonic Scheduling (DMS) and thus, in the case of implicit deadlines, according to Rate Monotonic Scheduling (RMS). Due to the implementation-specific details of the SCHED DEADLINE scheduler, schedulability cannot be tested according to Melani et al. [18]. The deadline of a thread is relative to its activation time on Linux. For the schedulability test, according to Melani et al., and similar tests, its deadline must be relative to the activation time of the source subtask. Thus, G-FPS is used for the event-based scheduling of container-based DAG tasks. Pathan's test [14] is applied to check the schedulability. As discussed in [3,4], the container-based control system uses the socket API for inter-service communication. The `select`, `poll`, or `epoll` syscalls can be used to wake up tasks, when a message is delivered. The ideal implementation of the execution model would require a syscall that blocks until a task has received messages from all of its predecessors. Such syscall is not available on Linux, which leads to unnecessary context switches and increased intra-task interference between the subtasks of a DAG, as illustrated on the left in Fig. 3. Our approach to minimizing the intra-task interference is described in the following and shown in Fig. 3. Subtasks that have more than one predecessor are decomposed into sequential virtual subtasks. The number of virtual

subtasks corresponds to the number of predecessors. For a subtask $v_{i,j}$ of a DAG task Π_i with predecessors $pred_{i,j} = \{pred_{i,j,1}, pred_{i,j,2}, ...\}$, the decomposition into sequential virtual subtasks is done in four steps:

1. First, the response time analysis (RTA) is performed using the original task set and the test of Pathan et al. [14]. Here, the worst-case response time $R_{i,j}$ is calculated for each subtask $v_{i,j}$.
2. The subtask $v_{i,j}$ is divided according to the cardinality of the predecessors into $n_{i,j}^{virt} = |pred_{i,j}|$ virtual sequential subtasks $v_{i,j}^k$. The first $k = 1, ..., n_{i,j}^{virt} - 1$ subtasks are assigned the WCET $C_{i,j}^k = \tau_{recv} + \tau_{proc}$, where τ_{proc} corresponds with the message processing time and τ_{recv} corresponds with the delay, needed to fetch the message from the messaging system. The last virtual subtask is assigned the WCET $C_{i,j}^{|pred_{i,j}|} = \tau_{recv} + \tau_{proc} + \tau_{task} + n_{i,j}^{sub}\tau_{send}$. τ_{send} denotes the time needed to send a message.
3. The predecessors $pred_{i,j}$ of the subtask are sorted in descending order according to their response times $R_{i,j}$. The virtual subtasks are ordered in ascending order for $k = 2, ..., n_{i,j}^{virt}$ and the predecessors $pred_{i,j,k}$ and $v_{i,j}^{k-1}$ are assigned to the subtask. The first virtual subtask receives only $pred_{i,j,1}$ as predecessor.
4. Finally, we evaluate schedulability by applying Pathan's RTA [14] to the transformed DAG tasks.

If a DAG extends across the boundaries of a system, schedulability under global FPS is evaluated by the RTA of Peng et al. [13]. This method considers that subtasks not executed on the same processor cores cannot interfere with each other.

4 Execution Models for the Container-Based PLC

4.1 Cyclic Execution Model

Scheduling: To support independent and DAG applications simultaneously and guarantee schedulability for high processor loads, the cyclic execution model uses the decomposition method of Jiang et al. [11] and global EDF. This execution model is suitable for implementing cyclic DAG tasks with implicit or constrained deadlines ($D_i \leq T_i$). The interaction between the communication system and application subtasks follows the synchronous interaction pattern. Thus, the execution time of a subtask results in

$$\tau_{total} = n_{pub}\tau_{recv} + \tau_{task} + n_{sub}\tau_{send} + \tau_{aug}. \tag{1}$$

The task's execution time is augmented with the latency τ_{aug}, which is the time needed until all messages are transmitted to the successors. n_{pub} is the number of predecessors and n_{sub} is the number of successors. In the schedulability test, only the actual execution time of the subtasks is considered.
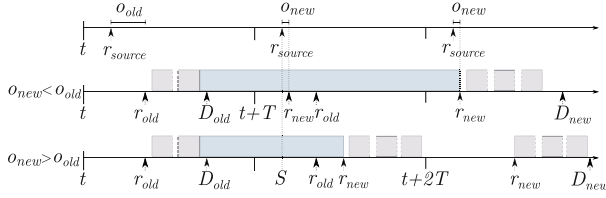
Fig. 4. Concept for reconfiguration by reassigning deadlines and offsets.

Runtime Reconfiguration: The reconfiguration, that is, adding a subtask v_{i,n_i+1} to a DAG task Π_i with subtasks $\{v_{i,1}, ..., v_{i,n_i}\}$ and period T_i, is done in three steps. v_{source} is any source subtask of Π_i, e.g. $v_{i,1}$. The update strategy is exemplified for one subtask $v_{i,j}$, but is applied simultaneously to all subtasks of the DAG. To simplify the notation, $T = T_i$ is the period of DAG task Π_i and $r^k_{source} = r^k_{i,1}$ is the request time of the k-th instance of the source subtask. The relative offset $o = o_{i,j}$ of the subtask $v_{i,j}$ refers to r^k_{source}, which is increased by T for each invocation of the DAG task: $r^k_{source} = r^{k-1}_{source} + T$. The request time r^k of the k-th instance of subtask $v_{i,j}$ is $r^k = r^k_{source} + o$. The deadline D of subtask $v_{i,j}$ is relative to its request time. For r^k_{source} and r^k, the notations r_{source} and r are used. First, new deadlines and offsets are assigned to subtasks by the decomposition procedure. The old and new offsets and deadlines of subtask $v_{i,j}$, are denoted o_{old} and o_{new}, and D_{old} and D_{new}. The next step is to notify the subtasks about their new deadlines and offsets. A message is sent to each subtask, containing the new deadlines and offsets and a global synchronization point. The global synchronization point is derived from the request time r^k_{source} of the source subtask: $S = r^k_{source} + xT$. $x \in \mathbb{N}$ can be freely chosen. While, e.g., Xenomai natively supports the allocation of offsets between threads, on Linux, this can only be done using a timed sleep. If the request time of a subtask is shifted to the left, that is ($o_{new} < o_{old}$), the corresponding subtask would have to be executed a second time within one period. Since the deadline of the subtask may already be exceeded at this point, and the Constant Bandwidth Server (CBS) may have no remaining bandwidth, the subtask might be throttled. For this reason, the reallocation follows the strategy shown in Fig. 4. If $o_{new} < o_{old}$, the subtasks are paused until $S + T + o_{new}$, and the new deadline is assigned. If $o_{new} > o_{old}$, the subtasks are paused until $S + o_{new}$ and D_{new} is assigned. The third step includes the deployment of the new subtask. A consistent state transfer is necessary if a subtask is updated, i.e., replaced. First, the schedulability test is used to check whether the new subtask can be executed parallel, i.e., with identical offsets and deadlines, to the subtask to be replaced. If this is the case, the subtask starts. Otherwise, the deployment follows the procedure described above. Next, the necessary communication channels are initialized. The new subtask does not process incoming messages and leaves them in the message queue. A time-stamped image of the state of the internal variables of the component to be replaced is successively transmitted to the new subtask. Once the state is completely transmitted, the actual state is reconstructed using the messages
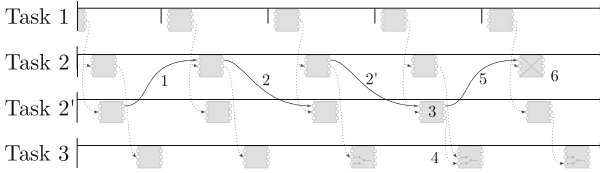
Fig. 5. Example of an update to replace a stateful subtask.

in the message queue. The new subtask starts executing its program logic and transmitting the output messages to its successors. The successors always apply the messages of the new subtask during the update process, as far as these are available. Finally, the original subtask terminates. Figure 5 illustrates the principle flow of a stateful reconfiguration. The DAG task consists of the source subtask task 1, task 2, and task 3, which are executed in this order. The stateful task 2 is replaced by task 2' at runtime. The reassignment of offsets and deadlines has already been done, and the required communication links have been established. (1) Task 2' requests the transfer of task 2's internal state and starts receiving messages from task 1. (2) The state's transmission begins, which extends over two cycles. (3) State reconstruction is conducted based on the messages in the message queue and the transmitted state. In this example, only one cycle is needed for this. Task 2' begins with the execution of the program. (4) Task 3 receives messages from tasks 2 and 2'. Messages from task 2 are dropped, and messages from task 2' are applied. Furthermore, task 2 is notified that the deployment of task 2' has been completed (5). (6) Task 2 terminates.

4.2   Event-Based Execution Model

Based on the results of the previous sections, the event-based execution model of the container-based control system is presented. The assignment of priorities and the execution at runtime is performed according to the procedures explained in Sect. 3. The developed model ensures that only one event-based activation is necessary for each subtask since all other precedence messages have already been delivered and can be processed without interrupting the subtask. To check schedulability, the RTA of Pathan et al. [14] and to check schedulability on a distributed system, the RTA by Peng et al. [13] are used. In event-based systems, it is unpredictable when an event may occur. For this reason, reconfiguration is done either offline or without consistent state transmission.

5   Validation of a Sample Use-Case

The real-time performance of the execution models was evaluated based on a sample application. A production line model is controlled by a single DAG task with eight services and seven deployment units (see Fig. 6). The WCETs of the PLC subtasks are shown in Table 1.

(a) Model production line[1]   (b) Corresponding DAG task
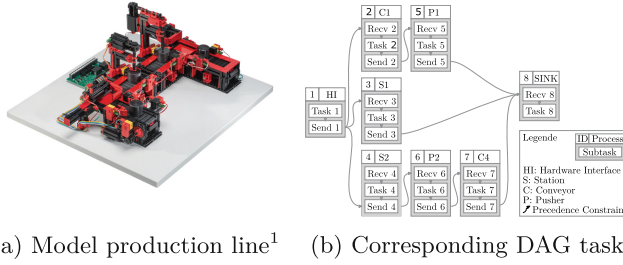
Fig. 6. Production line model with two processing stations.

Table 1. WCETs of the subtasks of the exemplary application for validation

Task	T1	T2	T3	T4	T5	T6	T7	T8
WCET [μs]	476	676	698	777	794	697	683	725

Validation of the Event-Based Execution Model: To validate the event-based execution model, two instances of the DAG task Π_1 and Π_2 were executed at a rate of $T_1 = 5$ ms and $T_2 = 10$ ms. The temporal behavior was recorded over a time span of six hours on a Raspberry Pi 3 equipped with PREEMPT_RT-Patch and Linux version 5.2.21. Figure 7 depicts the measured execution of Π_1 on the test system compared to the previously calculated response times of the subtasks. Only tasks 1, 2 and 3 exceeded the calculated response times. However, this was caused by WCET overruns and the jitter of the test system. If the scheduling method is to be used for safety-critical applications, the WCET must therefore be estimated sufficiently pessimistically. The second DAG task Π_2 also did not exceed the calculated response time.
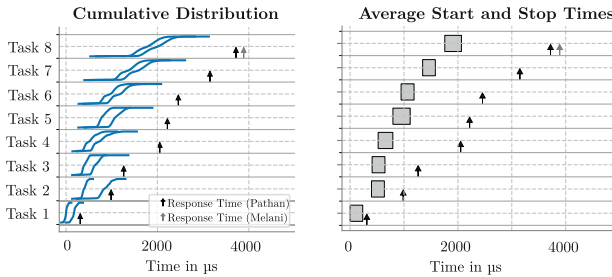
Fig. 7. Comparison between the calculated response times of DAG Π_1 with $T_1 = 5$ ms and the actual system behavior.

Validation of the Cyclic Execution Model: The cyclic execution model was also validated by running the application consisting of Π_1 and Π_2 ($T_1 = T_2 = 8$ ms) with implicit deadlines for six hours on the test system. Figure 8 shows the resulting artificial deadlines and measured behavior of Π_1 and Π_2. The precedence constraints were not violated at any time. The cyclic and event-based dependent execution models' evaluation shows that both can be employed under real-time requirements.
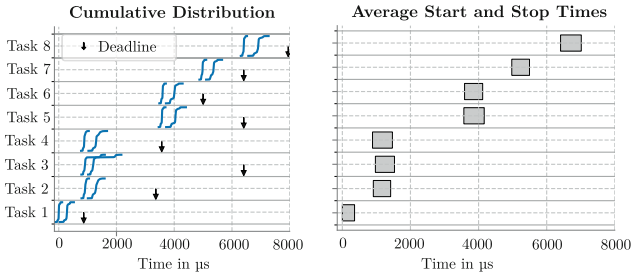
Fig. 8. Comparison of the calculated response times of DAG Π_1 with $T_1 = 8$ ms and the actual system behavior.

6   Conclusion and Future Work

In this work, we developed concepts for event-based and cyclic execution models. Both execution models support applications with independent tasks and applications, where the tasks' execution must occur in an explicitly specified order. This order is modeled as DAG. The execution models allow the development, deployment, and dynamic reconfiguration of distributed control applications. As part of future work, we plan to extend the used socket-based messaging system to support Time Sensitive Networking.

References

1. Lechler, A., Verl, A.: Software defined manufacturing extends cloud-based control. In: Proceedings of the ASME 12th International Manufacturing Science and Engineering Conference - 2017, New York, NY. The American Society of Mechanical Engineers (2017)
2. Mazzara, M., Meyer, B.: Present and Ulterior Software Engineering. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67425-4
3. Tasci, T., Melcher, J., Verl, A.: A container-based architecture for real-time control applications. In: Conference Proceedings ICE/IEEE ITMC, Piscataway, NJ, pp. 1–9. IEEE (2018)
4. Tasci, T., Fischer, M., Lechler, A., Verl, A.: Predictable and real-time message-based communication in the context of control technology. In: Weißgraeber, P., Heieck, F., Ackermann, C. (eds.) Advances in Automotive Production Technology – Theory and Application. A, pp. 264–271. Springer, Heidelberg (2021). https://doi.org/10.1007/978-3-662-62962-8_31
5. Cucinotta, T., Mancina, A.: A real-time service-oriented architecture for industrial automation. IEEE Trans. Ind. Inform. **5**(3), 267–277 (2009)
6. Dai, W., Vyatkin, V., Christensen, J.H., Dubinin, V.N.: Bridging service-oriented architecture and IEC 61499 for flexibility and interoperability. IEEE Trans. Ind. Inform. **11**(3), 771–781 (2015)
7. Tsai, W.T., Lee, Y.-H., Cao, Z., Chen, Y., Xiao, B.: RTSOA: real-time service-oriented architecture. In: Second IEEE International Workshop on Service-Oriented System Engineering 2006, Los Alamitos, California, pp. 49–56. IEEE Computer Society (2006)
8. RT-CASEs: Container-Based Virtualization for Temporally Separated Mixed-Criticality Task Sets. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik GmbH, Wadern/Saarbruecken, Germany (2019)
9. Telschig, K., Schonberger, A., Knapp, A.: A real-time container architecture for dependable distributed embedded applications, pp. 1367–1374. IEEE (2018)
10. Saifullah, A., Ferry, D., Li, J., Agrawal, K., Chenyang, L., Gill, C.: Parallel real-time scheduling of DAGs. IEEE Trans. Parallel Distrib. Syst. **25**(12), 3242–3252 (2014)
11. Jiang, X., Long, X., Guan, N., Wan, H.: On the decomposition-based global EDF scheduling of parallel real-time tasks. In: RTSS 2016, Piscataway, NJ, pp. 237–246. IEEE (2016)
12. Rivas, J.M., Javier Gutierrez, J., Carlos Palencia, J., Harbour, M.G.: Deadline assignment in EDF schedulers for real-time distributed systems. IEEE Trans. Parallel Distrib. Syst. **26**(10), 2671–2684 (2015)
13. Peng, X., Han, M., Deng, Q.: Response time analysis of typed DAG tasks for G-FP scheduling. In: Guan, N., Katoen, J.-P., Sun, J. (eds.) SETTA 2019. LNCS, vol. 11951, pp. 56–71. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-35540-1_4
14. Pathan, R., Voudouris, P., Stenstrom, P.: Scheduling parallel real-time recurrent tasks on multicore platforms. IEEE Trans. Parallel Distrib. Syst. **29**(4), 915–928 (2018)
15. Qamhieh, M., Fauberteau, F., George, L., Midonnet, S.: Global EDF scheduling of directed acyclic graphs on multiprocessor systems. In: Proceedings of the 21st International Conference on Real-Time Networks and Systems, RTNS 2013, pp. 287–296. Association for Computing Machinery, New York (2013)

Software-Defined Manufacturing for the Entire Life Cycle at Different Levels of Production

Sebastian Behrendt[1(✉)], Michael Martin[1], Alexander Puchta[1], Robin Ströbel[1], Johannes Fisel[2], Marvin C. May[1], Philipp Gönnheimer[1], Jürgen Fleischer[1], and Gisela Lanza[1]

[1] wbk Institute of Production Science, Karlsruhe Institute of Technology, Kaiserstraße 12, 76131 Karlsruhe, Germany
`sebastian.behrendt@kit.edu`
[2] Robert Bosch GmbH, Robert-Bosch-Campus 1, 71272 Renningen, Germany

Abstract. Increasingly volatile markets, higher numbers of product variants and more sophisticated customer demands lead to a soaring complexity of productions themselves and their operations. An enabling technology that allows to cope with this increased complexity is digitization, as it enables data capturing and data driven analysis in production. Software-defined manufacturing (SDM) empowers to fully use the potential of digitization by decoupling physical production hardware and the associated control software. This enables an increase in the versatility of existing resources through automated generation of software for instantiations and interventions in production control. With the aim of enlarging the abstraction and decoupling capabilities of SDM, this paper presents a concept to use SDM at different abstraction levels of the production over the whole life cycle. The different abstraction levels, i.e. machine, production system and production network, are decoupled based on a service-oriented approach that defines interactions between these abstraction levels. The requirements to implement this approach are determined for the different levels where special notice is given to changing requirements over the life cycle of the production. With respect to the requirements, recommendations are given considering the integration of the concept in existing productions. Finally, potential benefits of this concept are discussed.

Keywords: Software-defined manufacturing · brownfield · production system · production network · production life cycle

1 Introduction

Today, the environment in which manufacturing companies operate is characterized by volatility, uncertainty, complexity and ambiguity (VUCA world) [1, 2]. The requirements which arise in this environment are decisive for the success of companies. Flexibility and adaptability are enablers to cope with these requirements and crucial success factors that can ensure competitiveness [3].

Progressive digitization has evolved from a trend to an enabler in the global production network and offers companies the opportunity to deal with growing demands in

the VUCA-world [4]. A new paradigm that offers potential to exploit the full scope of digitization and remove the existing barriers is Software-defined Manufacturing (SDM) [5, 6]. Inspired by software-defined networking, SDM aims to decouple hardware from software, with the goal of increasing flexibility and changeability of both hardware and software. Therefore, SDM empowers increasing efficiency in planning and operating and reduces the complexity of the systems at the same time [5–7]. The full potential of SDM can be realized when it is used across all levels of production [3]. Starting from machines, through the production system, and up to the global production network, a continuity can be created by implementing standards and by using the same technologies as well as the same mindset. However, SDM should not only be applicable across all levels, but also along the entire production life cycle. The production life cycle is oriented on the product life cycle (Introduction, Growth, Maturity and Decline) [8]. During the life cycle of a production, requirements change and adjustments must be made which can be supported by SDM through a continuous adaption process which includes the development, planning and operation of a production [3, 6].

Based on existing approaches and technologies, it is now necessary to establish SDM across all levels and along the entire life cycle of a production in order to counteract the VUCA world effects. The remainder of this paper addresses existing approaches (Sect. 2), presents an architecture that aims to realize SDM over all layers of production (Sect. 3) which is illustrated in an exemplary use case (Sect. 4). Finally, the presented architecture is discussed and the paper concluded (Sect. 5).

2   State of the Art

The SDM paradigm aims at "adapting an entire production purely via software" [9]. Similar to other software-defined approaches such as software-defined networking [10], SDM pursues its goal by utilizing abstraction layers to decouple control software from production hardware [5]. SDM promises to increase flexibility and changeability of production and allows to handle the complexity of modern, digitized production sites [7]. Currently, different concepts for SDM architectures are discussed in the literature, that can be differentiated in modular service-oriented approaches [6, 7, 9] and centralized control approaches [11, 12]. Since modular service-oriented approaches suit the idea of abstraction, these approaches are presented in the following.

An essential foundation for the realization of SDM is a digitized production, where different assets in the production are cyber physical systems (CPS). A CPS refers to a system with integrated digital and physical capabilities that collects data and allows to control its physical components by software [13]. There exists a multitude of communication protocols with service-oriented architectures that can interact with and control CPS, such as OPC UA [14] or MQTT [15]. Both provide a basis for standardized communication across all levels of an enterprise. In their basic form, however, they do not consider the specific changes in a CPS over the life cycle. A promising concept that allows to create standardized data models that comprise all data to describe a CPS and allow for communication with the physical part of the CPS is the asset administration shell (AAS) [16, 17]. The AAS is promoted as a basis for a digital twin (DT) [16].

Despite multiple, sometimes contradictory definitions of a DT in the literature [18], most researchers agree that a DT is the virtual representation of a physical asset that

synchronizes with the real system [19]. In most cases, a DT incorporates models that allow making predictions about the real system [20]. Stark et al. [21] state that a DT consists of data and models that utilize this data to mimic the behavior of the real system. A definition relevant in regard of SDM is presented by Kritzinger et al. [18] who states that a DT must support automatic information flow from and to the real system. Thus, a DT needs to be able to control the physical system.

Besides SDM, there exists a plethora of approaches that aim to improve production by using CPS or digitization in general. The concept of cyber-physical production systems, which is proposed in [22], aims at a production with autonomous, interconnected and cooperative CPS. The concept targets a self-organized and highly flexible production. Similar approaches that aim for an intelligent, flexible and changeable production are biological [23], reconfigurable [24], fractal [25] or holonic [26] manufacturing systems. However, the above-mentioned concepts lack a conceptual software architecture that allows to use the CPS as proposed. This paper aims to close this gap by introducing an SDM architecture that allows for abstraction to reduce the complexity in production and enabling service-oriented manufacturing.

3 Approach

Realizing the SDM paradigm promises to improve the efficiency in planning and operation of production systems. The main enablers that are exploited to realize a software-defined production are *virtualization* and *abstraction.* As a result, operation and planning can be fully adopted by virtual representations of the production and interaction with the production is simplified by realizing abstraction with offered services.

In regard of SDM, virtualization describes the digital replication of a physical asset in production [7]. However, different requirements need to be fulfilled to achieve virtualization of the production system. The production assets, i.e. production resources such as machines, need to be CPS with virtual counterparts, i.e. DTs [27]. By combining these individual virtual counterparts of the assets in a model that also contains the connections of the assets, a virtual representation of the production system can be created. Similarly, virtual production systems can be combined to virtual production networks. Besides production resources, production systems and global production networks are also referred to as assets in the following.

In order to realize a virtually operated production, it is crucial that assets allow for a bidirectional information flow between physical and virtual world. Thus, assets need DTs that suit the definition of Kritzinger [18]. The two components of the DTs, i.e. data and behavioral models, need to be separated to realize abstraction.

Abstraction refers in this context to the idea that interactions with the production are separated from the production's technical infrastructure. Thus, abstraction implements the concept of separation of concerns [7]. Abstraction can be achieved by a reference model that contains references to the data of the DT and the real production resources in a standardized form [6].

We propose a service-oriented, modular architecture for planning and operation of production that allows to realize the SDM paradigm with virtualization and abstraction.

The architecture of an individual asset, as depicted in Fig. 1, contains three distinct layers: an application layer, reference layer and infrastructure layer. The application
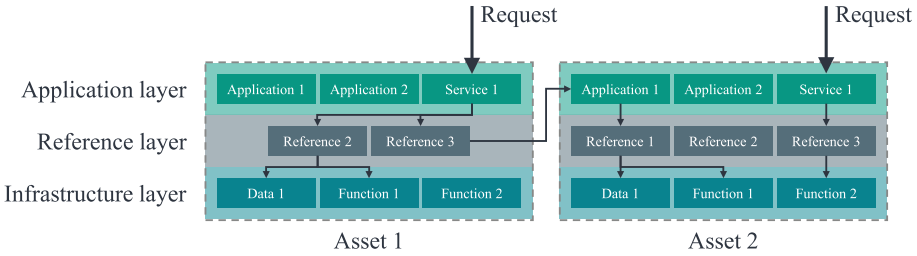
Fig. 1. Visualization of the proposed service-oriented architecture with the possible connections between application, reference and infrastructure layers of two exemplary assets.

layer provides services and applications for operation and planning of the asset, such as a service to produce a product or a simulation model of the asset as an application. The infrastructure layer contains all data to describe the asset and functions that can be performed by the asset. Application layer and infrastructure layer are connected through the reference layer. A reference is thereby a connection of a standardized key to infrastructure components that relate to this key. For example, the key *processing time* of a machine allows a standardized reference to all processing time data values of this machine. The reference model defines thereby the keys of each asset. References can either be requests for data or requests to perform a function. Besides references to infrastructure components, references to the application layer of other assets are also possible. This allows the creation of an architecture where hierarchical levels of the production can be represented, as shown in Fig. 2. The advantage of this architecture is that the interaction of humans or machines with the production is limited to the application layer and the technical infrastructure is not necessary to be considered. Moreover, production layers are decoupled so that interactions on a certain production level do not directly depend on other production levels. This approach makes it possible, for example, to trigger successive simulations from the upper levels to the levels below, thus extending the functionality of communication protocols such as OPC UA. Moreover, it is possible to use different communication protocols concurrently.
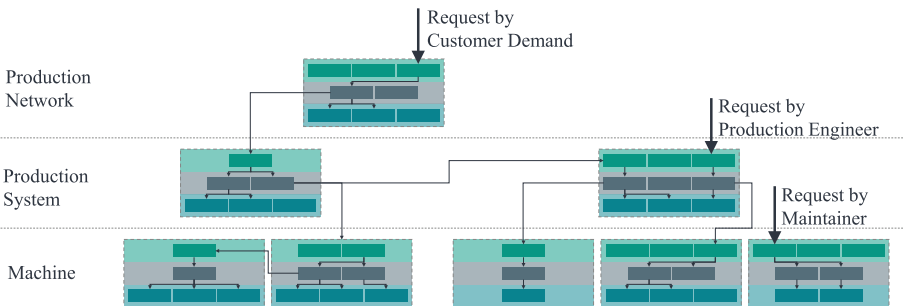
Fig. 2. Visualization of the proposed service-oriented architecture in context of the three production levels machine, production system and production network with three exemplary requests

Subsequently, the specific characteristics of the production levels machine, production system and production network are described with respect to the SDM architecture.

3.1 Machine

Due to a continuous standardization of interfaces and information models, the connection of new machines to the SDM architecture will become increasingly easy. As shown in [28] by a survey of representative companies in the mechanical and plant engineering sector, almost all companies work with machines and machine controls from at least two manufacturers and 81% of the machines in operation are older than six years. Therefore, brownfield machines with varying standards, interfaces and ages must be considered for a consistent integration in production systems. Since these machines have an advanced degree of wear and tear, the machine life cycle plays an important role. For connection of these machines, programmable logic controller (PLC) data, as well as additionally retrofitted sensors, must be considered. These sources can be accessed via various edge devices and nodes leading to an inhomogeneous data landscape. Multiple protocols and data formats, as mentioned in Sect. 2, must be considered. To deal with this complexity and, thus, be able to introduce retrofitted brownfield machines to SDM, relevant data sources should be made accessible to the reference layer with the help of a data-grabber [29]. Furthermore, to add the data sources to an information model, their signal types must be determined. This can be done by intelligent signal identification algorithms [30], which can be integrated into crawler tools for production communication networks. Furthermore, it can be assumed that there is no axis model available for brownfield machines, which is required to represent the machine using a DT. Based on this, a holistic integration of brownfield machines only unfolds its full potential when the system parameters of all machine axes are known. Therefore, reference runs with high information content as presented in [31] can be used to generate data for the determination of axis components. If both new and existing machines have been enabled for integration into the SDM architecture, they are available to host services and applications that utilize the machine's infrastructure layer.

3.2 Production System

Operating a production system efficiently demands a high effort in planning due to the complexity of the systems [32]. This is caused by the necessary consideration of technical details of the used production resources and by changing requirements to the production system over its life cycle [3, 24]. With SDM and the presented architecture, these limitations can be overcome.

Since the SDM architecture allows for abstraction, production resources can be treated as black boxes with certain capabilities offered by services. Thereby, technical details of these resources can be neglected. This simplifies the management of the resources in a production system as planners only need to consider a subset of information.

Pairing abstraction with virtualization simplifies the production planning task for production systems even more. The DTs of production systems, mostly modelled as material

flow simulations [18, 20], allow for detailed forecasting, optimization and autonomous control. The reference layer simplifies the generation of DTs as the access to required data is improved by standardized interfaces. Moreover, data analysis services can be combined with real time data sources, as e.g. MES systems, by the reference layer to synchronize DT and real production system.

The implementation of this approach into existing production systems requires three steps. At first, the requirements for virtualization of the production system and its resources need to be satisfied. Hence, the assets need to be digitized to gather data and to allow for virtual control. Next, the reference model can be created with references to the data and the production resources. At last, services and applications can be built upon the reference layer by using existing architectures such as REST. Possible services for a production system are the production of products or the analysis of the production system. Moreover, applications could range from simulation tools to data dashboards presenting the performance of the production system.

3.3   Production Network

By using the SDM paradigm also across companies in the global production network, new opportunities for interaction and collaboration in production networks as well as the design of production networks become possible. Production networks are systems that usually have grown over years and combine different locations and production systems [33]. For this reason, the focus on the global production network level is on brownfield planning.

The starting point for using SDM on the production network level is as well a reference model that provides the necessary information and enables abstraction and virtualization. Accordingly, the first requirement for SDM from the production network perspective is a unique reference between the reference models at the different levels, given by the reference layer. While the various sites, systems and machines represent internal company information, the production network level also requires information on suppliers, customers, service providers, etc. In the future, this data can also be provided via information models with an external interface. If suppliers do not yet provide information models, the data must be collected from the company's own databases in a transitional phase. Here, ERP and MES systems can be used. This allows, for example, to import customer and order data into the information model to perform order allocation or aggregate production planning tasks. In summary, the application layer for the production network receives information from the underlying levels (linked via the reference layer), information models of external parties and the company's own systems (ERP, MES, etc.). Since production networks are complex structures with high variety, the reference layer needs a detailed logic which can be modeled by asset administration shells and enhanced semantically through ontologies [16].

The infrastructure layer is contrasted with the application layer. Like the production system, the production network also focuses on a DT. However, this requires the right level of abstraction to be able to fulfill necessary tasks, on the one hand, and to keep the computing time and the effort required as low as possible, on the other hand. Since the use cases can be very diverse, it is important to define as precisely as possible which

information is required for each individual application. If there are continuous reference models across all levels, only the corresponding one must be referenced.

4   Exemplary Use Case

In the following a specific use case for the service-oriented architecture in the context of the three production layers is introduced. In detail, the architecture aims to realize a changeable production environment. This goal is enabled by the architecture because it can facilitate the planning and operation of today's digitalized production.

The use case comprises of the introduction of a new product variant in an existing production which is a common but still challenging scenario in production. Parallel to the introduction, the quantity of requested products by the customers changes as well. To react to these drivers of change, the existing production needs to be optimized and reconfigured for the new product und quantities. The use case shows how the service-oriented architecture is used in production planning on different levels of production.

First, a service is used on the production network level to find the optimal locations to produce the new product variant. To do so, historic production data of the different locations and production systems is needed. With the architecture, the data can be quickly gathered from information models. Heterogeneous data standards at different sites (identification, addresses, format, etc.) can be used in a unified service via the reference layer, thus supporting interoperability. The data is then analyzed and compared to the production capabilities needed to produce the new product variant. Out of this, planning algorithms can find a suitable match of production plant capabilities and product requirements and recommend it to the user. Allocation planning can thus be automated by using data directly from the information models.

Thereafter, production planning on a production system level needs to be performed. At first, material flow simulations are performed to analyze the performance of the production system under the new circumstances. The proposed architecture allows to create a service to retrieve a data-based description of the production system from the information model and automatically create a simulation model based on this data. As the structure of the obtained data is standardized by the reference layer, this service is applicable for different production systems and allows to reduce simulation modeling time dramatically. Similar services are also possible for other production planning tasks, such as scheduling of orders or reconfiguration planning. If the result of the simulation shows an acceptable performance, the architecture can be used to request the production system to produce the new product variant in the demanded quantity. The reference layer should link this request to the ERP system of the production to automatically trigger the production process. Again, this request could be standardized to allow the connection of different ERP systems to simplify the interaction of production planners with the IT infrastructure of the production.

If the production of the new variant requires a changeover of some machines, the architecture can also be used to support this process. A service can be imagined, that allows to request a changeover of a machine. Depending on the capability of the machine for automated changeover, the service either requests maintenance personal to perform the changeover or directly communicates the needed changeover to the machine via

suitable communication protocols. Ideally, the interface of this service is standardized for all machines and the reference layer distinguishes the associated requests to suit the needs of the machine.

The use case illustrates how the architecture can be used for production planning and operation on different levels of production. The possibility to create standardized services and requests with the architecture does not only simplify but also reduces the required time for production planning as these services can be reused in other scenarios. Moreover, interoperability of different tools, such as simulation or data analysis, is enhanced since it enables a standardized representation of a consistent data base.

5 Discussion and Conclusion

The presented SDM approach promises various advantages and opportunities by extending the SDM concept with consideration of the production life cycle on different levels of production. By decoupling the physical production and associated software with abstraction, a clear application layer can be implemented leading to simplified supply chain and production planning. In turn, this helps to reduce the complexity in production planning and control and, thus, promises to increase the efficiency in production. Consistent matching of product requirements and currently available production capabilities can be achieved even before the production of initial samples. Furthermore, the architecture promotes the use of virtualization that also reduces the required time for production analyses and forecasts. Lastly, interoperability in planning and operation of production is enabled by services and the use of a reference layer.

However, the presented SDM architecture and SDM in general still face various challenges in the implementation. For example, depending on the requested services, there are specific minimum requirements for the required data quality. Thus, standardization and digitization of production assets is required. In order to be able to map the entire global production network, end-to-end networking between all layers and models must be ensured. Furthermore, isolated use of individual models must be made possible at every level. For this purpose, it has to be possible to abstract individual layers so that individual models can be used. This should be possible, even if there is no reference to one of the other layers. In summary, the management of the reference layer must be simple to avoid additional administrative work. Moreover, existing implementations need to be extended to allow an integration into a unifying architecture. Future research shall implement the introduced architecture with various layers and demonstrate its applicability to address the presented challenges.

References

1. Bennett, N., Lemoine, G.J.: What a difference a word makes: understanding threats to performance in a VUCA world. Bus. Horiz. **57**(3), 311–317 (2014)
2. Mack, O., Khare, A., Kramer, A.: Managing in a VUCA World, 1st edn. Springer-Verlag, s.l. (2015)
3. ElMaraghy, H., Wiendahl, H.-P.: Changeable manufacturing. In: Laperrière, L., Reinhart, G., Chatti, S., Tolio, T. (eds.) CIRP Encyclopedia of Production Engineering, pp. 157–163. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-642-20617-7_6674
4. Wee, D., Cattel, J., Breunig, M.: Industry 4.0 How to navigate digitization of the manufacturing sector (2015)
5. Lechler, A., Verl, A.: Software defined manufacturing extends cloud-based control. In: Proceedings of the ASME 12th International Manufacturing Science and Engineering Conference. The American Society of Mechanical Engineers, New York (2017)
6. Neubauer, M., Ellwein, C., Frick, F.: Kontinuität als neues Paradigma. Computer Automation 2022 (2022)
7. Thames, L., Schaefer, D.: Software-defined cloud manufacturing for industry 4.0. Procedia CIRP **52**, 12–17 (2016)
8. Wulf, S., Redlich, T., Krenz, P., Wulfsberg, J.P.: Die Entwicklung der Offenheit im Lebenszyklus von Produktionsnetzwerken. Zeitschrift für wirtschaftlichen Fabrikbetrieb **108**(12), 911–916 (2013)
9. Lecher, A., Kircher, C., Verl, A.: SDM - Software Defined Manufacturing. wt Werkstattstechnik online (5) (2021)
10. Kreutz, D., Ramos, F.M.V., Esteves Verissimo, P., Esteve Rothenberg, C., Azodolmolky, S., Uhlig, S.: Software-defined networking: a comprehensive survey. Proc. IEEE **103**(1), 14–76 (2015)
11. Lopez, F., Shao, Y., Mao, Z.M., Moyne, J., Barton, K., Tilbury, D.: A software-defined framework for the integrated management of smart manufacturing systems. Manuf. Lett. **15**, 18–21 (2018)
12. Nayak, N.G., Durr, F., Rothermel, K.: Software-defined environment for reconfigurable manufacturing systems. In: 2015 5th International Conference on the Internet of Things (IOT 2015). Seoul, South Korea, 26–28 October 2015, pp. 122–129. IEEE, Piscataway (2015)
13. Baheti, R., Gill, H.: Cyber-physical systems. Impact Control Technol. **12**(1), 161–166 (2011)
14. Leitner, S.-H., Mahnke, W.: OPC UA-service-oriented architecture for industrial applications. ABB Corp. Res. Center **48**(61–66), 22 (2006)
15. Ewert, D., Jung, T., Tasci, T., Stiedl, T.: Assets2036 – lightweight implementation of the asset administration shell concept for practical use and easy adaptation. In: Weißgraeber, P., Heieck, F., Ackermann, C. (eds.) Advances in Automotive Production Technology – Theory and Application. A, pp. 153–161. Springer, Heidelberg (2021). https://doi.org/10.1007/978-3-662-62962-8_18
16. Bader, S.R., Maleshkova, M.: The semantic asset administration shell. In: Acosta, M., Cudré-Mauroux, P., Maleshkova, M., Pellegrini, T., Sack, H., Sure-Vetter, Y. (eds.) SEMANTiCS 2019. LNCS, vol. 11702, pp. 159–174. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-33220-4_12
17. Tantik, E., Anderl, R.: Integrated data model and structure for the asset administration shell in industrie 4.0. Procedia CIRP **60**, 86–91 (2017)
18. Kritzinger, W., Karner, M., Traar, G., Henjes, J., Sihn, W.: Digital Twin in manufacturing: a categorical literature review and classification. IFAC-PapersOnLine **51**(11), 1016–1022 (2018)

19. Negri, E., Fumagalli, L., Macchi, M.: A review of the roles of digital twin in CPS-based production systems. Procedia Manuf. **11**, 939–948 (2017)
20. Benfer, M., Peukert, S., Lanza, G.: A framework for digital twins for production network management. Procedia CIRP **104**, 1269–1274 (2021)
21. Stark, R., Kind, S., Neumeyer, S.: Innovations in digital modelling for next generation manufacturing system design. CIRP Ann. **66**(1), 169–172 (2017)
22. Monostori, L.: Cyber-physical production systems: roots expectations and R&D challenges. Procedia CIRP **17**, 9–13 (2014)
23. Ueda, K., Vaario, J., Ohkura, K.: Modelling of biological manufacturing systems for dynamic reconfiguration. CIRP Ann. **46**(1), 343–346 (1997)
24. ElMaraghy, H.A.: Flexible and reconfigurable manufacturing systems paradigms. Int. J. Flex Manuf. Syst. **17**(4), 261–276 (2005)
25. Tharumarajah, A.: Comparison of the bionic, fractal and holonic manufacturing system concepts. Int. J. Comput. Integr. Manuf. **9**(3), 217–226 (1996)
26. Babiceanu, R.F., Chen, F.F.: Development and applications of holonic manufacturing systems: a survey. J. Intell. Manuf. **17**(1), 111–131 (2006)
27. Babiceanu, R.F., Seker, R.: Big Data and virtualization for manufacturing cyber-physical systems: a survey of the current status and future outlook. Comput. Ind. **81**, 128–137 (2016)
28. Netzer, M., Begemann, E., Gönnheimer, P., Fleischer, J.: Study on digitalization in the german mechanical engineering sector. wt Werkstattstechnik online **111**(07–08), 526–530 (2021)
29. Hillenbrand, J., Gönnheimer, P., Gerlitz, E., Fleischer, J.: Design and implementation of a holistic framework for data integration in industrial machine and sensor networks. Procedia CIRP **104**, 1771–1776 (2021)
30. Gönnheimer, P., Karle, A., Mohr, L., Fleischer, J.: Comprehensive machine data acquisition through intelligent parameter identification and assignment. Procedia CIRP **104**, 720–725 (2021)
31. Gönnheimer, P., Ströbel, R., Netzer, M., Fleischer, J.: Generation of identifiable CNC reference runs with high information content for machine learning and analytic approaches to parameter identification. Procedia CIRP **107**, 734–739 (2022)
32. Wiendahl, H.-P., Scholtissek, P.: Management and control of complexity in manufacturing. CIRP Ann. **43**(2), 533–540 (1994)
33. Lanza, G., et al.: Global production networks: design and operation. CIRP Ann. **68**(2), 823–841 (2019)

Dynamic Safety Distance Determination for Human Robot Coexistence in Industrial Applications

Marc Fischer$^{(\boxtimes)}$, Lars Klingel, Armin Lechler, Alexander Verl, and Michael Neubauer

Institute for Control Engineering of Machine Tools and Manufacturing Units, University of Stuttgart, 70174 Stuttgart, Germany
`marc.fischer@isw.uni-stuttgart.de`

Abstract. The coexistence of humans and robots in manufacturing requires safety. Typically, safety functionalities for coexistence are based on speed and separation monitoring according to EN ISO 10218. Thereby, a robot should reduce speed or stop completely when an obstacle is too close to avoid collisions. Nowadays, despite this standard, speed and separation monitoring are still realized with static worst-case safety assumptions about the kinematics of humans and robots. Over the years, different static techniques have evolved in the industrial environment like fences, light fences, or camera-based systems where critical zones can be configured. The latter is admittedly more flexible in the configuration but still static during the run-time of the manufacturing system. The static worst-case assumptions result in large unnecessary distances between humans and robots, which leads to inefficient use of space. Therefore, factories are larger than they need to be, which leads to higher costs. This work aims to enhance the use of space in human-robot coexistence applications to make factory layouts more efficient. Therefore, a dynamic minimum distance calculation based on the DIN ISO 15066 with the kinematic information of the coexisting human and industrial robots is provided. It is shown in a simulation that this approach of a dynamic safety distance calculation leads to a reduction of the required space.

Keywords: Safety · Human-Robot Collaboration · Distance Monitoring

1 Introduction

Human-robot collaboration (HRC) is an important research topic in the current aim to use autonomous and adaptive manufacturing due to its flexible application possibilities [16,24]. A main disadvantage of traditional industrial robotic systems is the necessity of fences and static safety configurations with high space consumption, which makes the integration in adaptive and flexible production systems challenging [16]. The usage of collaborative robots is not limited to the direct interaction of human and robots, it also includes flexible safety functions,

easy relocation, and less space consumption [24]. HRC requires safety to be applicable in industrial environments. The related work presents different safety methods for different levels of human-robot interaction. The level of interaction is often differentiated into coexistence, cooperation, and collaboration [3, 16, 24]. In a coexistence, humans and robots share the same environment but do not interact directly. In a cooperation, humans and robots work in the same workspace but on different tasks, and in a collaboration, the human and robot execute a task together. The safety methods are differentiated in collision avoidance and collision detection [20, 23, 27] or similar categories [28, 30]. Collision detection is required for collaboration when humans and robots interact directly or share the same workspace for cooperation because of the possible contact between a workpiece and a human. Thereby the impact of an injury through a collision must be limited. Collision avoidance is sufficient for coexistence because no direct contact between humans and robots is required and demanded. Besides the research, standards exist for handling safety in HRC. The EN ISO 10218 [13] presents three safety methods for collaborative interaction: Hand-guided control, speed and separation monitoring (SSM) as well as power and force limiting.

Collaborative robots like Bosch APAS, Universal Robot, Kuka LBR iiwa and Franka Emika are designed for direct interactions between humans and robots. Therefore, safety is gained by limiting the injury while colliding with lightweight structures or sensorized skins. However, the payload is typically limited to less than 20 kg due to the lightweight structures. Thus, these kinds of robots are flexible in safety but limited to specific use cases with low payload requirements. According to the number of robot installations in 2020, the market share of collaborative robots is small with five percent [1]. Using traditional industrial robotic systems for human-robot collaborations could enhance the market share of collaborative robots because their limitation to specific use cases could be revoked. Thereby, the most common applications of industrial robots are handling and welding [1] which do not require direct interaction between humans and robots. Instead, a safe coexistence between humans and robots is sufficient. New safety functions are required to enable adaptive and flexible application of traditional industrial robotic systems in manufacturing during the coexistence with humans.

To the best of the authors knowledge, there are only a few products on the market which try to make traditional robotic systems' safety more flexible. The SafetyEye from Pilz uses a camera-based vision system where static danger zones can be defined. No fences are required, but the flexibility is limited to the configuration time [18]. The system enables SIL3 solutions. Another product is the INXPECT Radar sensor, targeting static danger zones with SIL2 or PLd. These solutions are more flexible during configuration but still require much space. Therefore, this paper presents a flexible safety method based on speed and separation monitoring. In contrast to the other products and research presented in Sect. 2, this new approach monitors the position and speed of humans and robots to calculate the required separation. The distance is lower compared to worst-case assumptions where no speed or position is monitored. The monitoring

of robots and humans enables a more efficient space usage which is shown in a simulation in 4. Furthermore, the boundary conditions for the design and implementation of such a system is shown in 5.

2   Related Work

The problem of safety in human-robot coexistence has been addressed for many years. In contrast to the low number of certified products on the market, different approaches can be found in research, which are based on different sensors and safety functions. Various literature reviews give an overview of the different approaches. Eight reviews have been identified [3,12,20,22,23,27,28,30] by searching with `((human AND robot) OR (human-robot)) AND (safety OR safe) AND (review OR survey))` in Web of Science[1] and Scopus[2]. The reviews state different approaches for collision avoidance like estimating the intention of a human to modify the robot movement [12,20]. Moreover, distance determination between humans and robots are used to change the trajectory of the robot, as in [9,10]. All papers related to SSM in these reviews have been analyzed. Within these approaches, many of them rely on the distance between the robot and obstacle like [29,31]. Others use additional sensors for tracking the human or obstacle like [5,7,11]. Many researches define static zones around the robot, like [5,25]. Others take the velocities into account, like [9] who presents collision avoidance by changing the trajectory of the robot with control barrier functions that include the position and velocity of the human.

Four publications are similar to this work's approach. [2] track humans with a Kinect V2 3D-RGB camera and calculate the required safety distance based on TS/ISO 15066. They only monitor the tool center point of the robot instead of a more complex model. By simulation, they show fewer safety stops compared to traditional safety systems. [15] want to enhance the productivity of collaborative tasks by minimizing the degraded state of the robot due to the safety functions. They modify the standard SSM by calculating a safety threshold in real-time, based on the position and velocity of humans and robots. However their focus lays on cooperation and collaboration. Furthermore they present a risk classification based on the SSM. [17] track the movement of humans and the robot and transfer them into a physical simulation. The humans are tracked with multiple Kinect cameras. Spheres extend the joint model. The robot's position and future trajectory are tracked too. The simulation software checks a collision based on the current human position and future robot trajectory. Possible movements of humans are not considered. The main contribution of [17] is the camera-based human motion tracking. In [26] the dynamic calculation of SSM based on the tracking of humans and the robot is analyzed. They present a so called trajectory-dependent safety distance, which refers to the robots trajectory. Details on the human modeling are not given. They show less required safety distance when the robot moves away from an obstacle. All four works do not focus

---

[1] https://www.webofscience.com/.
[2] https://www.scopus.com.

on the reeducation of required space. Furthermore [17] and [26] use lightweight robots. Only in [15] with the ABB IRB1200 a industrial robot is used, but the IRB1200 is still a lightweight robot which is limited to a low payload.

3   Accurate Speed and Separation Monitoring

The SSM increases safety for HRC by monitoring the separation of humans and robots based on the position and speed of both, humans and robots. This section explains the SSM according to the norm and introduces the modification of this work's approach where monitored values over worst-case assumptions are used. This can lower the required safe space.

3.1   Calculation According to the Norm

The EN ISO 10218 names SSM as a possible safety function to enable HRC but does not detail the implementation of such a safety function. However, the ISO/TS 15066:2016 [14] is a technical specification that guides the implementation of SSM. The required separation distance according to the ISO/TS 15066:2016 sec. 5.5.4.2.3 is (1).

$$S(t) = S_h(t) + S_r(t) + S_s(t) + C + Z_d + Z_r \tag{1}$$

The functions and variables are as follows [14]:

– S_h is the contribution to the protective separation distance attributable to the operator's change in location
– S_r is the contribution to the protective separation distance attributable to the robot system's reaction time
– S_s is the contribution to the protective separation distance due to the robot system's stopping distance
– C is the intrusion distance, as defined in ISO 13855; this is the distance that a part of the body can intrude into the sensing field before it is detected
– Z_d is the position uncertainty of the operator in the collaborative workspace, as measured by the presence sensing device resulting from the sensing system measurement tolerance
– Z_r is the position uncertainty of the robot system, resulting from the accuracy of the robot position measurement system.

The separation distance S must be calculated by considering all human and robot parts. The technical specification leaves this open to the user. Therefore a model for humans and robots is required first.

3.2   Human and Robot Modeling

Some approaches use two points for the hand of the human and the tool center point of the robot, like [8], which is not accurate enough as other parts of human and robot could be closer. Other approaches like [4] are based on geometrical models, but the formula can not be used with them. Others use point clouds generated from the sensor's output like [7]. Others apply a joint link model like [9] where humans and robots are modeled by joints and the links between the joints. As shown in Fig. 1, this work also uses a joint model extended with spheres around the joints. This model is a good compromise between accuracy and calculation effort.
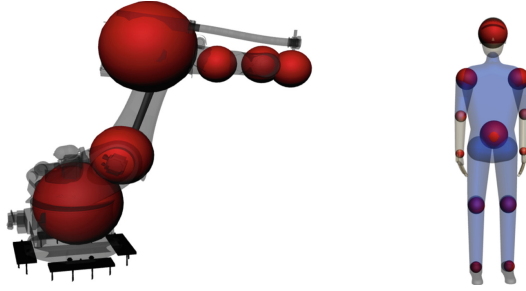
Fig. 1. Joint model of the human and robot

With the joint model, the separation distance as the minimum of the distances between each link of the robot j_j and each link of the human j_i by $S = min(S_{i,j})$ can be calculated.

3.3   Implementation Details and Monitoring Differences

The contribution of the variables to the required separation distance is based on the possibility of monitoring the operator and robot. If no monitoring is possible, a worst-case assumption must be made.

S_h is expressed in [14] as $S_h(t) = \int_{t_0}^{t_0+T_r+T_s} v_h(t)dt$ where $v_h(t)$ is the velocity of the human in the direction of the robot. T_r is the reaction time of the whole system including the time for detecting the human, processing the signals, and activating the robot stop. A static time is assumed, due to the real-time requirement with deterministic behavior for safety functions. The time depends on the used sensors, algorithms, hardware and robot and must be identified empirically. T_s is the time for stopping the robot. The stopping time of the robot relies on the velocity, pose and payload of the robot and is identified empirically by the robot manufacturer. If the velocity of the human can be monitored the velocity is only known for the current time t_o. Some approaches [12] try to estimate the intention of the human operator which could be used to make a

estimation of the future velocity v_h. In this work a more conservative method by assuming a maximum acceleration additional to the the current speed so that $S_h(t) = \int_{t_0}^{t_0+T_r+T_s} v_h(t_0) + \int_{t_0}^{t_0+T_r+T_s} a_{h,max} dt$ whereby v_h is limited by $|v_h(t)| < v_{h,max}$ is used. If the velocity can't be monitored the worst case of $v_h = v_{h,max} = 1.6\frac{m}{s}$ must be assumed according to ISO/TS 15066:2016.

The required separation distance attributable to the robot reaction time is $S_r = \int_{t_0}^{t_0+T_r} v_r(t)$ according to [14]. If the robot's velocity can be monitored, a potential acceleration must be considered according to the technical specification. In contrast, if the trajectory is known and the position and velocity are monitored, the future position and velocity can be determined based on the dynamics calculation of the robot control. If the velocity can't be monitored the maximum $v_r = v_{r.max}$ must be assumed.

The required separation distance attributable to the robot stop time is $S_s = \int_{t_0+T_r}^{t_0+T_r+T_s} v_r(t)$ according to [14]. The stopping distance is measured by the robot manufacturer for each axis and relies on the current velocity, payload and pose. With these distances the robot pose for $t = t_0 + T_r + T_s$ can be determined.

The remaining variables C, Z_d, and Z_r depend on the chosen sensors and are provided by the sensor manufacturer or must be determined empirically.

4   Simulation

To validate the potential of this work's model-based SSM approach, a coexistence scenario is simulated which lays the foundation for a comparison to existing approaches, namely a light fence, SafetEye, and the sole monitoring of the robot. The SSM is used for each approach and determines the required separation distance.

4.1   Scenario

Industrial robots occur in various industrial applications. Nevertheless, they are most often used for handling tasks [1]. This is why the benefits of this work's approach is shown by the example of such a robot handling task. This work contributes to the safe coexistence of heavy-weight industrial robots and humans in highly automated plants. Therefore a scenario is shown where a human walks by an industrial robot, which performs a handling task. A KUKA KR500 R2830 is chosen, which performs a pick and place task. A visualization[3] of this simulation scenario is shown in Fig. 2.

The simulation is carried out in a virtual commissioning tool, where the kinematics of the robot and the human are modeled. A virtual CNC is used and integrated into the virtual commissioning tool to generate a realistic trajectory of the robot. The safety calculations of the different approaches are described in the next section.

---

[3] A video of the simulated scenario can be found at https://youtu.be/ErpZITy9dUw.
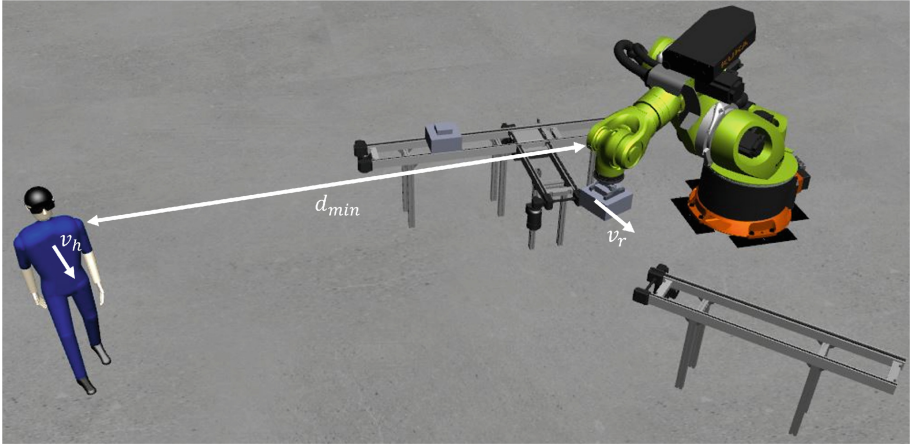
Fig. 2. Simulation scenario

4.2   Implementation Details

The parameters for the different safety approaches are identified from the manufacture manuals or they are estimated. The values C and T_r are different for each approach.

– For the simulation of a **light fence** a Pilz PSEN_opII4H [19] with a Beckhoff TwinSafe [6] safety PLC and an EtherCAT fieldbus is used. According to the DIN EN 13855 the intrusion distance is calculated to $C = 8(d - 14)$ with d in mm whereby in our case $d = 30$mm so that $C = 0.128$. The reaction time T_r is calculated as $T_r = t_{\text{sensor}} + t_{\text{io}} + t_{\text{bus}} + t_{\text{logic}} + t_{\text{bus}} + t_{\text{io}} = 0.0393$s with the values from [6,19].
– The values for the **SafetyEye** simulation are provided in [18] with $C = 0.208$ and $T_r = 0.3$.
– For the **sole monitoring of the robot** and the **model-based SSM** $C = 0.1$ and $T_r = 0.3$ based on values presented in [21] are assumed.

The position uncertainty of robot and human are estimated to $Z_d = 0.1$ and $Z_r = 0.001$. The maximum velocity of a human is set to $v_h = v_{h,max} = 1.6\frac{\text{m}}{\text{s}}$ as suggested by [14]. The maximum velocity of the robot $v_{r,max}$ could be set to the physical limits but depending on the application these values are never reached. Therefore, the maximum velocity occurring during the application satisfy the worst case assumption.

4.3   Evaluation

The results[4] of the simulation are shown in Fig. 3. The minimum occurring separation distance $d_{min} = min(d_{i,j})$ is the minimum euclidean distance between

---

[4] The source code of the calculation can be found at https://github.com/iswunistuttgart/robotsafety-dynssm.

each link of the robot and the human. If the required separation distance is lower than d_{min} the use case can be considered as safe. The required separation distance for the light S_{lf} and the SafetyEye S_{se} is constant because the worst-case assumption is used for the human and the robot motion. Due to the higher values of the constants C and T_r for the SafetyEye approach, a higher distance between human and robot is required. Therefore, high sensor accuracy and low reaction time are key elements for realizing small separation distances.
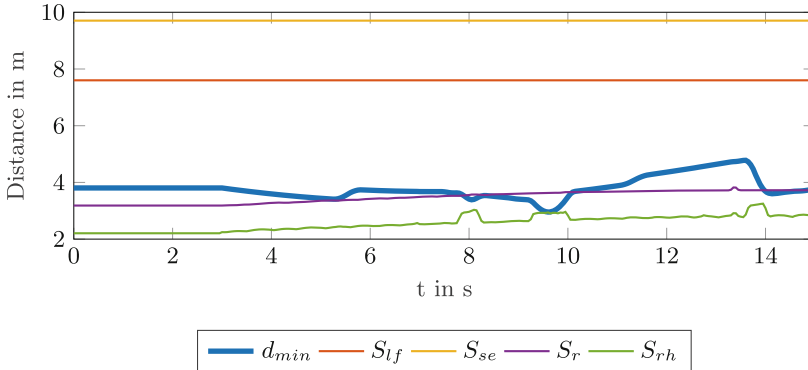
Fig. 3. Comparison of different safety methods

In contrast to the constant values, the sole monitoring of the robot S_r and the monitoring of robot and human S_{rh} is variable due to the monitoring of the actual motion of the human and robot. Monitoring the robot motion halves the required separation distance compared to the light fence. The reduction through the monitoring of the human is smaller but still the smallest required separation distances because the velocities of the human and robot are lower compared to the worst-case assumptions. Thus, saving space and minimizing safety distances requires monitoring humans and robots.

5 Requirements and Safety Discussion

The main problem with complex safety functions for HRC is the safe perception of the human and the robot [26]. In related work, different sensors are used and combined, but the certification of the concepts is not considered. In other areas like autonomous driving, sensor fusion is a recommended way to perceive complex environments. Therefore an accurate examination of possible sensors and model extraction must be done. The authors propose a redundant but diverse perception with different sensors and algorithms to gain reliable monitoring of human movement. Furthermore, the robot's motion must be determined safely. A common approach is to use a sensor before and after the drive so that the positional differences of the mechanic can be detected. Most robots have only one

sensor at the drive. As shown in the simulation, the processing time is also a key influence in the calculation. The SafetyEye requires a higher separation distance than the light fence due to its higher processing time. Therefore, the sensors and algorithms should have low latencies and deterministic timing behavior.

6   Conclusion and Future Prospects

In this work it was shown how human and robot modeling combined with a dynamic safety calculation based on a SSM can help to safely overcome conservative distance estimations. Based on these results, the complete approach by combining two appropriate sensors for the perception of humans should be developed. The combination of such a human tracking and the presented model-based SSM approach should be integrated into an industrial robotics environment to help reduce space and make production systems more flexible.

References

1. World robotics 2021. Technical report, International Federation of Robotics (2021)
2. Araiza-lllan, D., de San Bernabe Clemente, A.: Dynamic regions to enhance safety in human-robot interactions. In: IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA), pp. 693–698. IEEE (2018)
3. Arents, J., Abolins, V., Judvaitis, J., Vismanis, O., Oraby, A., Ozols, K.: Human-robot collaboration trends and safety aspects: a systematic review. J. Sens. Actuator Netw. **10**(3), 48 (2021)
4. Balan, L., Bone, G.: Real-time 3D collision avoidance method for safe human and robot coexistence. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 276–282. IEEE (2006)
5. Bdiwi, M., Pfeifer, M., Sterzing, A.: A new strategy for ensuring human safety during various levels of interaction with industrial robots. CIRP Ann. **66**(1), 453–456 (2017)
6. Beckhoff Automation GmbH & Co. KG: Operation Manual EL6910 TwinSAFE Logic Terminal (2020)
7. de Gea Fernández, J., et al.: Multimodal sensor-based whole-body control for human-robot collaboration in industrial settings. Robot. Auton. Syst. **94**, 102–119 (2017)
8. Ferraguti, F., Bertuletti, M., Landi, C.T., Bonfe, M., Fantuzzi, C., Secchi, C.: A control barrier function approach for maximizing performance while fulfilling to ISO/TS 15066 regulations. IEEE Robot. Autom. Lett. **5**(4), 5921–5928 (2020)
9. Ferraguti, F., et al.: Safety barrier functions and multi-camera tracking for human-robot shared environment. Robot. Auton. Syst. **124**, 103388 (2020)
10. Flacco, F., Kroger, T., de Luca, A., Khatib, O.: A depth space approach to human-robot collision avoidance (2012)

11. de Gea Fernández, J., et al.: iMRK: demonstrator for intelligent and intuitive human-robot collaboration in industrial manufacturing. KI - Künstl Intell **31**(2), 203–207 (2017)

12. Halme, R.J., Lanz, M., Kämäräinen, J., Pieters, R., Latokartano, J., Hietanen, A.: Review of vision-based safety systems for human-robot collaboration. Procedia CIRP **72**, 111–116 (2018)

13. ISO: 10218-1:2011, Robots and robotic devices - Safety requirements for industrial robot - Part 1: Robots (2011)

14. ISO/TS: 15066:2016, Robots and robotic devices: Collaborative robots (2016)

15. Liu, Z., et al.: Dynamic risk assessment and active response strategy for industrial human-robot collaboration. Comput. Industr. Eng. **141**, 106302 (2020)

16. Matheson, E., Minto, R., Zampieri, E.G.G., Faccio, M., Rosati, G.: Human-robot collaboration in manufacturing applications: a review. Robotics **8**(4), 100 (2019)

17. Morato, C., Kaipa, K.N., Zhao, B., Gupta, S.K.: Toward safe human robot collaboration by using multiple Kinects based real-time human tracking. J. Comput. Inf. Sci. Eng. **14**(1) (2014)

18. Pilz GmbH & Co. KG: SafetyEYE Bedienungsanleitung-21743-DE-22 (2018)

19. Pilz GmbH & Co. KG: PSEN opII4H-Serie Bedienungsanleitung-1003501-DE-08 (2019)

20. Proia, S., Carli, R., Cavone, G., Dotoli, M.: Control techniques for safe, ergonomic, and efficient human-robot collaboration in the digital industry: a survey. IEEE Trans. Autom. Sci. Eng. **19**, 1–22 (2021)

21. Ramer, C.: Arbeitsraumüberwachung und autonome Bahnplanung für ein sicheres und flexibles Roboter-Assistenzsystem in der Fertigung. Dissertation, FAU, Erlangen (2018)

22. Reddy, A., Bright, G., Padayachee, J.: A review of safety methods for human-robot collaboration and a proposed novel approach. In: Gusikhin, O., Madani, K., Zaytoon, J. (eds.) ICINCO 2019, pp. 243–248. SCITEPRESS - Science and Technology Publications Lda, Setúbal, Portugal (2019)

23. Robla-Gomez, S., Becerra, V.M., Llata, J.R., Gonzalez-Sarabia, E., Torre-Ferrero, C., Perez-Oria, J.: Working together: a review on safe human-robot collaboration in industrial environments. IEEE Access **5**, 26754–26773 (2017)

24. Rodriguez-Guerra, D., Sorrosal, G., Cabanes, I., Calleja, C.: human-robot interaction review: challenges and solutions for modern industrial environments. IEEE Access **9**, 108557–108578 (2021)

25. Tashtoush, T., et al.: Human-robot interaction and collaboration (HRI-c) utilizing top-view RGB-d camera system. Int. J. Adv. Comput. Sci. Appl. **12**(1) (2021)

26. Vicentini, F., Giussani, M., Tosatti, L.M.: Trajectory-dependent safe distances in human-robot interaction. In: Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA), pp. 1–4. IEEE (2014)

27. Villani, V., Pini, F., Leali, F., Secchi, C.: Survey on human-robot collaboration in industrial settings: safety, intuitive interfaces and applications. Mechatronics **55**, 248–266 (2018)

28. Wang, N., Zeng, Y., Geng, J.: A brief review on safety strategies of physical human-robot interaction. In: ITM Web of Conferences, vol. 25, pp. 01015 (2019)

29. Weitschat, R., Aschemann, H.: Safe and efficient human-robot collaboration part II: optimal generalized human-in-the-loop real-time motion generation. IEEE Robot. Autom. Lett. **3**(4), 3781–3788 (2018)

30. Zacharaki, A., Kostavelis, I., Gasteratos, A., Dokas, I.: Safety bounds in human robot interaction: a survey. Saf. Sci. **127**, 104667 (2020)

31. Zlatanski, M., Sommer, P., Zurfluh, F., Madonna, G.L.: Radar sensor for fenceless machine guarding and collaborative robotics. In: IEEE International Conference on Intelligence and Safety for Robotics (ISR), pp. 19–25. IEEE, Piscataway, NJ (2018)

Integrated Framework for Safety Management and Software-Assisted Safety Assessment in Fluid Production

Chee H. Koo[1], Bernd Neuschwander[2], Marian Vorderer[1], Patrick Stehle[2], Roman Kretzschmann[2], Timur Tasci[3], Urs Leberle[1(✉)], and Alexander Verl[3]

[1] Corporate Sector Research and Advanced Engineering, Robert Bosch GmbH, Robert-Bosch-Campus 1, 71272 Renningen, Germany
`Urs.Leberle@de.bosch.com`
[2] Pilz GmbH & Co. KG, Felix-Wankel-Straße 2, 73760 Ostfildern, Germany
[3] Institute for Control Engineering of Machine Tools and Manufacturing Units (ISW), Seidenstraße 36, 70174 Stuttgart, Germany

Abstract. In the factory of the future, the concept of fluid production allows production processes to be adapted more frequently and efficiently to fulfill rapidly-changing product requirements and to prevail in an increasingly challenging market environment. Production resources, machinery and humans are integrated seamlessly to achieve common production goals. Rapid changes in system configurations are also enabled by highly modular, self-descriptive and reusable machinery modules called Mechatronic Objects. The high reconfigurability in fluid production poses new challenges for the safety management of future production systems. Every modification done to an existing system needs to be risk assessed and approved during the commissioning to allow its safety-compliant operation. As current industrial practices for risk assessment are still labor-intensive and time-consuming, new methods are needed to quickly integrate the aforementioned highly modular Mechatronic Objects into fluid production systems. In this paper, we propose a safety management and assessment framework to address the safety-related challenges for fluid production. This paper proposes a new modeling method to describe the different types of Mechatronic Objects and derives different stages for the assessment and approval of fluid production systems. This allows a more efficient and accelerated commissioning phase, a seamlessly integrated production and the reduction of manual efforts needed for the system commissioning. The proposed framework, the methods and the implemented software tools are validated using a case study for a highly reconfigurable assembly system at the research campus ARENA2036 at the University of Stuttgart.

Keywords: fluid production · safety management · safety assessment · cyberphysical production system

1 Introduction

Future production in the automotive industry needs to be adapted more frequently and in an efficient manner to fulfill rapidly-changing product require-

ments in order to prevail in the increasingly challenging market environment. In this future production, production resources, machinery and humans are integrated seamlessly to achieve common production goals. If production changes are required, the system configuration can be adapted flexibly and dynamically to address new production requirements. This highly reconfigurable nature of the production systems is made possible thanks to the application of Plug-and-Produce (PnP) technology and the usage of Cyber Physical System (CPS) within the production. Relevant approaches and implementations to achieve this vision are currently studied within the research project FluPro (Fluid Production) at the research campus ARENA2036 of the University Stuttgart.

Safety assessment represents one of the many important aspects to guarantee production safety and to enable the successful operation of fluid production [1]. According to the Machinery Directive [2] in the European Union, every machinery or production system needs to be risk assessed and the results documented before the CE marking can be issued, which represents the manufacturer declaration for the system conformity according to related safety standards and regulations. The procedure for risk assessment according to ISO 12100 [8] has to be done after every system modification to ensure that possible emerging risks are identified, eliminated or reduced to an acceptable level. Current industrial practices depend mostly on manual efforts of highly-experienced safety engineers and the created document-based assessment results are hardly reusable. This time-consuming and labor-intensive safety-related procedures are counter-productive, especially in consideration of the main objectives of fluid production to allow an efficient and seamlessly integrated production.

In this paper, we propose a safety assessment framework called the Fluid Production Safety 4A-Framework (FluPro-S4A) to provide an assisted risk assessment and documentation process within fluid production. The goal of this framework is to facilitate the manual efforts needed to risk assess and approve the operational safety of production systems, which will lead to an accelerated system commissioning. This paper proposes a new modeling method to describe the different types of assets (i.e. production equipment and machinery) within fluid production and derive the integrated framework FluPro-S4A alongside relevant software tools to enable a semi-automated and assisted assessment during system commissioning.

2   Related Work

Tools or assistance systems with more advanced capabilities for risk assessment are desired among experts based on a study conducted by [7]. In the field of Human-Robot-Collaboration (HRC), some of the methods and tooling worth mentioning include the formal method SAFER-HRC for hazard analysis [4,15], the rule-based system [5], the simulation-based approach [6,16] and the robot reachability analysis [14,17]. The author in [9] focuses more on robot behaviour analysis and proposes an automated configuration method for different robot states. These mentioned methods deal mostly with concrete aspects within the HRC applications to facilitate the risk assessment procedure.

In the field of reconfigurable systems, similar approaches can also be found as frequent system changes might lead to an increased effort for risk assessment. The authors [12] provide a certification concept for modular production lines, whereas [13] proposes a digital certificate for the CE-conformity of I4.0 production lines. Toolings and methods to facilitate risk assessment in various domains such as manufacturing have similar motivation to assist humans during the assessment and decision-making process. Some of the examples include the concept AutoSafety for assisted risk assessment for adaptable production systems [10], the decision tree analysis method [11], the framework for the assessment of complex systems [3] or the runtime analysis of failure rates for automated guided vehicles [18].

The contributions in this paper lie mainly on the whole integrated process during the commissioning of fluid and reconfigurable production systems, where high-level modeling and safety approval procedures are proposed. With the methodical foundation provided by the aforementioned publications, we demonstrate how such computer-aided methods can be integrated seamlessly into production systems.

3   Modeling and Integrated Process for the System Commissioning

Considering the context within fluid production, several terminologies are firstly introduced to describe the hierarchy and the types of production assets:

– A *System* (S) is a collective term to describe the whole production layout that is made up of one or more production cells or machinery (*Base*) to achieve common production goals.
– A *Base* (B_S) represents a standalone production unit (cell or machinery) that is made up of modular production assets (*Mechatronic Object*) that offer distinct production skills. This represents the starting point for the system safety assessment introduced later in this paper.
– A *Mechatronic Object* ($MO_{S,B}$) describes a production asset that is integrated into the *Base*. Two categories of MO are introduced based on our modelling: *MO Process* offers production skills, while *MO Measure* represents implemented safety devices/measures for the purpose of risk mitigation.

The introduced terminologies can be further described using the presented meta-model in Fig. 1 to illustrate the relationships between different production assets in the fluid production. With this asset relationship, we focus specifically in this paper on the assessment and commissioning of *one* Base that possesses a combination of different Mechatronic Objects (MOs).

To derive our integrated framework for the safety management, we firstly analyze the commissioning process of fluid production systems and identify its relevant connections to safety. The commissioning of a fluid system can be divided into four phases before the production starts: (1) *planning phase*, (2) *external implementation*, (3) *internal implementation*, and (4) *documentation*.
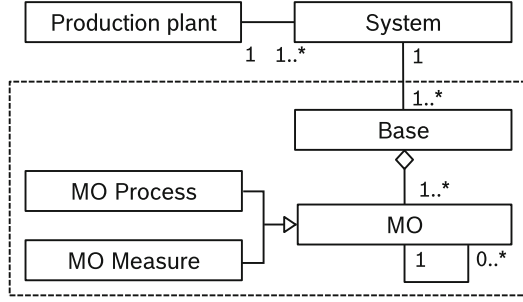
Fig. 1. Meta-model to describe the asset structure for fluid production

During the *planning phase*, the required modular MOs that are needed are selected. Relevant tasks that can be done in this phase include the layout planning of MOs, the analysis of production cycle times and the derivation of production cost. In the second phase *external implementation*, further fine-tuning of the selected MOs can be carried out to fulfill production requirements and to prepare for the system modification. Here, the selected MOs must undergo a safety pre approval process to ensure that the MOs are safe according to the Machinery Directive [2] and risk assessed based on ISO 12100 [8]. Preapproved MOs will bring along its safety-related digital descriptions and can be integrated.

During the third *internal implementation* phase, a temporary production shutdown is required to conduct the system modification. The selected MOs are now integrated into the production environment. A system risk assessment and a safety approval will take place (approval of the complete *Base* described in this paper). The interlinking of the integrated MOs will be risk assessed and, if necessary, optimisation/mitigation suggestions will be made. The *documentation* phase ensures that requirements based on the Machinery Directive [2] are fulfilled and represents the final step before the new system configuration is commissioned. This required documentation contains e.g. operating manuals, conformity declarations, involved production costs and hazard/risk assessment documents.

4   Fluid Production Safety 4A-Framework (FluPro-S4A)

Based on the modeling and the commissioning process explained in Sect. 3, a framework called FluPro-S4A is proposed for the integrated safety management/assisted risk assessment of fluid production systems (Fig. 2). During these presented assessment procedures, the acquired up-to-date system model that represents the production system will be used as foundation for all the conducted assessments. The FluPro-S4A framework constitutes four main steps described as follows:

1. *Change **A**cknowledgement*: This step represents the starting point of the framework and is usually triggered manually by the operator after a system

modification or automatically by a monitoring system. The up-to-date system model will be checked and its changes acknowledged for the subsequent steps. (see Sect. 4.1)

2. *Asset Assessment*: If system changes are identified, every integrated asset will be assessed to ensure its validity. This is to ensure that the assets fulfill the basic requirements for operation. (see Sect. 4.2).

3. *System Assessment*: During this overall system assessment, the whole production system (i.e. system model) will be assessed using our developed methods (see Sect. 4.3) to ensure that all identified risks are properly addressed and mitigated by the available safety measures.

4. *Approval Assistance*: In the final step (see Sect. 4.4), relevant methods will be applied to show how the assessment results can be post-processed to support safety engineers during decision-making.
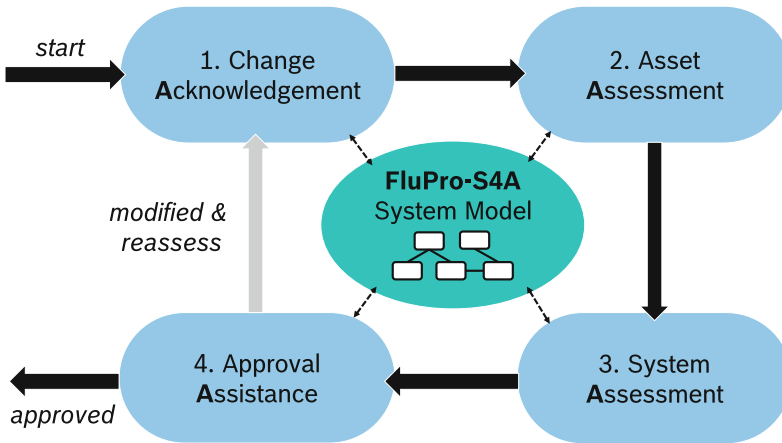
Fig. 2. The proposed FluPro-S4A framework for an integrated assessment

Using the FluPro-S4A framework, the operator can also improve the production system configuration gradually based on the generated assessment results. As can be seen in Fig. 2, if adjustments to the production system are made, a reassessment can be triggered to reevaluate the system configuration. The system model will be updated digitally and the aforementioned four assessment steps will be conducted again. In the following subsections, detailed descriptions will be provided for the four assessment steps within FluPro-S4A framework.

4.1   Step 1: Change Acknowledgement

A fluid production system is described by the modeling method explained in Sect. 3 and represented digitally using a *system model*. Through this semantic data model of the system, the structural interconnections between the assets

and the base system can be described and visualized digitally, which enables the change identification and acknowledgement of the fluid production by comparing both the states before (S) and after (S') a system change. This change acknowledgement analysis will look through both the base and the asset level to identify the type of system change. The three possible types of system changes are *base structural changes* (changes at the level of production systems), *asset structural changes* (changes within a Base) or *asset configuration changes* (changes of a MO/asset itself). Each of these types will have different implications and will be assessed differently (further details will be provided in the subsequent Sect. 4.3).

4.2   Step 2: Asset Assessment

The main focus of this assessment step is to ensure that the integrated production assets have properly gone through the safety preapproval stage before being integrated into the overall fluid production. The status of an asset can be issued during the preapproval stage and is represented by its *approval status*. A confirmation of the approval status is required to guarantee its validity and to avoid data incompleteness during the *System Assessment*. Operators will be guided by the framework to properly deal with invalid assets during the asset integration. We define four types of approval status with the following explanation:

1. Status *"approved"*: The asset has been checked during the preapproval stage and its risks are eliminated inherently. The Asset can be integrated within the fluid production without additional safety measures.
2. Status *"pre-approved"*: The asset has undergone the preapproval procedures and possesses residual risks that cannot be fully eliminated by itself. Further safety measures have to be applied and confirmed during the overall *System Assessment*.
3. Status *"expired"*: An *expired* asset loses its approval validity and needs to be reapproved before being integrated into the fluid production. Possible reasons for an "*expired*" status are the necessity for an inspection after a predefined timeframe.
4. Status *"awaiting approval"*: The asset has not gone through the safety preapproval stage and requires preapproval to ensure its data completeness during the overall *System Assessment*.

4.3   Step 3: System Assessment

The overall *System Assessment* ensures that possible risks due to the combination of all integrated production assets within the fluid production are properly identified and mitigated. Using the methods proposed in this paper, the process can be conducted in a semi-automated manner to provide assistance to users during the commissioning phase and consists of two parts with different purposes: (1) the identification of emerging risks based on the combination of assets, (2) the integrated process to provide safety evidences for the approval.

Identification of Emerging Risks Due to Combination of Assets

During the consideration of emerging risks (i.e. risks resulting of the overall system configuration), several factors need to be taken in account. Firstly, risks might arise due to visible root causes (e.g. movement of a mechanical components) or due to non-visible reasons (e.g. timing behaviour). Secondly, possible relevant risks might not be completely identified or implied digitally. This means that user role remains crucial during the hazard/risk identification process. We propose the following methods to assist users in identifying emerging risks for fluid production:

1. *Simulation*: The application of simulation-based approaches, such as mechanical analysis using physics engine or non-mechanical analysis, allows a better understanding of the system behavior after integration. Possible emergent failures or risks can be identified using these methods.
2. *Logical implication*: Logical implication is useful where information is incomplete or provided in a rather abstract way. The identified risks from this implication will contain uncertainty and should be confirmed manually by users or experts. An exemplary scenario to apply this method is "if an asset with kinematics (i.e. robot) is integrated, possible crushing points will occur in the working space".
3. *Guided questionnaire*: Guided questionnaire can be applied in situations where simulation-based approaches and logical implication are not possible or too complex to be modelled digitally. As an asset might have specified conditions that have to be fulfilled to guarantee its safe operation, a questionnaire to guide the operators during the identification and the analysis of emergent risks can be helpful.

Proposal of an Integrated Process for the Fluid Production

Based on the possible types of system changes for fluid production stated in Sect. 4.1 and the proposed methods for the identification of emerging risks, an integrated process for the overall *System Assessment* can be defined (visualized in Fig. 3) to categorize, identify and assess risks related to a concrete system configuration (represented by *System Model*).

As can be seen in Fig. 3, a series of steps are defined within this *System Assessment* based on our proposed methods for fluid production. Previously identified risks from the preapproval stage that are provided by every integrated asset will be updated to the risk list. By comparing both the previous (S) and the current state ($S^{'}$), the involved types of changes can be identified. Different software modules can be activated throughout the process to guide the user in completing the safety approval/risk assessment procedures.

4.4   Step 4: Approval Assistance

The final step within the FluPro-S4A framework focuses on methods to provide assistance and to increase the efficiency of the approval process. Current
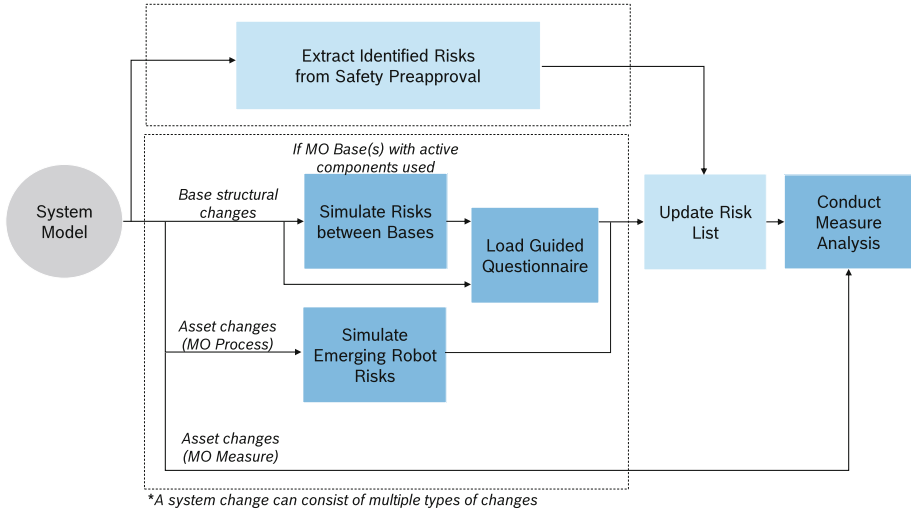
Fig. 3. The proposed exemplary procedures based on our addressed case study for fluid production

legislative requirements only allow semi-automated, but not fully automatically generated results so far. Future work will be done to clarify and to enhance the decision making by a software.

5 Case Study

For a clearer understanding, a case study for the assembly of a small electrical control unit, comparable to those widely used in the automotive industry, is demonstrated (see Fig. 4). This case study shows a possible scenario in fluid production, in which previously unused MOs can be integrated seamlessly into production based on the required production changes. We model the system using our presented approach to describe the *Base* and the *MOs* (see "3" in Fig. 4). Besides, we also demonstrate the application of different software tools for the integrated safety management within the fluid production:

- The *preapproval tool* can be used to assist the user during the hazard analysis and the preapproval of individual MO. The tool is integrated into the communication infrastructure and will update the digital model with the relevant hazard/risk descriptions of the MO for further analysis in the subsequent steps. (see "2" in Fig. 4)
- The *system safety assessment/approval tool* called AutoSafety based on [10] can be used to assess the overall system configuration. Data can be extracted from the preapproval phase and be further analyzed. Different software analysis modules and simulations can be implemented and applied based on our approaches proposed in Sect. 4.3. (see "4" in Fig. 4)
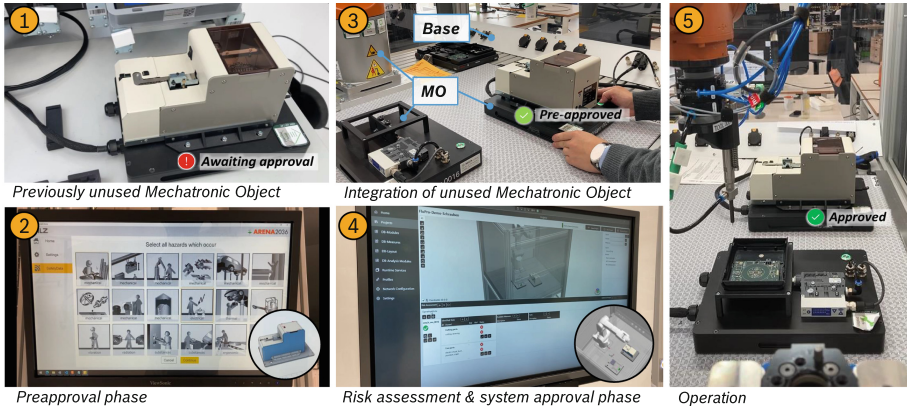
Fig. 4. Demonstration of the presented approach for an assembly process and the usage of software tools for the preapproval and risk assessment/system approval phases

With the new system configuration completely risk assessed, documented and finally approved by the responsible safety engineer under the assistance of the aforementioned tools, the new production process can now operate with minimal downtime and reduced manual effort.

6   Conclusion and Future Work

This paper presents an integrated approach to enable the dynamic safety management and risk assessment for fluid, reconfigurable production scenarios. By using our presented framework called Fluid Production Safety 4A-Framework (FluPro-S4A), modular I4.0 production assets and its data can be modelled considering both functional and safety-related aspects. Besides, different assessment/approval steps are presented to capture system changes, to support risk assessment, and to analyze the respective safety impact in a systematic and seamless manner. This framework contributes to a more efficient safety management, enables interoperability of vendor-independent production assets with safety guarantees and lays the foundation for the successful execution of fluid production systems for future factories. Future work includes further optimization of the approval procedures and the development of methodologies to better identify emerging risks for fluid production.

References

1. Fries, C., et al.: Fluid manufacturing systems (FLMS). In: Weißgraeber, P., Heieck, F., Ackermann, C. (eds.) Advances in Automotive Production Technology – Theory and Application. A, pp. 37–44. Springer, Heidelberg (2021). https://doi.org/10.1007/978-3-662-62962-8_5

2. Machinery directive 2006/42/EC. Directive, The European Parliament and the Council of the European Union (2006)

3. Akerlund, O., et al.: Isaac, a framework for integrated safety analysis of functional, geometrical and human aspects. Proc. ERTS **2006**, 1–11 (2006)

4. Askarpour, M., Mandrioli, D., Rossi, M., Vicentini, F.: SAFER-HRC: safety analysis through formal vERification in human-robot collaboration. In: Skavhaug, A., Guiochet, J., Bitsch, F. (eds.) SAFECOMP 2016. LNCS, vol. 9922, pp. 283–295. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45477-1_22

5. Awad, R., Fechter, M., van Heerden, J.: Integrated risk assessment and safety consideration during design of HRC workplaces. In: 22nd International Conference on Emerging Technologies and Factory Automation (ETFA), pp. 1-10. IEEE (2017)

6. Huck, T.P., Ledermann, C., Kröger, T.: Virtual adversarial humans finding hazards in robot workplaces. arXiv preprint: arXiv:2103.00973 (2021)

7. Huck, T.P., Münch, N., Hornung, L., Ledermann, C., Wurll, C.: Risk assessment tools for industrial human-robot collaboration: novel approaches and practical needs. Saf. Sci. **141**, 105288 (2021)

8. International Organization for Standardization (ISO): ISO 12100: Safety of machinery - General principles for design - Risk assessment and risk reduction (2010)

9. Koch, T.: Approach for an automated safety configuration for robot applications. Procedia CIRP **84**, 896–901 (2019)

10. Koo, C.H., Schröck, S., Vorderer, M., Richter, J., Verl, A.: A model-based and software-assisted safety assessment concept for reconfigurable PnP-systems. Procedia CIRP **93**, 359–364 (2020)

11. Pfeifer, M., Varro, W.: Sicherheitsnachweis für modulare Anlagen, vol. 19, pp. 114–115 (2019)

12. Popper, J., et al.: Safety on modular machines: Whitepaper sf-3.1 (2018)

13. Richter, D.: Sicherheit von vernetzten modularen Industrieanlagen erfordert eine dynamische Sicherheitsarchitektur 4.0 (Safety and Security) (2018)

14. Sallez, Y., Berger, T., Bonte, T.: The concept of "safety bubble" for reconfigurable assembly systems. Manufact. Lett. **24**, 77–81 (2020)

15. Vicentini, F., Askarpour, M., Rossi, M.G., Mandrioli, D.: Safety assessment of collaborative robotics through automated formal verification. IEEE Trans. Rob. **36**(1), 42–61 (2019)

16. Shin, H., Choi, J., Choi, J., Rhim, S.: Physical safety analysis of robot considering impactor shape. In: 2017 2nd International Conference on Robotics and Automation Engineering (ICRAE), pp. 1-5 (2017)

17. Andrisano, A.O., Leali, F., Pellicciari, M., Pini, F., Vergnano, A.: Hybrid reconfigurable system design and optimization through virtual prototyping and digital manufacturing tools. Int. J. Inter. Des. Manufact. (IJIDeM) **6**, 17–27 (2012)

18. Jaradat, O., Punnekkat, S.: Using safety contracts to verify design assumptions during runtime. In: Casimiro, A., Ferreira, P.M. (eds.) Ada-Europe 2018. LNCS, vol. 10873, pp. 3–18. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-92432-8_1

Part B: Data-Driven Technologies

Improvement of the Scheduling of Automotive Testing Processes Based on Production Scheduling Methods

Leon Stütz[(✉)] , Timo König , Roman Bader, and Markus Kley

Institute of Drive Technology Aalen (IAA), Aalen University, Beethovenstraße 1, 73430 Aalen, Germany
leon.stuetz@hs-aalen.de

Abstract. Increasing challenges in the automotive industry are caused by shorter development times for products, greater diversity of variants and increasing cost pressure. Testing plays an elementary role within the product development process (PDP). There are already many publications that deal with the early phases of the PDP, but relatively few that address testing. Inefficient scheduling leads to suboptimal use of development and testing resources.

Automotive testing is characterized by high momentum and process complexity. The complexity of testing is determined, among other things, by the number of test rigs in a test field, the number and diversity of test objects, the type of testing and the preparatory setups. In addition, complex testing processes at the component and system level require a large number of human and material resources, whose time availability and sequence must be coordinated with the testing process. The sequence planning is subject to a high inherent dynamic because unexpected changes and disturbances of the process can occur during the testing. These changes require a rescheduling of the testing process. If done manually, the rescheduling results in high costs.

Based on known production planning methods, a solution approach is derived for improved utilization of test field resources for the automotive sector. The planning is optimized with a multitude of product - and process-related dependencies and restrictions using mixed-integer linear programming, a standardized method from operations research. The test field is simulated via a discrete event simulation. The proposed method considers the availability of essential resources.

Keywords: testing optimization · production scheduling · automotive testing · mixed-integer linear programming

1 Introduction

The well-known product development processes from industry and research are at the center of new advances in academia. Validation plays an essential role within product development [1]. Validation activities aim to compare the requirements and goals set for the product with the current status of the product as part of the product development

process (PDP). To ensure operationally and functionally safe vehicles, the process steps of simulation (virtual validation), test rig testing (test rig validation) and real road tests (vehicle validation) are run through in the course of testing [2].

Due to shorter development times, increasing cost pressure and the higher number of product variants, a methodical approach in product development is indispensable to achieve reliable and reproducible development results. Research work to date in the area of electrified powertrains has increasingly focused on early phases of PDP, which is why rig testing must be given priority to ensure high efficiency and reliability [3].

Overall vehicle reliability and efficiency is essentially determined by the five categories of product, process, environment, method and people. Testing is a subset of the process category and a key factor along with research, development, simulation and design [4]. To ensure a methodical approach to testing, the V-model of system development can be used (see Fig. 1). The approach required for planning the testing processes is discussed. In particular, approaches to data-driven production optimization are applied in the derivation of a relevant method for the allocation of test rig fields [5, 6].
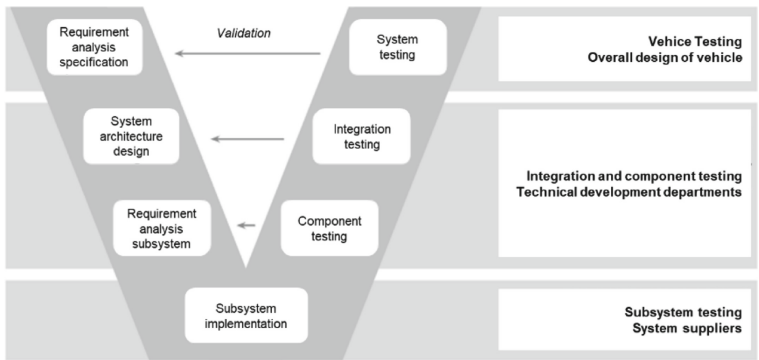
Fig. 1. V-model of system development [7]

In development phases, unforeseen interruptions and unavailable testing resources can negatively impact the testing processes. This has an impact on the time schedule. The proposed approach of simulating a virtual test field with a discrete event simulation in combination with a mathematical optimization model as a reference order generator is intended to counteract this problem in advance of testing so that critical time schedules can be identified. The simulation also enables an early estimation of critical testing processes with regard to completion times, considering disturbances that may occur.

2   State of the Art

The problem of interest with most industrial processes is to find the most efficient way to produce a set of products or services in a given time period using a limited set of resources. Due to the large potentials in resource savings, the scheduling of processes in industrial environments has attracted an increasing amount of attention from academia and industry.

The field of PDP is of interest for academia and industry because it is one of the main ways to achieve competitive advantage for a company. The performance of a product and its cost are defined in its development. The optimization of these two parameters is necessary for cost management [8, 9]. For many manufacturing companies, innovation, design and successful management of new product development often present major challenges [10, 11]. Long development times, prohibitive development and manufacturing costs, and poor quality have been common results for many of these organizations. The primary factor contributing to such unsuccessful results is the use of traditional sequential new product development by these organizations [12, 13]. Conversely, the literature over the past three decades clearly shows that, through their lean manufacturing practices, world class organizations, such as Toyota, have dominated competition not only in the area of manufacturing but also in the area of innovation, design, development and commercialization of new technologies [13–15]. Although the scheduling potential has been known for a long time, in recent years the substantial advances of related modeling and solution techniques, as well as the rapidly growing computational power have enabled new solutions for the existing problem [16].

Scheduling of automotive testing processes and production scheduling methods have many overlaps, mainly in the field of production planning. Manufacturing resources provide values for cost, quality, time and environmental impacts, which multiply with their usage within a manufacturing task for a specific part [17]. Accordingly, the planning of testing processes also aims to reduce the necessary resources (human resources, time, etc.) to conduct the testing task.

In most cases enterprise resource planning systems are used for production planning in combination with integrated manufacturing execution systems and advanced planning and scheduling systems. In production planning, especially in the scientific field, control algorithms are also taken into account in order to consider the constantly updated planning according to current production conditions [18, 19]. The production including the production control system is considered as a control loop, so that the current production status can be taken into account by data acquisition and if changes occur can be counteracted [20].

To enable the best possible planning, mathematical optimization algorithms are used, for instance mixed integer linear programming (MILP), which are mainly adapted to computational efficiency and the quality of results in terms of the minimization of setup and total flow time [21, 22]. Discrete event simulations can also be used to create a digital twin of a production environment to test, as well as validate the production control systems at hand with the virtual machine models [23, 24]. These simulation models are also used for production simulation in order to check and evaluate defined processes in production environments [25]. Similarly, the effects of alternative scenarios and different framework conditions on the production processes can be simulated to match the real system underlying the simulation model with the findings [26].

The scope of this paper is to transfer an approach from virtual testing of production environments, essentially consisting of virtual machines and a reference job generator, to a digital twin of a test field. Thus, the test field consists of virtual test rigs. The test field is also simulated as realistically as possible via a discrete-event simulation. The transfer will enable an estimation of temporal conditions and critical time schedules.

3   Test Field Modelling

3.1   Physical Test Field

The investigated test field consists of four component test rigs and one system test rig (see Fig. 2). The testing task, which is derived from a system context, is divided into sub-tasks at the component level and then assigned to the component test rigs. This enables simultaneous testing on several test rigs. It is also possible to substitute individual testing tasks, e.g. if a rush job is to be included in the test plan. Through the efficient combination of component and system test rigs, the test time can be reduced and the test rig utilization, through the more flexible use of the test rigs, can be significantly increased.
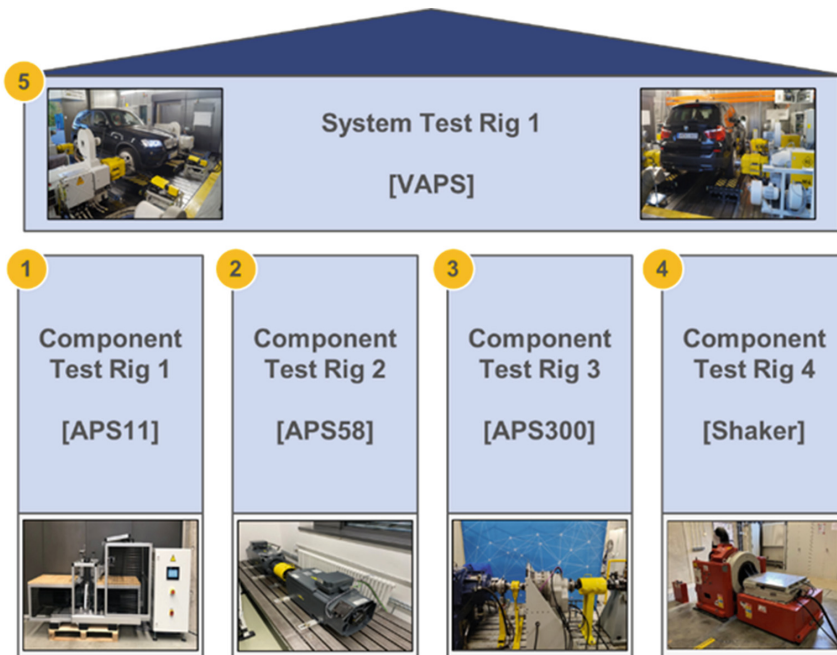
Fig. 2.   Test field: component and system test rigs

3.2   Virtual Test Field

In modern test fields, accurate knowledge of test times and, above all, identification of critical time schedule is essential to ensure optimal testing in complex test fields. The dynamic fluctuations, which should be considered in real time if possible, pose further challenges. These problems can be addressed with a control system that allows an up-to-date consideration of the status and counteracts deviations from the schedule in combination with a virtual test field (see Fig. 3). The aim is therefore to replicate a real test field as realistically as possible and as detailed as necessary.
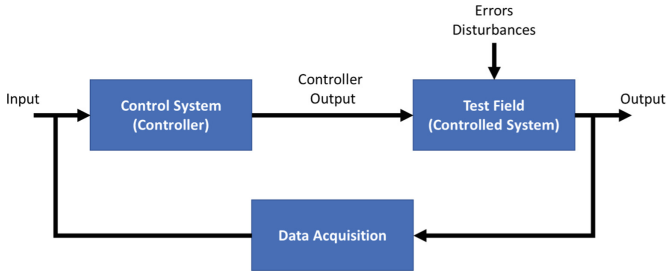
Fig. 3. Visualization of the structure of a control system to simulate virtual testing

The structure of the basic control loop is therefore transferred to a model for the simulation of testing processes and essentially consists of a reference order generator and virtual test rigs in a defined test field. Relevant data during the testing process can be recorded via test rig data (TRDA) and operating data (ODA) acquisition. The level of detail of the virtually represented test rigs corresponds to a very abstract level, because ultimately only higher-level variables, such as the test times, are relevant for a process planning. The reference order generator generates jobs for the virtual test rig. Based on defined testing scenarios a mathematical model is defined. In the following use case a MILP is used for a mathematical optimal order release. The reference order generator is connected via different communication processes with the virtual test rig. The virtual test rigs consist of a testing process simulation and an availability simulation, which are both dependent on the operating calendar of the test rigs. The availability simulation considers malfunctions, non-availability and operational reasons (e.g. personnel availability). The calendar, the testing process simulation as well as the availability simulation are connected via a simulation logic. The causes limit the availability and result in an average utilization rate. This is maintained by means of stochastic modeling averaged over a longer period of time (see Fig. 4).

4 Scheduling Results

In this section, the functionality of the developed model for simulating virtual testing processes and identifying critical time schedules during the process is presented. The virtual test field, which is affected by disturbances and unavailability, among other things, requires several reschedules during the process. The exemplary testing scenario visualized in Fig. 5 is composed of an initial state of testing orders and a variation, which in this use case represents a rush order, that the static testing process becomes a dynamic process with adjustments required. The initial state consists of test orders (A and B and C) for component testing and one order for system testing (ABC). System testing can only be started once the individual component tests have been completed. In the scenario, a total of four component test rigs and one system test rig are considered, as described in Sect. 3.1. The release of the test order (initial state) was on 30.05.2022 at 7.00 am. Different shift models (e.g. single-shift or three-shift) are considered for the test rigs, depending on the degree of automation.
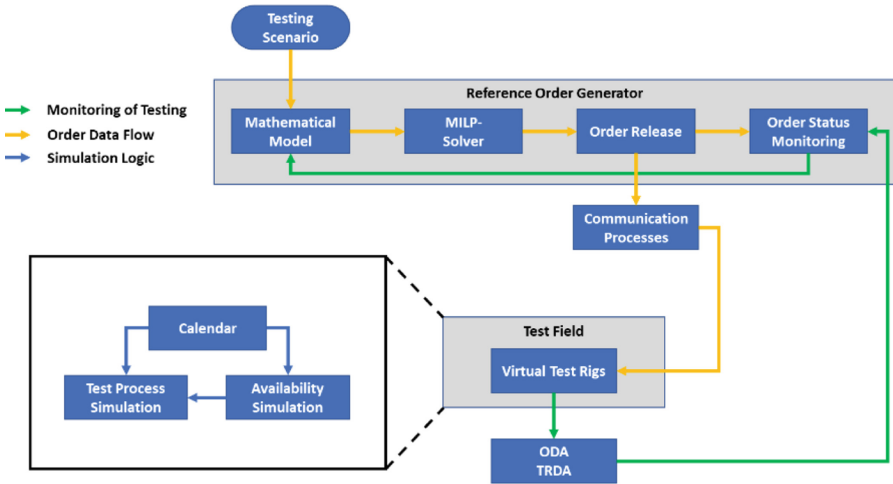
Fig. 4. Structure of the model for the simulation of virtual test procedures
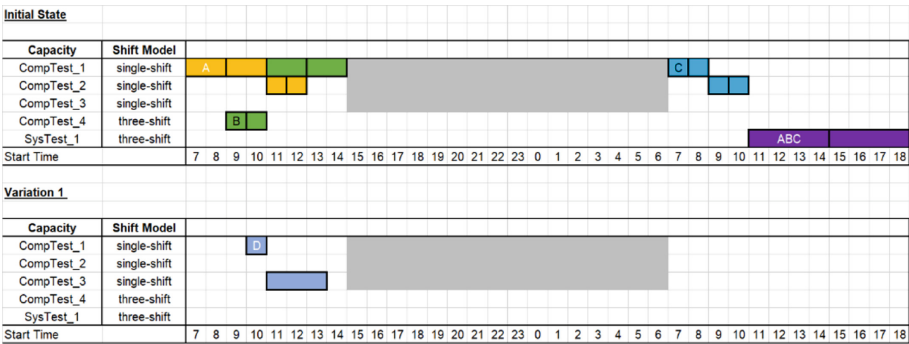
Fig. 5. Exemplary testing scenario with the defined test field

Due to delays in ongoing testing processes caused by disruptions and personnel absences, as well as the rush order, an update of the planning is essential to represent the current situation. The control process is performed with the MILP solver. The control strategy underlying the simulation model aims to complete all test orders as early as possible. So, the interval between the planned completion time and the latest permitted completion time is maximized. Several test specimens can be tested at the individual test rigs. Each individual test specimen can be passed on separately to the next test step. Product changes in the meantime are not permitted, so that the setup time that has to be considered for a change is minimized as far as possible. In the model, a degree of utilization that may have an impact on the actual test time can be considered for the test rigs.

Figure 6 below represents the results of the simulated planning scenario with multiple readjustments. In each case, the planned completion dates are shown as a function of the rescheduling steps. The necessary regulations during the testing process and the

resulting new schedule can be clearly seen in the individual figures. Adjustments at a single testing step also have a significant impact, here especially on system testing.
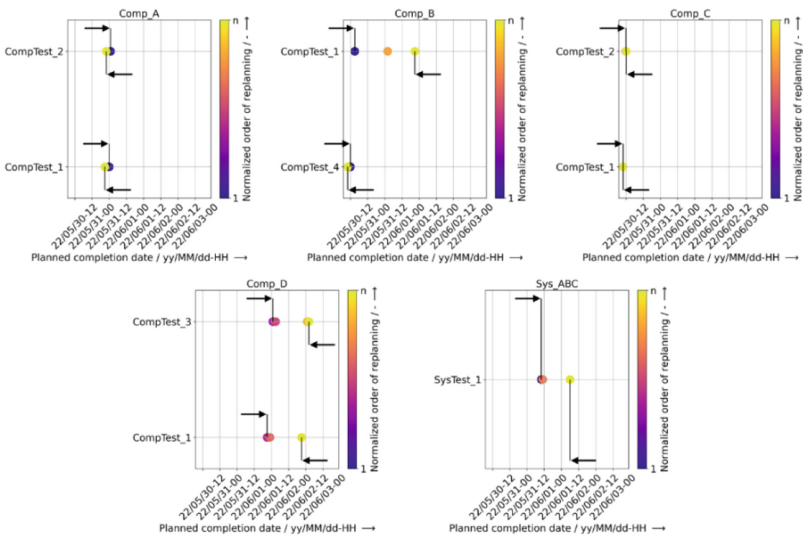
Fig. 6. Results of the simulated testing scenario

5 Conclusion and Outlook

In the context of this publication, it is shown that a simulation model for the virtual representation of test rigs enables an estimation and identification of critical time schedules with regard to the completion date, considering disturbances and personnel unavailability. Scheduling is based on MILP rule algorithms, so that an early completion of all test jobs is targeted. The component and system test rigs of the Institute of Drive Technology Aalen provided a sufficient reference environment to demonstrate the behavior and interaction of the individual test rigs with regard to the required replanning. The influence on each other and the influence of disturbances and unavailability can be seen, so that an improvement of the resulting test schedule is demonstrated when using the simulation model and MILP algorithm compared to basic planning methods.

The shown research results also demonstrated the advantages of the MILP algorithm with regards to the inclusion of rush orders into an existing order management scheme. The uncertainty of an upcoming rush order did not interfere with the completion enddate for the previously integrated orders. The implications for practice suggest high degree of flexibility in the order management planning of testing processes, adjusting to the rapidly changing incoming orders and substitution effects in the daily business of testing facilities.

cooperation project under the funding code ZF4113823LF9. The exemplary simulations were carried out, among others, with the help of the road-to-rig vehicle test rig "VAPS", which was funded by the DFG (funding code: INST 52/19-1 FUGB). The funding of the test rig was enabled by the building ZiMATE (funding code: BW 6710001).

References

1. Institute of Electrical and Electronics Engineers, IEEE Standard for software verification and validation. Institute of Electrical and Electronics Engineers, New York (1998)
2. Klos, W., Schenk, M., Schwämmle, T., Müller, M., Bertsche, B.: Antriebsstrangerprobung bei der Daimler AG. Moderne Erprobungsmethodik. Internationales Symposium für Entwicklungsmethodik (2011)
3. Schenk, M.: Adaptives Prüfstandsverhalten in der PKW-Antriebstrangerprobung. Dissertation, Universität Stuttgart, Stuttgart (2017)
4. Karthaus, C.A.: Methode zur Rückführung von Erprobungswissen in die Produktentwicklung am Beispiel Fahrzeugtriebstrang. Dissertation, Institut für Konstruktionstechnik und Technisches Design, Universität Stuttgart, Stuttgart (2020)
5. Palm, D., et al.: Datengetriebene Produktionsoptimierung in der Montage. Zeitschrift für wirtschaftlichen Fabrikbetrieb **113**(7–8), 518–521 (2018). https://doi.org/10.3139/104.111954
6. Braun, A., et al.: Der Weg zur Industrie 4.0-Roadmap. Zeitschrift für wirtschaftlichen Fabrikbetrieb **113**(4), 254–257 (2018). https://doi.org/10.3139/104.111888
7. Ungermann, J.: Zuverlässigkeitsnachweis und Zuverlässigkeitsentwicklung in der Gesamtfahrzeugerprobung, Dissertation, ETH Zurich, Zurich (2009)
8. Dekker, H., Smidt, P.: A survey of the adoption and use of target costing in Dutch firms. Int. J. Prod. Econ. **84**(3), 293–305 (2003). https://doi.org/10.1016/S0925-5273(02)00450-4
9. Ibusuki, U., Kaminski, P.C.: Product development process with focus on value engineering and target-costing: a case study in an automotive company. Int. J. Prod. Econ. **105**(2), 459–474 (2007). https://doi.org/10.1016/j.ijpe.2005.08.009
10. Gerhard, D., Brem, A., Voigt, K.-I.: Product development in the automotive industry: crucial success drivers for technological innovations. Int. J. Technol. Mark. **3**(3), 203–222 (2008). https://doi.org/10.1504/IJTMKT.2008.019922
11. Liang, J.C.: An integrated product development process in the automotive industry. Int. J. Prod. Dev. **8**(1), 80 (2009). https://doi.org/10.1504/IJPD.2009.023750
12. Liker, J.K., Morgan, J.M.: The toyota way in services: the case of lean product development. AMP **20**(2), 5–20 (2006). https://doi.org/10.5465/AMP.2006.20591002
13. Blackburn, J.D.: Time-Based Competition: The Next Battleground in American Manufacturing. Business One Irwin, Homewood (1991)
14. Unger, D.W., Eppinger, S.D.: Comparing product development processes and managing risk. IJPD **8**(4), 382 (2009). https://doi.org/10.1504/IJPD.2009.025253
15. Rohde, J., Jahnke, U., Lindemann, C., Kruse, A., Koch, R.: Standardised product development for technology integration of additive manufacturing. Virtual Phys. Prototyping **14**(2), 141–147 (2019). https://doi.org/10.1080/17452759.2018.1532801
16. Floudas, C.A., Lin, X.: Mixed integer linear programming in process scheduling: modeling, algorithms, and applications. Ann. Oper. Res. **139**(1), 131–162 (2005). https://doi.org/10.1007/s10479-005-3446-x

17. Reiff, C., et al.: A process-planning framework for sustainable manufacturing. Energies **14**(18), 5811 (2021). https://doi.org/10.3390/en14185811
18. Berwing, K., Lütkehoff, B., Fischer, M., Schröter, M., Steinlein, F.: Produktionsregelung: Perspektiven der Produktionssteuerung im Kontext der Digitalisierung (2022). https://bit.ly/3xB2M5W
19. Pritschow, G., Brinzer, B.: Produktionsregelung für die variantenreiche Serienfertigung. Zeitschrift für wirtschaftlichen Fabrikbetrieb **97**(12), 627–632 (2002). https://doi.org/10.3139/104.100598
20. Kletti, J., Schumacher, J.: Die perfekte Produktion. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-45441-1
21. Allahverdi, A.: The third comprehensive survey on scheduling problems with setup times/costs. Eur. J. Oper. Res. **246**(2), 345–378 (2015). https://doi.org/10.1016/j.ejor.2015.04.004
22. Steger-Jensen, K., Hvolby, H.-H., Nielsen, P., Nielsen, I.: Advanced planning and scheduling technology. Prod. Plan. Control **22**(8), 800–808 (2011). https://doi.org/10.1080/09537287.2010.543563
23. Pfeiffer, A., Kádár, B., Monostori, L.: Evaluating and improving production control systems by using emulation. In: Proceedings of the Twelfth IASTED International Conference on Applied Simulation and Modelling, 3–5 September 2003, Marbella, Spain, Anaheim. Acta Press (2003)
24. Gyulai, D., Pfeiffer, A., Kádár, B., Monostori, L.: Simulation-based production planning and execution control for reconfigurable assembly cells. Procedia CIRP **57**(7), 445–450 (2016). https://doi.org/10.1016/j.procir.2016.11.077
25. Delgado Sobrino, D.R., Ružarovský, R., Václav, Š., Cagáňová, D., Rychtárik, V.: Developing simulation approaches: a simple case of emulation for logic validation using tecnomatix plant simulation. J. Phys. Conf. Ser. **2212**(1), 12011 (2022). https://doi.org/10.1088/1742-6596/2212/1/012011
26. Schuh, G., Reuter, C., Hauptvogel, A., Brambring, F.: Cyber-physische Produktionssteuerung. Zeitschrift für wirtschaftlichen Fabrikbetrieb **109**(12), 964–967 (2014). https://doi.org/10.3139/104.111245

Mixed-Integer Programming Model for Scheduling of Modular Automotive Body-In-White Production Systems

Jan M. Gelgfren[1,2]([✉]), Hélène Arvis[1], Simon Hagemann[1], and Sigrid Wenzel[2]

[1] Mercedes-Benz Group AG, Bela-Barenyi-Straße 1, 71059 Sindelfingen, Germany
jan.markus.gelfgren@mercedes-benz.com
[2] Department Organisation of Production and Factory Planning, University of Kassel, Kurt-Wolters-Straße 3, 34125 Kassel, Germany

Abstract. The turbulent transition in the automotive industry towards electric vehicles is highly challenging, both in product development and in production. Conventional, linear assembly lines struggle to meet the flexibility requirements imposed by this market development. This paper presents the first mixed-integer programming (MIP) model which is tailor-made for the scheduling of modular body-in-white (BIW) production systems. The main novelty of the presented approach lies in using precedence graphs for modelling the joining steps of the BIW production and allowing for different capabilities at each workstation of the production system. Thus, the presented approach captures the characteristics of BIW production systems in more detail than comparable models available in the literature. The application of the presented approach to an exemplary production scenario with five jobs and 22 operations underlines the strengths of the proposed model but also indicates potential for future improvements.

Keywords: Production Scheduling · Modular Production · Dynamic Production · MIP Modelling · Job-Shop Scheduling

1   Introduction

The transition in the automotive industry towards electric vehicles - amplified by governmental policies - poses a great challenge, both in product development and in production [1]. This transformation further accelerates the trend in recent years towards a greater product diversity [2]. Linear assembly lines, which are currently well-established in the automotive industry, cannot meet the new requirements in a satisfactory manner [2]. Therefore, more flexible production systems are necessary [2].

Traditionally, the manual final assembly accommodates the largest product diversity [3, pp. 1–4]. However, the escalated demand for electric and hybrid vehicles causes a notable increase in diversity of vehicle bodies too [4]. As a result, not only the final assembly but also the highly automated body-in-white (BIW) production needs to become more flexible to meet the market demands [5].

Modular production addresses these issues [6]. In this production paradigm, the logistics is decoupled from the actual production by breaking up the assembly line into individual, modular production units and automated guided vehicles (AGVs) which transport the parts between the workstations [6]. This radically improves the flexibility of the production system [6].

The scheduling of modular production systems is intimately connected with the job-shop scheduling problem (JSP). The JSP refers to the NP-hard problem of scheduling n jobs of different processing times on m different workstations (WS) of varying processing power, the objective being to minimize the makespan [7]. The traditional JSP only considers the scheduling of the operations on the workstations, but in modular production, the dynamic transportation durations between the workstations have to be considered as well [8].

Mixed-Integer programming (MIP) modelling is a widely used approach for scheduling tasks [9]. For instance, in 2014, 14 out of the 40 papers published in the *Journal of Scheduling* used different MIP approaches, more than any other technology [9]. Due to its comprehensive modelling capabilities, its potential to find global optima and its prevalence in the literature, the authors have chosen a MIP approach to schedule the modular BIW production. The main novelty of this MIP formulation lies in including precedence graphs for modelling the joining steps of the BIW production and allowing for different workstation capabilities in the production system.

2   State of the Art

Consistently with the chosen approach, the literature review is limited to MIP approaches that encompass both the scheduling of the workstations and the AGVs. Bilge et al. [10] presented the first MIP model for the joint production and transportation scheduling for flexible production systems. However, it could not be solved due to its size and non-linearity. Instead, the authors developed an iterative procedure to determine a feasible workstation schedule. Using this schedule, a time window is calculated during which job delivery is possible. A heuristic schedules the AGVs to deliver a job within this time schedule.

Fontes et al. [8] proposed a MIP model which takes into consideration the dynamic transportation durations between the workstations, thus making it interesting for flexible production systems in general. They use precedence variables to sequence both the production operations and the AGV transportation tasks. Their modelling approach uses one set of chained decision variables for the workstations and another for the AGVs. These sets are interconnected through the completion time constraints both for the workstations and the AGVs. Since all AGVs are considered identical, the model does not explicitly consider the AGVs. This way, the use of an additional index can be avoided, thus heavily reducing the number of decision variables and constraints.

Homayouni et al. [11] enhanced the model of Fontes et al. [8] by adding more detail to the modelling of the workstations. Due to the NP-hard nature of the problem, the calculation time exponentially escalates with an increasing problem

size. Therefore, they proposed a local search-based heuristic to find solutions to the problem.

This paper extends the work of Fontes et al. [8] and Homayouni et al. [11] by including the following features, necessary for the BIW production:

- The operations of one BIW are not serial, but are formed as a precedence graph [12], exemplified in Fig. 1. For instance, the order of the side walls is indifferent, but both side walls need to be added before the roof can be assembled.
- In the BIW production, many different joining technologies are used [13, pp. 36–39]. Since not all workstations can be assumed to be able to perform all technologies, each workstation can have its own capabilities. For example, one workstation can weld, another can glue, a third can weld and clinch etc.
- The parts are considered as an input. This is implemented as a stepping stone for future work, in which logistical bottlenecks might lead to parts shortages.
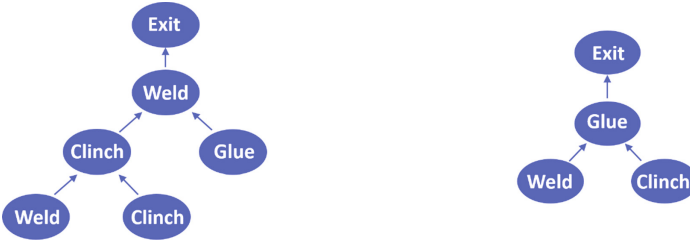
Fig. 1. Examples of precedence graphs.

3   Proposed Mixed-Integer Programming Model

3.1   Notations

Table 1 summarizes all the notations used in the model in Sect. 3.2. Additionally, a directed precedence graph of operations is considered instead of a linear job. To this effect, a topological order $\succeq$ is introduced. For two nodes u and v, this order describes the following relations:

- u precedes v if a chain of directed edges from u to v exists: $u \prec v$
- u succeeds v if a chain of directed edges from v to u exists: $u \succ v$
- u is indifferent to v if no such chains exist:
    - $u \succeq v \implies u$ succeeds or is indifferent to v
    - $u \preceq v \implies u$ precedes or is indifferent to v

This partial order introduces the notion of predecessors and successors for each operation. For any given operation, each predecessor creates a required part whereas each successor requires a created part.

Table 1. Notations used for the model formulation.

Sets, Indices and Definitions

J	Set of jobs. A job is a set of operations needed to produce a BIW
f, t	Dummy jobs, first and last respectively. These jobs have no meaning for the production but are only needed in the model
J_i	Set of operations for job $i \in J' = J \cup \{f, t\}$. An operation is a node in the precedence graph. Each job includes an "exit" operation e_i, corresponding to the travel operation toward the unloading dock U
i, k	Indices for jobs, $i, k \in J'$
j, l	Indices for operations $j, l \in J_i$ for $i \in J'$
L	Loading dock for the parts before they enter the modular production system
U	Unloading dock for the finished vehicle bodies
W	Workstations where operations are executed. Since the "exit" operation is attributed to the unloading dock U, $U \in W$. No operation is attributed to the loading dock L, which is why $L \notin W$
$P_k(l)$	Required parts for the operation l of job k
$C_k(l)$	Created parts by the operation l of job k
$Pred_k(l)$	Predecessor operations of l in the graph of operations of job k
B_k	Primary parts necessary for the job k. Primary parts are parts as found in the storage – in contrast to parts which have been created during the BIW production by joining other parts

Parameters

P_{ij}^a	Process time for operation j of job $i \in J, j \in J_i$ on workstation a
T_a^b	Travel times between the workstations a and b with $a, b \in W \cup \{L\}$
n_i	Number of operations of job $i \in J$, $n_i = |J_i|$
v	Number of AGVs
M	A large positive integer

Decision Variables

$w_{ij}^{kl,a}$	Binary variable equal to 1 if the operation l of job k directly succeeds operation j of job i on workstation $a \in W \setminus \{U\}$
$x_{ij,q}^{kl,p}$	Binary variable equal to 1 if the part p for operation l of job k is transported by the same AGV directly after the part q for operation j of job i
u_{ij}^p	Binary variable equal to 1 if the part p for operation j of job i is the first to be transported by the AGV
z_{ij}^p	Binary variable equal to 1 if the part p for operation j of job i is the last to be transported by the AGV
c_{max}	Production makespan, continuous variable
c_{ij}	Completion time of operation j of job i, continuous variable
v_{ij}^p	Arrival time of part p for operation j of job i, continuous variable
m_{ij}^a	Binary variable equal to 1 if operation j of job i is done by workstation $a \in W$

3.2   Model

The main novelties of this model are taking into account a graph of operations for each job, rather than a linear succession of operations, considering that different workstations can have different capabilities and using parts as input. We consider that each operation can have several predecessors, but only one direct successor.

The input of the model consists of the needed parts, the workstations of the production system and their capabilities, the number of AGVs in the production system and the jobs that are to be produced and their precedence graphs. The objective of the model is to minimize the total production makespan (Eq. 1), defined as the latest completion time of every operation (Eq. 2), that is the time at which every job has been completed and unloaded.

Each operation is carried out on one workstation and has an immediate predecessor (Eq. 3) and an immediate successor (Eq. 4) on the workstation in question. The topological order, as explained in Sect. 3.1, ensures that the predecessors and successors follow the logic of the precedence graph. The dummy jobs f and t ensure the coherence of these constraints for the first and last operations executed on the workstations as boundary conditions.

The AGVs assure the movement of jobs between the workstations. The number of AGVs mobilized is limited by the number of AGVs available (Eq. 5). In addition, we ensure that the same number of AGVs finish the day as start the day (Eq. 6). Concerning the parts, primary parts for the operations are initially stored at the factory entrance, or loading dock. These parts must be transported (Eq. 7), either as the first operation of an AGV, or after another operation. Other transported operations must also have a predecessor on the AGV, or be the first transported (Eq. 8). Symmetrically, the last operation of a job must be transported (Eq. 9) as it represents the unloading of the job, and transported operations are either the last operations transported by their AGV, or have a successor (Eq. 10). To ensure the fluidity of these transports, a flow constraint (Eq. 11) is also present. In addition, not all of the operations are to be transported: an operation either follows its predecessor on the same workstation or retrieves the part created by this predecessor via an AGV (Eq. 12).

The model then ensures that the workstations are properly attributed to each job, in particular assigning the same workstation to two consecutive operations (Eq. 13) and a single workstation to each operation (Eq. 14), as well as assigning each operation to a workstation of the correct category (Eq. 15). Furthermore, dummy jobs are assigned to a single workstation (Eq. 16), and the unloading dock is only assigned to the dummy exit variable of each job (Eq. 17).

In order to define the schedule, one must define the completion and arrival times for each operation. The completion of an operation depends on the arrival time of all of the required parts and on the processing time on the assigned workstation (Eq. 18). Since a workstation can only process one operation at a time, the completion of the previous operation on the same workstation is a necessity. The operation can only be declared complete after this previous operation and its own processing on that workstation (Eq. 19). After the completion of the predecessor operations, the required parts are transported towards the workstation

of the current operation, which defines the arrival time. (Eq. 20). If an operation contains primary parts, those parts must be fetched from the loading area and brought to the designated workstation (Eq. 21).

In order to drop off parts for an operation, the AGV carrying them has to have dropped off the previous load and picked up the required parts of the current operation at the workstation at which they were created (Eq. 22). Finally, should an operation be transported after another and require primary parts, the designated AGV must pick up the parts at the loading area after having dropped off the previous operation (Eq. 23).

$$\min c_{max} \tag{1}$$

subject to

$$c_{max} \geq c_{ie_i}, \qquad \forall i \in J \tag{2}$$

$$\sum_{a \in W \setminus \{U\}} \left[\sum_{j \preceq l \in J_k} w_{kj}^{kl,a} + \sum_{\substack{i \in J_k \setminus \{k\} \\ j \in J_i}} w_{ij}^{kl,a} + \sum_{j \in J_f} w_{fj}^{kl,a} \right] = 1, \qquad \forall k \in J \cup \{t\}, l \in J_k \setminus \{e_k\} \tag{3}$$

$$\sum_{a \in W \setminus \{U\}} \left[\sum_{j \succeq l \in J_k} w_{kl}^{kj,a} + \sum_{\substack{i \in J_k \setminus \{k\} \\ j \in J_i}} w_{kl}^{ij,a} + \sum_{j \in J_t} w_{kl}^{tj,a} \right] = 1, \qquad \forall k \in J \cup \{f\}, l \in J_k \setminus \{e_k\} \tag{4}$$

$$\sum_{k \in J, l \in J_k, p \in P(l)} u_{kl}^p \leq v \tag{5}$$

$$\sum_{k \in J, l \in J_k, p \in P_k(l)} u_{kl}^p = \sum_{k \in J, l \in J_k, p \in P_k(l)} z_{kl}^p \tag{6}$$

$$u_{kl}^p + \sum_{\substack{j \preceq l \in J_k \\ q \in P_k(j)}} x_{kj,q}^{kl,p} + \sum_{\substack{i \in J \\ i \neq k}} \sum_{\substack{j \in J_i \\ q \in P_i(j)}} x_{ij,q}^{kl,p} = 1, \qquad \forall k \in J, \forall l \in J_k, \forall p \in P_k(l) \cap B_k$$

$$\tag{7}$$

$$u_{kl}^p + \sum_{\substack{j \preceq l \in J_k \\ q \in P_k(j)}} x_{kj,q}^{kl,p} + \sum_{\substack{i \in J \\ i \neq k}} \sum_{\substack{j \in J_i \\ q \in P_k(j)}} x_{ij,q}^{kl,p} \leq 1, \qquad \forall k \in J, \forall l \in J_k, \forall p \in P_k(l) \tag{8}$$

$$z_{ke_k}^p + \sum_{\substack{i \in J \\ i \neq k}} \sum_{\substack{j \in J_i \\ q \in P_i(j)}} x_{ke_k,p}^{ij,q} = 1, \qquad \forall k \in J, \forall p \in P_k(e_k) \tag{9}$$

$$z_{kl}^p + \sum_{\substack{j \succeq l \in J_k \\ q \in P_k(j)}} x_{kl,p}^{kj,q} + \sum_{\substack{i \in J \\ i \neq k}} \sum_{\substack{j \in J_i \\ q \in P_i(j)}} x_{kl,p}^{ij,q} \leq 1, \qquad \forall k \in J, \forall l \in J_k, \forall p \in P_k(l) \tag{10}$$

$$u_{kl}^p + \sum_{\substack{j \preceq l \in J_k \\ q \in P_k(j)}} x_{kj,q}^{kl,p} + \sum_{\substack{i \in J \\ i \neq k}} \sum_{\substack{j \in J_i \\ q \in P_i(j)}} x_{ij,q}^{kl,p} - \sum_{\substack{j \succeq l \in J_k \\ q \in P_k(j)}} x_{kl,p}^{kj,q} - \sum_{\substack{i \in J \\ i \neq k}} \sum_{\substack{j \in J_i \\ q \in P_i(j)}} x_{kl,p}^{ij,q} - z_{kl}^p = 0, \qquad \forall k \in J, l \in J_k, \forall p \in P_k(l) \tag{11}$$

$$\sum_{a \in W} w_{kl'}^{kl,a} + \sum_{\substack{j \preceq l \in J_k \\ q \in P_k(j)}} x_{kj,q}^{kl,p} + \sum_{\substack{i \in J \\ i \neq k}} \sum_{\substack{j \in J_i \\ q \in P_i(j)}} x_{ij,q}^{kl,p} = 1, \qquad \forall k \in J, l \in J_k, \forall l' \in Pred_k(l),$$

$$\forall p \in P_k(l) \cap C_k(l') \qquad (12)$$

$$m_{ij}^a + m_{kl}^a \geq M(w_{ij}^{kl,a} - 1) + 2, \qquad \forall i, k \in J, j \in J_i, l \in J_k,$$
$$a \in W \setminus \{U\} \qquad (13)$$

$$\sum_{a \in W} m_{ij}^a = 1, \qquad \forall i \in J, j \in J_i \qquad (14)$$

$$m_{ij}^a \leq \mathbb{1}_{cat(ij) \in cat(a)} \qquad \forall i \in J, j \in J_i, a \in W \qquad (15)$$

$$\sum_{j \in f} m_{fj}^a \leq 1, \sum_{j \in t} m_{tj}^a \leq 1, \qquad \forall a \in W \qquad (16)$$

$$m_{ke_k}^U = 1, \; m_{kl}^U = 0, \qquad \forall k \in J, l \in J_k \setminus \{e_k\} \qquad (17)$$

$$c_{ij} - v_{ij}^p - \sum_{a \in W} m_{ij}^a P_{ij}^a \geq 0, \qquad \forall i \in J, j \in J_i, p \in P_i(j) \qquad (18)$$

$$c_{kl} - c_{ij} - P_{kl}^a \geq M(w_{ij}^{kl,a} - 1), \qquad \forall i, k \in J, j \in J_i, l \in J_k, a \in W \setminus \{U\}$$
$$(\text{if } i = k : l \succeq j) \qquad (19)$$

$$v_{kl}^p - c_{kl'} - T_a^b \geq M(m_{kl'}^a + m_{kl}^b - 2), \qquad \forall k \in J, l \in J, l' \in Pred_k(l),$$
$$\forall p \in P_k(l) \cap C_k(l'), \; a, b \in W \qquad (20)$$

$$v_{kl}^p - \sum_{a \in W} m_{kl}^a T_L^a \geq 0, \qquad \forall k \in J, \forall l \in J_k, \forall p \in P_k(l) \cap B_k$$
$$(21)$$

$$v_{kl}^p - v_{ij}^q - \sum_{a \in W} m_{ij}^a T_a^b - \sum_{a \in W} m_{kl}^a T_b^a \geq M(x_{ij,q}^{kl,p} + m_{kl'}^b - 2), \qquad \forall i, k \in J, j \in J_i, l \in J_k, b \in W,$$
$$l' \in Pred_k(l), p \in P_k(l) \cap C_k(l'),$$
$$q \in P_i(j) \qquad (22)$$

$$v_{kl}^p - v_{ij}^q - \sum_{a \in W} m_{ij}^a T_a^L - \sum_{b \in W} m_{kl}^b T_L^b \geq M(x_{ij,q}^{kl,p} - 1), \qquad \forall i, k \in J, i \neq k, j \in J_i, l \in J_k,$$
$$p \in P_k(l) \cap B_k, q \in P_i(j) \qquad (23)$$

4   Results and Discussion

The model reliably calculates the minimal makespan for examples up to five jobs and 22 operations in total. Figure 2 shows an example with three workstations, two AGVs and five jobs with precedence graphs as shown in Fig. 1 (job 1 has the precedence graph on the left, jobs 2, 3, 4 and 5 the one on the right). The production system is a small-scale version of a real modular production system where each workstation has different capabilities. WS1 can glue and clinch, WS2 can weld and WS3 can clinch. Each coloured block means that the workstation or the AGV is currently active during this time interval: a workstation is processing an operation and an AGV is bringing a part to a workstation. The colour of the

block corresponds to the job, to which the operation or part belongs. The light blue "empty drive" boxes mean that the AGV is on its way to pick up a part but is currently not carrying anything. As all jobs need a gluing operating (cf. Fig. 1) and WS1 is the only workstation which can glue, the makespan cannot be shorter than the combined execution time of the gluing operations on WS1. Since WS1 is uninterruptedly executing the gluing operations of the jobs, it is evident that the makespan is minimal in this example.
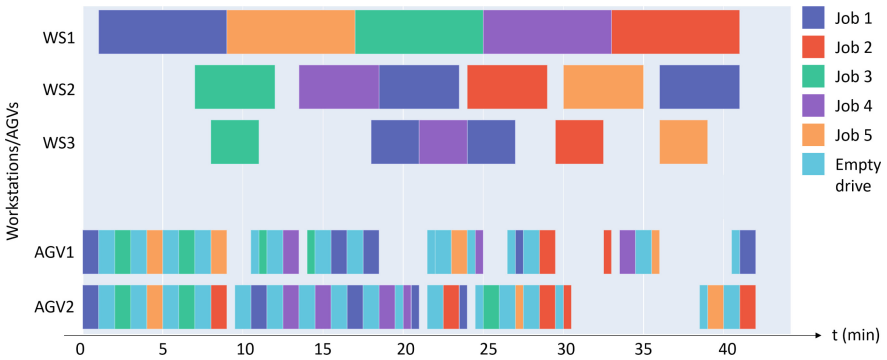
Fig. 2. An example producing five jobs, using three workstations and two AGVs.

This small example shows the major novelties of this model. The joining steps of the jobs are modelled as precedence graphs. This represents the flexibility of modular production much more adequately than serial operations. Also, the workstations have different capabilities and the parts are considered as an input.

The model deterministically calculates the minimal makespan for small examples but due to the NP-hard nature of the problem, it struggles with larger, more realistic use cases. The example from Fig. 2 is solved in less than a minute on a regular laptop with an Intel i5 processor (2.8 GHz) with 8 GB RAM. However, if the size of the problem (number of operations, jobs, AGVs and workstations) reaches a real-life BIW production magnitude, no deterministic exact solution is found within reasonable time. Therefore, in further work, the quality of the solutions which can be attained in acceptable time and an adequate definition of acceptable time will be researched. In addition, the authors will look into the possibility of adding further constraints posed by the production which might mitigate the calculation time increase for growing problem sizes.

Each AGV can only carry one part at a time. Since many operations require several parts, either several AGVs are necessary or the AGVs need to do multiple laps. This is obviously not ideal and will be looked into in future work.

Figure 3 shows the makespan and the AGV utilization rate for the example in Fig. 2 as a function of the number of AGVs. The makespan is significantly better with two AGVs in comparison to one single AGV. However, more than two AGVs do not improve the makespan at all. Therefore, two AGVs are deployed. To further shorten the makespan, a different production system is necessary.
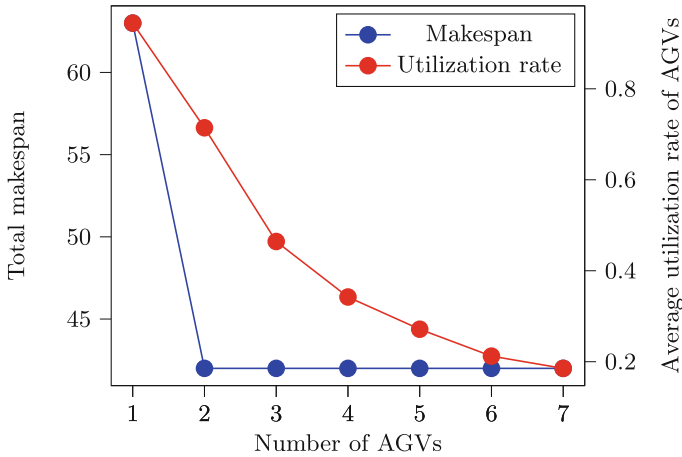
Fig. 3. Makespan (blue) and average AGV utilization rate (red) as a function of the number of AGV in the example from Fig. 2.

5   Conclusion

In this paper, the authors present a MIP model for the the modular BIW production. The model is inspired by the models by Fontes et al. [8] and Homayouni et al. [11] but adapted to meet the requirements posed by the BIW production. In particular, the model considers that the BIW production operations are not serial, but arranged as a precedence graph. In addition, the whereabouts of the parts as well as the different capabilities of the workstations are included.

The proposed approach is applied to an exemplary production scenario with five jobs and 22 operations which highlights the strengths of the model but also indicates its current limitations, arising from the NP-hard nature of the problem. In future work, the authors will further analyze the performance of the model, assess the quality of its solutions and explore possibilities to reduce the numerical complexity of the model in order to facilitate the scheduling of larger scenarios.

References

1. Casper, R., Sundin, E.: Electrification in the automotive industry: effects in remanufacturing. J. Remanufact. **11**(2), 121–136 (2020). https://doi.org/10.1007/s13243-020-00094-8
2. Kampker, A., et al.: Agile low-cost montage. In: Internet of Production für agile Unternehmen: AWK Aachener Werkzeugmaschinen-Kolloquium 2017, pp. 231–259. Apprimus Verlag, Aachen (2017)

3. Zenner, C.: Durchgängiges Variantenmanagement in der Technischen Produktionsplanung. PhD thesis. Universität des Saarlandes, Saarbrücken (2006)
4. Moon, D.H., Kim, D.D., Shin, Y.: Automotive body shop design problems using meta-models considering product-mix change and reconfiguration strategy. Appl. Sci. **11**(6), 2748–2768 (2021). https://doi.org/10.3390/app11062748
5. Hagemann, S., Stark, R.: Automated body-in-white production system design: data-based generation of production system configurations. In: ICFET 2018: Proceedings of the 4th International Conference on Frontiers of Educational Technologies, pp. 192–196. Association for Computing Machinery, Moscow (2018). https://doi.org/10.1145/3233347.3233373
6. Mayer, S., Endisch, C.: Adaptive production control in a modular assembly system based on partial look-ahead scheduling. In: 2019 IEEE International Conference on Mechatronics (ICM), pp. 293–300. IEEE, Ilmenau (2019). https://doi.org/10.1109/ICMECH.2019.8722904
7. Stastny, J., Skorpil, V., Balogh, Z., Klein, R.: Job shop scheduling problem optimization by means of graph-based algorithm. Appl. Sci. **11**(4), 1–16 (2021). https://doi.org/10.3390/app11041921
8. Fontes, D.B.M.M., Homayouni, S.M.: Joint production and transportation scheduling in flexible manufacturing systems. J. Global Optim. **74**(4), 879–908 (2018). https://doi.org/10.1007/s10898-018-0681-7
9. Ku, W.-Y., Beck, J.C.: Mixed integer programming models for job shop scheduling: a computational analysis. Comput. Oper. Res. **73**, 165–173 (2016). https://doi.org/10.1016/j.cor.2016.04.006
10. Bilge, Ü., Ulusoy, G.: A time window approach to simultaneous scheduling of machines and material handling system in an FMS. Oper. Res. **43**(6), 1058–1070 (1995). https://doi.org/10.1287/opre.43.6.1058
11. Homayouni, S.M., Fontes, D.B.M.M.: Production and transport scheduling in flexible job shop manufacturing systems. J. Global Optim. **79**(2), 463–502 (2021). https://doi.org/10.1007/s10898-021-00992-6
12. Klindworth, H., Otto, C., Scholl, A.: On a learning precedence graph concept for the automotive industry. Eur. J. Oper. Res. **217**(2), 259–269 (2012). https://doi.org/10.1016/j.ejor.2011.09.024
13. Walla, W.: Frühzeitige Produktbeeinflussung bezüglich Produktionsanforderungen im Karosserierohbau der Automobilindustrie. PhD thesis. Karlsruher Institut für Technologie (KIT), Karlsruhe (2015)

Robotic Assembly Line Balancing with Multimodal Stochastic Processing Times

Dawid Stade[(✉)] and Martin Manns

University of Siegen, Mechanical Engineering Department, PROTECH-Institute of
Production Technology, 57076 Siegen, Germany
`dawid.stade@uni-siegen.de`

Abstract. In this paper, a genetic algorithm for the robotic assembly line balancing problem (RALBP) is developed that supports multimodal stochastic processing times and multiple parallel-working robots per workstation. It has the objective to minimize the amount of workstations at a given production rate and probability limit for violating the cycle time (PL). The algorithm is evaluated on the BARTHOLD data set in a range of 1 % to 50 % for PL using an experimentally determined and a normal distribution for the task times. The increase of PL results in a shift of tasks from rear to front stations, because more tasks can be assigned to each station. The shift using normal distributed task times is stronger. This demonstrates the importance of realistic stochastic distribution assumptions. For practical applicability, more constraint types have to be included in the future.

Keywords: Assembly lines · Genetic algorithms · Robots · Stochastic process times

1 Introduction

Robots play an important role in high-volume assembly lines, because they are reliable and repetitively accurate [1]. One widely used joining technology for robotic assembly lines is resistance spot welding (RSW) [2]. RSW welds car body sheets rapidly and forms a welding nugget in between the parts. Shape and size of the nugget affects the strength of the welded joint. To avoid excessive destructive tests to inspect weld quality and to react to disturbances during the welding process, feedback control systems (FCS) are used (see [3]). Since the weld strength cannot be measured directly, correlating electrical variables such as the dynamic resistance are commonly employed. The FCS determines the time of current termination based on experimentally determined values of the measurement variable, this leads to variation in the welding times.

Figure 1 shows the distribution of weld times for one weld point. The FCS is programmed to extend the weld time up to a preset prolongation limit [4]. This leads to a major weld time mode at $-0.5\,\text{ms}$ to $0.5\,\text{ms}$ deviation from the predefined weld time of $450\,\text{ms}$, a center mode at about $7.5\,\text{ms}$ to $8.5\,\text{ms}$ and a closing mode at about $18.5\,\text{ms}$ to $19.5\,\text{ms}$ deviation. Similar distributions but

different positions and heights of the modes can be found for other weld points. Gaussian Kernel-Density Estimation (KDE) is used to approximate a Gaussian Mixture Model (GMM). Its probability density function is shown in the figure.

In order to take into account the effects of processing time distributions on the robotic assembly line in early planning phases, this work introduces a computational approach to consider multimodal stochastic processing times in the robotic assembly line balancing problem (RALBP). In this approach, the planner specifies a probability limit for exceeding the cycle time in the stations, which is then respected by the algorithm. The probability of exceeding the cycle time is thus known in contrast to deterministic approaches and can be taken into consideration in subsequent planning steps so that, for example, the utilization of the stations can be optimized in a controlled manner. The proposed approach is inspired by the balancing of labour-intense assembly lines with normal distributed processing times, see ref. [5,6].
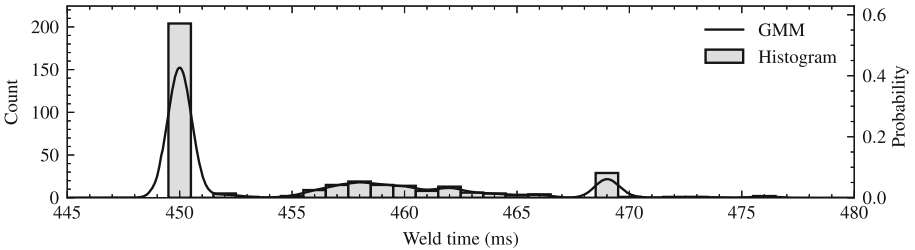
Fig. 1. Weld time distribution of a weld joint with 357 samples from car body construction

2 State of Research

The RALBP, first described by Rubinovitz and Bukchin [7], is an extension to the popular assembly line balancing problem (ALBP). The goal is to assign a collection of tasks with their corresponding processing times and precedence relationships to workstations and to find a suitable robot for each station to optimize an objective function without violating resource and technological constraints.

As Chutima [8] shows, the RALBP is predominantly solved using metaheuristics followed by exact algorithms and heuristics, because they show the highest effectiveness for practical sized problems.

Levitin et al. [9] present a genetic algorithm (GA) for the RALBP with the objective to minimize the cycle time at a given amount of serially arranged workstations (type II). The GA manipulates the order of the tasks, in which they are distributed to the stations. To avoid violation of their precedence constraints, the authors introduce a special reordering procedure to correct invalid solutions and a fragment reordering crossover as well as mutation operator. The actual assignment to stations and the selection of a robot is done in the decoding procedure.

3   Objective

The objective of this paper is to develop a method to address multimodal stochastic processing times that result from RSW process control using RALBP. In the context of car body construction, multiple robots per workstation are working on the same part in parallel. Multimodal processing time distributions have to be experimentally determined. The objective of the method is to minimize the amount of workstations. In addition, the influence of the distribution assumptions that are chosen to represent the task time variation will be investigated.

4   Genetic Algorithm

The proposed approach is based on the GA developed by Levitin et al. [9] and introduces a new decoding procedure. The assumptions for the RALBP made by Rubinovitz and Bukchin [7] are modified as follows:

1. The precedence relationships for all tasks are known and cannot be changed. They are represented by a precedence matrix $\boldsymbol{P}_{T \times T}$, where $p_{ij} = 1$ means, task i must be a predecessor of task j.
2. The duration of a task is distributed and approximated by variations of the probability density function in Fig. 1. Tasks cannot be subdivided.
3. The only limitations for assigning tasks to stations are the probability limit PL for violating the given cycle time and the precedence constraints in $\boldsymbol{P}$.
4. K robots are assigned to each station. There are no limitations for assigning a task to a robot other than the cycle time violation probability limit.
5. Loading, unloading and material handling times as well as setup and tool changes are neglected or are included in the task times.
6. The line is balanced for a single product.
7. There are no limitations for the availability of robots.

The notation used in this paper is described in Table 1.

4.1   Solution Representation

This paper uses a partition-oriented encoding scheme following Kim et al. [10]:

– Vector $\boldsymbol{v}$: Contains a permutation of the task elements. This vector is generated by the genetic algorithm.
– Vector $\boldsymbol{s}$: Contains pointers to elements in $\boldsymbol{v}$ to partition it into stations.
– Matrix $\boldsymbol{R}$: $K \times T$ binary matrix which contains the assignments of tasks to robots (if task i is assigned to robot k, r_{ki} is 1, otherwise 0).

4.2   Steps of the GA

The steps of the algorithm can be summarized as follows, for further details the reader is referred to [9]: i) generate a random valid initial population ii) decode and sort the population by their fitness using the decoding procedure in Sect. 4.3 iii) create offspring of some of the best chromosomes and replace the worst with them iv) randomly mutate some of the chromosomes v) jump back to step ii) for G times or stop after not improving for H times.

Table 1. Notation for the genetic algorithm

C	Cycle time
T	Number of tasks
e_i	Expectation value of the processing time of task i ($i \in (1, \ldots, T)$)
X_i	Random variable of the processing time of task i ($i \in (1, \ldots, T)$)
$\boldsymbol{P}_{T \times T}$	Binary matrix containing all precedence constraints of the problem statement ($p_{ij} = 1$ means, task t_i must be a predecessor of task t_j)
$\boldsymbol{v}$	Vector containing a sequence of tasks indices $v = (v(1), \ldots, v(T))$
$\boldsymbol{s}$, S	Vector of station indices, number of stations
K	Number of robots per station
$\boldsymbol{R}_{K \times T}$	Binary matrix to assign tasks to robots (if task t_i is assigned to robot k, r_{ki} is set to 1, otherwise 0)
G	Maximum number of generations
H	Limit of generations without an improvement of the best solution
PL	Given limit for the probability of violating the cycle time C
m	Control variable for decoding procedure ($1 \le m \le |v|$)
n	Pointer to the first task in the current station ($n \in \{1, \ldots, |v|\}$)
cp_k	Free capacity of the kth robot ($k \in (1, \ldots, K)$)
d	Robots ordered by their free capacity ($d = (d(1), \ldots, d(K))$)
t_i	Deterministic processing time of ith task ($i \in (1, \ldots, T)$)
$\mathcal{S}$	Free capacity in the last station

4.3 Decoding Procedure for the RALBP

The decoding procedure populates vector $\boldsymbol{s}$ and matrix $\boldsymbol{R}$, i.e. assign the tasks to workstations and robots. It considers a probability limit PL for violating the cycle time in each robot. The algorithm consists of the following steps:

D1 Set variables $m = 1$ and $n = 1$.
D2 Open a new station by adding the index of the current task m to station vector $\boldsymbol{s}$. Set $n = m$.
D3 Select element $v(m)$ if $m \le |v|$, otherwise stop the algorithm.
D4 Check, if the selected task element $v(m)$ fits into the station:
D4.1 If $n = m$, set $r_{1m} = 1$ and increase m to $m + 1$. Jump to step **D3**.
D4.2 Sort the robots by their descending free capacity $\boldsymbol{cp}$, see Eq. 1, which leads to a permutation $d = (d(1), \ldots, d(K))$.

$$cp_k = C - \sum_{a=n}^{m-1} r_{ka} \cdot e_{v(a)}, \; k \in (1, \ldots, K) \tag{1}$$

D4.3 For each robot $d(l)$ in $l \in (1, \ldots, K)$, do a monte carlo simulation to check if the probability to violate the cycle time of the tasks in the robot

including the random numbers $X_{v(m)}$ of the currently selected task $v(m)$ exceeds the given limit PL. If it does not, the current task fits on the robot. Break the loop and jump to step **D4.4**. If $l = K$ and no robot has been found to be able to complete task $v(m)$, jump to step **D2**.

D4.4 Set $r_{d(l)m} = 1$ and increase m to $m + 1$. Jump to step **D3**.

5 Results

The presented algorithm is evaluated on the basis of practical problems from car body construction. The *BARTHOLD* problem statement from the data sets by Scholl [11] is chosen because it matches the *F-Ratio* (0.7 to 0.8) and *WEST-Ratio* (15 to 25) (see Dar-El [12]), which we have found to be typical for car body construction problems. The cycle times and processing times of the problem are multiplied by 0.1 s, which results in cycle times of 56.4 s, 62.6 s, 70.5 s, and 80.5 s. It consists of 149 tasks with processing times within a range of 0.3 s to 38.3 s. For each test case, the probability limit of cycle time violations PL is gradually increased from 1% to 50%, the decoding procedure assigns the tasks in such a way that PL is not exceeded in the stations. A normal distribution ($\mathcal{N}$) is derived from the expectation value (454.98 ms) and standard deviation (6.67 ms) of the weld time distribution in Fig. 1 and compared to the approximated GMM. The *BARTHOLD* problem contains deterministic task times t_i. The random variables X_i are linearly scaled so that their expectation values match t_i.

The gradually increase of the probability limit PL allows to assign more tasks to each station, resulting in a shift from rear to front stations and an increasing free capacity in the last workstation. At a certain probability limit, the shift is sufficient to save one station. The position of this limit is essentially determined by the problem and the process time distribution of the tasks. The boxplot in Fig. 2 shows the free capacity of the last station $\mathcal{S}$ for all test cases, where the boxes contain the values of $\mathcal{S}$ for PL in a range from 1% to 50%.
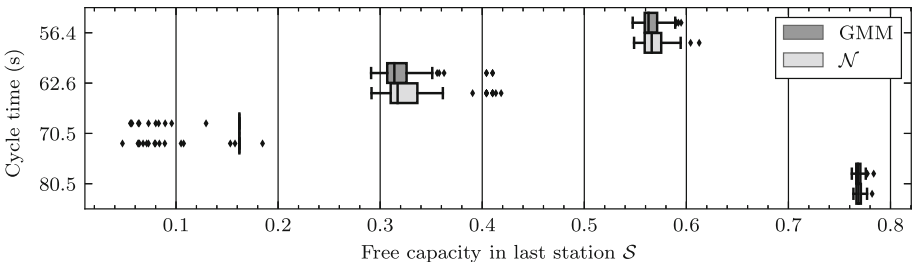
Fig. 2. Free capacity in the last station $\mathcal{S}$ for cycle times 56.4 s, 62.6 s, 70.5 s and 80.5 s (WEST-Ratios 14.8, 16.44, 18.5 and 21.14) and probability limits PL from 1% to 50%

In all scenarios except at a cycle time of 70.5 s, the range of $\mathcal{S}$ is wider with the normal distribution. The interquartile range is between 12 % ($C = 80.5\,s$)

and 56% ($C = 62.6\,s$) larger than with GMM. Normal distributed processing times therefore save more capacity.

6   Conclusion and Future Research

A genetic algorithm considering stochastic processing times and multiple parallel-working robots per workstation for the RALBP was presented. The algorithm was evaluated on the basis of practical test problems using an approximated GMM and a normal distribution to represent stochastic task times. The results show that with a normal distribution more capacity in the last station could be saved by increasing the probability limit for violating the cycle time, which can lead to savings of workstations at lower risk. It can be concluded from this that the choice of distribution and the underlying problem play a decisive role. Quality of the input data as well as appropriate and process adequate distribution assumptions strongly affect the results. Therefore inaccuracies may lead to false assumptions and therefore to lasting effects on following operations.

For practical application of the proposed approach, some features are missing and may be subject to future research: i) Support for in-station constraints, which result from geographical positions of the tasks and depend on the actual set of tasks being assigned to the station and ii) a procedure to preserve precedence constraints in the stations and on the parallel-working robots.

References

1. Sule, D.R.: Manufacturing Facilities: Location, Planning and Design, 2nd edn. PWS Publishing Company, Singapore (1998)
2. Stiebel, A., et al.: Monitoring and control of spot weld operations. SAE Techn. Paper Ser. (1986)
3. Podržaj, P., et al.: Overview of resistance spot welding control. Sci. Technol. Weld. Join. (2008)
4. Bosch Rexroth AG. Rexroth PSI 6xxx - UI regulation and monitoring
5. Ağpak, K., Gökçen, H.:. A chance-constrained approach to stochastic line balancing problem. Eur. J. Oper. Res. (2007)
6. Lolli, F. et al.: Stochastic assembly line balancing with learning effects. IFAC-PapersOnLine (2017)
7. Rubinovitz, J., Bukchin, J.: Design and balancing of robotic assembly lines. In: Proceedings of the Fourth World Conference on Robotics Research (1991)
8. Chutima, P.: A comprehensive review of robotic assembly line balancing problem. J. Intell. Manufact. (2020)
9. Levitin, G., Rubinovitz, J., Shnits, B.: A genetic algorithm for robotic assembly line balancing. Eur. J. Oper. Res. **168**(3), 811–825 (2006)
10. Kim, Y.K., Kim, Y.J., Kim, Y.: Genetic algorithms for assembly line balancing with various objectives. Comput. Industr. Eng. **30**(3), 397–409 (1996)
11. Scholl, A.: Data of assembly line balancing problems. In: Schriften zur Quantitativen Betriebswirtschaftslehre Technische Hochschule Darmstadt (1993)
12. Dar-El, E.M.: MALB-a heuristic technique for balancing large single- model assembly lines. AIIE Trans. **5**(4), 343–356 (1973)

Elementary Welding Operations
for Automatic Robot Programming

Ioan-Matei Sarivan[1(✉)], Jørgen S. Larsen[2], Ole Madsen[1],
and Brian V. Wæhrens[1]

[1] Department of Materials and Production at Aalborg University, Fibigerstræde 16,
Aalborg 9220, Denmark
`ioanms@mp.aau.dk`
[2] PDMTechnology A/S, Niels Jernes Vej 10, Aalborg 9220, Denmark

Abstract. Highly-skilled engineers are needed to reprogram a welding-robot upon a new variant or product introduction. This paper addresses the complexity of the welding-robot programming process by proposing and demonstrating the theoretical concept of "elementary welding operation" (EWO). The EWO is a fundamental building block for robot-welding programs containing most of the data necessary to generate a program automatically. The EWO is a descriptive data structure for the welding process, comprehensible by humans and machines. A technical proof of concept is put together and tested. The results indicate that the programming of the welding robot can be automated as part of a one-piece information flow based on a product model. Further development is required to ensure the system can handle complex weldment geometries, multi-pass welding, and proper validation of the robot program upon deployment.

Keywords: robot programming · welding robot · elementary welding operation · digitalisation · data structure

1 Introduction

Welding is the most significant application for industrial robotics, with over 50% of the existing industrial manipulators being commissioned in welding tasks [1]. The usage of welding-robots comes with many advantages in terms of workers' health, increased quality assurance, shorter lead-time, and reduced production cost. The inflexible nature of welding-robots is seen, however, as a problem in engineering-to-order (ETO) and manufacturing-to-order (MTO) production environments, where high variation across produced items occurs. Each product is unique and requires the welding robot to be reprogrammed for every new incoming order. High-skilled robotics engineers are required to reprogram the welding-robots in order for a new product to be produced. The reprogramming process requires human, financial, and time resources which drive ETO and MTO companies away from the prospect of using welding-robots. [2]

This paper proposes the theoretical concept of *elementary welding operation* (EWO). The EWO is a data structure meant to be used as a fundamental building block for the automatic programming of welding robots. The data contained inside an EWO is extracted automatically from a product instance. The data can be further processed to automatically program a welding-robot [3]. The digital integration of engineering and pre-production operations is obtained, and human input is reduced. The novelty of the implementation stands in having the geometry of the weldments programmed automatically based on unique product models. Experiments presented in this paper are limited to the validation of the weldment's geometry in an off-line programming (OLP) virtual environment using 6DOF welding-robots. The results show that EWO can potentially reduce the necessary time and cost involved in programming welding robots by automating this process.

The following meaning is attributed to the terms:

– Product Model: a parametric CAD entity that is configurable using a product configurator tool containing associated KBE (knowledge-based-engineering) and welding process data [3].
– Product Instance: a CAD entity obtained following the configuration of the Product Model using a product configurator by providing values for design parameters.

The paper is structured as follows: related work in the field of automatic robot programming and state-of-practice are presented in Sect. 2. Section 3 presents the overall structure of the elementary welding operation. Section 4 contains the technical description of a proof-of-concept implementation using the EWO concept. Results of experiments with the technical system are presented in Sect. 5. The system is bench-marked against the state-of-practice method. Conclusions and final considerations towards future development are mentioned in Sect. 6.

2   Related Work

Related work was found to be documented in the literature regarding systems that automatically extract welding geometry from CAD models and generate robot programs. The fundamental information used to program a welding robot is the geometry of the item to be welded, which is generally stored in a CAD model. The geometry of the item can be extracted from *STEP* (Standard for Exchange of Product Data) format CAD files by accessing the relevant data structures which define the item: vertices coordinates and vectors [4]. The STEP file format is generally used for CAD applications and is documented under the ISO10303 ISO standard.

In subtractive and additive manufacturing, parametric robot programming is a method widely used to handle large variations in the geometry of the item that is produced. Such methods are often found in systems like CNC machines [5] and 3D printers [6]. However, in the case of CNC and 3D printing machines, the orientation of the tool is often not an issue. Still, parametric programming

allows for facile adjustment of existing robot programs, where the orientation of the tool is relevant [7].

In order to facilitate the geometry extraction of the product from CAD models, Lobov and Tran (2020) propose an object oriented automated method for product design. The method makes available the necessary functions to create and extract the geometry of the intersection edges of the solid bodies where it is possible for a weldment to be located [8,9]. Prescott et al. (2020) further developed the method by making it possible for geometry data for the weldment to be added on top of the CAD model at the identified intersection edges. This data is also known as KBE data (knowledge-based-engineering) [10] and can also contain process information about the weldment (e.g., filler-wire material). The data extracted from the CAD-KBE models is used to generate welding-robot trajectories based on the identified weld paths.

Alongside the geometry data of the item, the process data (e.g., torch travel angle, welding angle, torch travel speed) is needed to perform a weldment. Schmidt et al. propose a hierarchical data structure where process data like the welding current and gas flow are stored [11]. Mohammed et al. use the structure of the STEP CAD files to store process data like the orientation of the torch, welding speed, gas flow, and others [12].

Several methods and system architectures for programming of welding-robots based on CAD models were identified in the literature as related work, focusing on creating an end-to-end seamless pipeline from CAD to robot program. Some of these initiatives require human intervention to validate the final robot program [13], and others are completely automatic [14,15]. The EWO method presented in this paper brings as a contribution a data structure meant to facilitate the automatic transfer of geometry data and process data from a CAD system to an OLP (offline-programming) system.

State-of-Practice

A state of practice study was done in collaboration with a partner ETO/MTO company where products vary order by order. The company is involved in the metal-fabrication industry, and they use welding-robots with the GMAW welding process. The robot-programming is handled by robotics engineers who use OLP software. The product instance is imported into the OLP software. Any overlaying process data is lost during the import procedures, including the locations of the weldments, and only the geometric CAD model is available in the OLP system. The robotics engineers must manually program the locations of the weldments and the process related data, using information available on the item's technical drafts, or using the product instance available in the source CAD system. For products where the geometry of the weldments is similar, and only the scale of the product differs, parametric robot programming is used, and older robot programs are edited to shorten the time needed to reprogram the robot. The programming process time varies between one to five days of work (considering a day of work has eight hours), depending on the complexity of the product.

3   Elementary Welding Operation

The proposed "elementary welding operation" EWO concept is designed to establish the data conduit between a product instance and the welding-robot program necessary for manufacturing that product. The EWO is part of a hierarchical data structure organizing the KBE data and the process data needed to program the welding-robot. The following assumptions are made regarding the welding-robot program:

1. The welding process goes through a number of quasi-stationary states
2. Each state is prismatic and can be represented using an EWO (Fig. 1)
3. A welding with complex geometry can be realised by combining a number of EWO (Fig. 2)

(a) Product instance with overlaid weldment KBE data

(b) Product instance imported in OLP tool
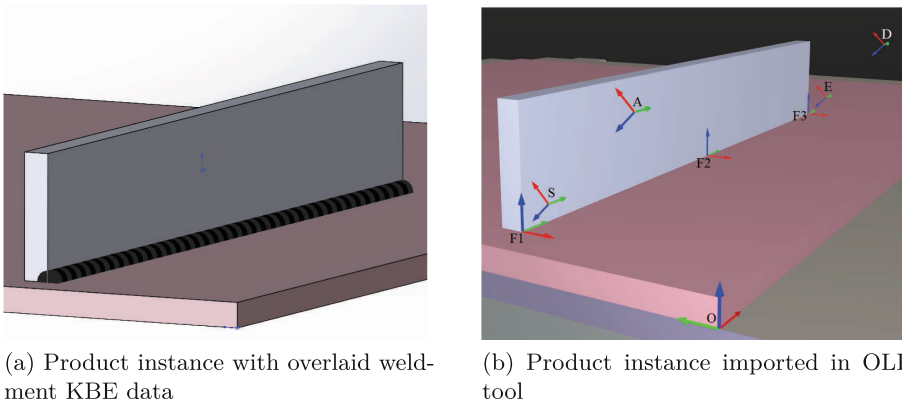
Fig. 1. Visual representations of a product instance (inside the SolidWorks CAD software) and a product instance imported in the RoboDK OLP software. The product is a fillet joint with a prismatic weldment.

The data structure containing EWO is represented in Fig. 3. The hierarchical root is named "Weldments" and is composed of at least one "Weldment." In the context of the data structure, a "Weldment" results from a continuous motion of the torch along the welding seam, together with the approach and departure motions before and after starting to weld. The approach and the departure motions are defined using the *approach pose* and a *departure pose* as seen in Fig. 3. A "Weldment" contains at least one Elementary Welding Operation. The EWO is always a straight line and represents the indivisible element that composes the process of performing a robotic weldment.
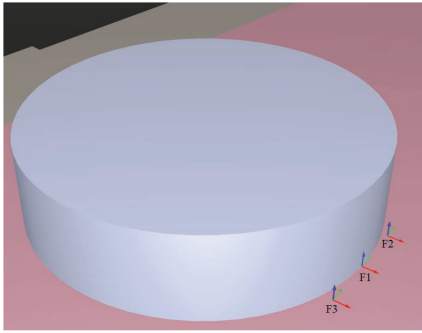
(a) Auxiliary frames for EWO contruction on a curved weldment seam
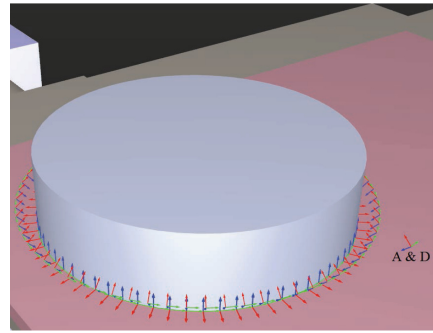
(b) Frames are added for each tessellation point of the curve where an EWO is located

Fig. 2. Visual representation of round shaped object. For each tessellation point of the curve, a coordinate frame is used to represent an EWO.
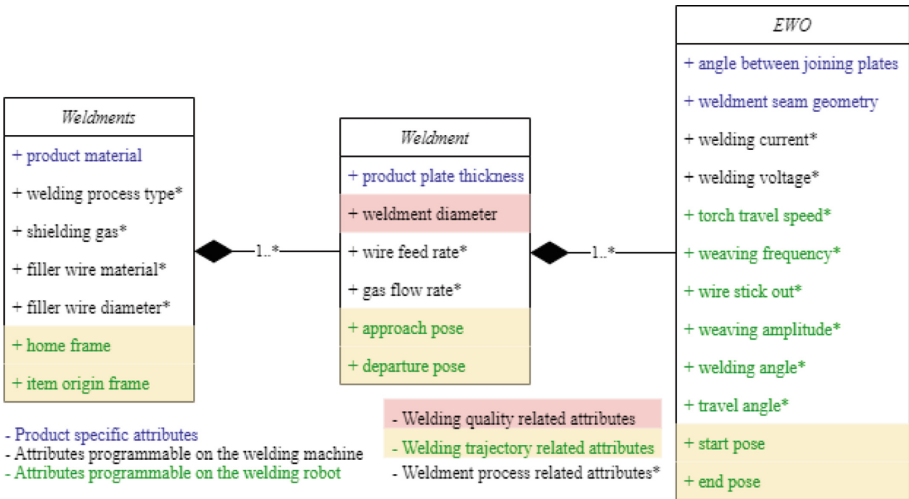
Fig. 3. UML diagram containing the data structure of welding program composed of EWOs. The attributes at each hierarchical level are classified using several criteria.

The attributes contained by each data class element of an EWO program can be observed in Fig. 3. The welding system considered in this paper is composed of a welding robot and a welding machine, each having their own controller. The attributes are organised at the level of the data structure as follows:

- "Weldments": attributes based on the material of the product which stay unchanged during the whole welding process.
- "Weldment": attributes that can vary from weldment to weldment based on the thickness of the plate, but kept constant while welding.

– "EWO": attributes that are variable while the welding torch is active and depend on the angle between the plates and the overall weldment seam geometry of the product.

The attributes are further classified into three categories. "Product," "Welding Machine," and "Welding Robot." The "Product" attributes are the properties of the item to be welded and are marked with blue text. These can be extracted from the CAD model and the overlaid KBE data. The attributes of the "Welding Machine" (black) and the "Welding Robot" (blue) are programmed on each machine based on the product's properties.

As observed in Fig. 3, the attributes related to the geometry of the robot trajectory are highlighter with yellow. The geometry attributes do not influence the quality of the weldment, only the path that is being followed by the welding torch. The geometry attributes are determined based on the geometry of the item that is extracted from the product instance. Process-related attributes do influence the quality of the weldment. A weldment quality parameter considered in this paper is the *weldment diameter*.

The main geometric characteristic of the EWO is a straight line. Therefore, when a weldment is composed of only one straight line, it contains one single EWO with a *start* and an *end pose*. When the weldment has a curved geometry, the "weldment" data structure is composed of n EWOs, where n is the number of tessellation points of the curve as observed in Subfigure 2b. For the intermediary EWOs contained in a "Weldment" data structure, only the start poses are considered for the welding-robot program. The values for the rest of the variables are contained in look-up tables or parametric rules, which are based on the product's material and plate thickness.

To configure the EWO *start* and *end pose*, the example in Fig. 1 is considered. The *item origin frame* represents the frame in which the geometry of the item is defined, marked with O in Subfigure 1a. The origin frame is imported from the product instance. The *approach pose* (marked with A in Subfigure 1a) and the *departure pose* (marked with D in Subfigure 1a) represent the approach/departure positions and orientations (poses) for the torch before and after engaging in a weldment. The *start pose* (marked with S) and *end pose* (marked with E) represent the start and end poses of the torch at the start and the end of the EWO.

To determine the approach/departure and the start/end poses for the torch, Horn's absolute orientation method is used [16]. The method makes it possible to determine the transformation matrix from one reference frame to another by having the coordinates of three points as input defined in both frames. First, reference frames are created along the weldment (marked with $F1$, $F2$ and $F3$ in Subfigure 1a). The positions of the frames are set using the coordinates of the tessellation points that compose the weldment. The coordinates of the tessellation points are expressed in relation to frame O. For a straight weldment, three frames are considered, at the start $F1$, at the middle $F2$, and at the end $F3$.

The frames have the Y axis along the weldment seam and the X axis along the horizontal surface. The linear transformation from frame O to frame $F1$

is determined by considering the matrices A and B. Matrix A contains the coordinates of frames $F1$, $F2$ and $F3$ represented in relation to frame O. The matrix B contains the coordinates of the frames in relation to frame $F1$. The matrices can be observed in Eq. 1. The x and z coordinates of $F2$ in relation to $F1$ are set to be 0, while the y coordinate is the distance between the two frames. This specifies that the y axis of the frame is along the weldment seam. The same procedure applies for $F3$.

$$A_O = \begin{bmatrix} x_{F1} & x_{F2} & x_{F3} \\ y_{F1} & y_{F2} & y_{F3} \\ z_{F1} & z_{F2} & z_{F3} \end{bmatrix} \quad B_{F1} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & dist(F1, F2) & dist(F1, F3) \\ 0 & 0 & 0 \end{bmatrix} \quad (1)$$

In the case of a weldment along a curved edge, the Y axis is tangent with the curved edge and makes contact with at least one tessellation point of the weldment. As observed in Subfigure 2a, frames $F2$ and $F3$ are placed along the curve's tangent, $F2$ on the positive side and $F3$ on the negative side, therefore the $dist(F1, F3)$ term of Matrix B in Eq. 1 will have the sign changed. The procedure is repeated for each tessellation point of the weldment's curve, as can be observed in Subfigure 2b.

With the input A and B matrices, the Horn's method yields the transformation matrix "T" from the "O" frame to the "F1" frame of each tessellation point. Therefore, the whole geometry of the weldment can be described mathematically in relation to "O". In order to express the start/end poses and the approach/departure poses the welding parameters can be used. The wire stick out s, the welding angle α, the travel angle β, and the approach distance d are considered. The matrix transformation from the frame "F" of the tessellation point to the start pose is displayed in Eq. 2. The same transformation can be used to determine the end pose and the approach/departure pose by replacing the s term with d. The results of the transformation applied for start/end poses, and approach/departure poses can be observed in Fig. 1 for a straight fillet weldment and Fig. 2 for a curve-shaped weldment.

$$\begin{matrix} {}^{F1}_{S}T = \end{matrix} \begin{bmatrix} -cos(\alpha) & -sin(\alpha)sin(\beta) & -sin(\alpha)cos(\beta) & s*sin(\alpha)cos(\beta) \\ 0 & cos(\beta) & -sin(\beta) & s*sin(\beta) \\ sin(\alpha) & -cos(\alpha)*sin(\beta) & -cos(\alpha)*cos(\beta) & s*cos(\alpha)cos(\beta) \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2)$$

4 Technical Implementation

To implement the EWO theoretical concept, the XML file format was chosen to store the data represented in Fig. 3. The SolidWorks CAD software was selected to create product instances based on product models that store KBE data and process data [3]. For robot programming, the RoboDK OLP tool was selected as welding programs can be created for a variety of welding-robot vendors.
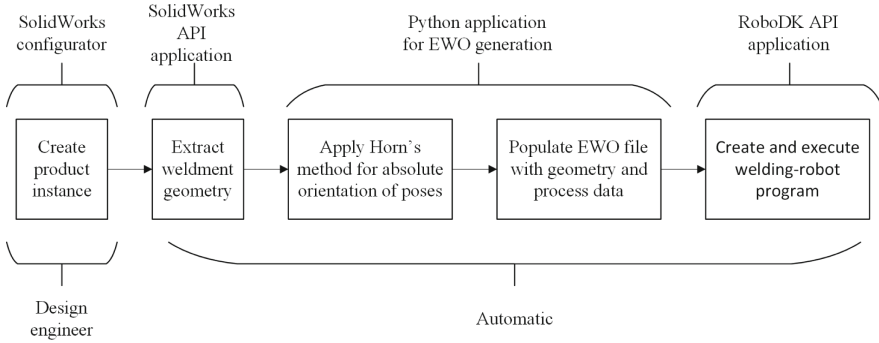
Fig. 4. Implementation pipeline using elementary welding operations.

The EWO concept enables seamless data transfer between the two software environments by using C#-Python application specially developed for this purpose. The implementation is summarised in Fig. 4.

With the EWO data structure in place, the RoboDK API is used to create TCP targets. A collision check is executed for each target in order to create a collision-free weldment trajectory for the welding torch. The process is automatic and yields the robot program that can be either executed in real-time or transferred to the robot's controller for execution.

5 Experiments and Results

Experiments were conducted using the presented technical implementation to test the viability of the EWO theoretical concept. The testing criteria used to evaluate the results are the number of engineers needed to operate the required software and the time required to generate a welding program for a welding robot. The implementation was compared against the state-of-practice method at the partner company. By using the implementation pipeline from Fig. 4, weldment geometry data is automatically extracted from the product instance (Fig. 5, Subfigure 5a) which is turned into TCP goal poses for the welding torch (Subfigure 5b).

The EWO method allows for the automatic import of most of the geometry and process data (welding angle and travel angle) needed to generate the welding-robot program without further intervention from engineers to fine-tune the positioning of the welding torch. Once the product instance is generated, the design engineer only needs to initiate the Python software described in the previous section, which will automatically generate the welding-robot program based on the EWO data. The time required to generate the program can take up to a few minutes, depending on the complexity of the product and the performance of the computer running the software.
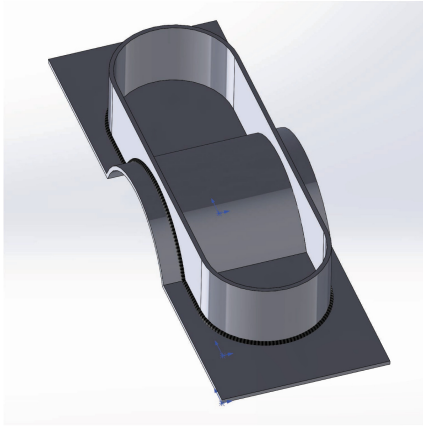
(a) Product instance generated using a product configurator [3]. The overlaid KBE data can be observed.
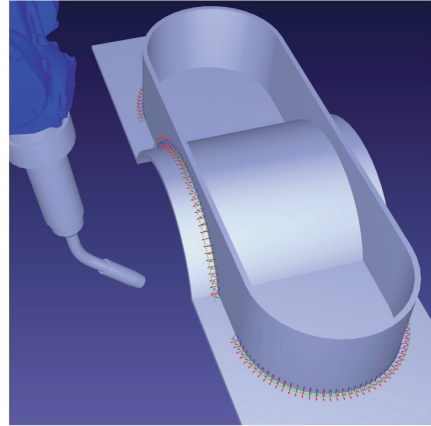
(b) Visual representation of the product inside the OLP system with EWO data.

Fig. 5. EWO welding program is obtained inside the RoboDK OLP system using the welding geometry and process data imported from the product instance.

6   Discussion and Conclusion

In this paper, a novel method for structuring geometry and process data for automatic programming of welding robots was presented. The initiative is aligned with previous efforts for creating digital tools that reduce the trade-off between automation and flexibility in MTO, and ETO environments [19]. Experiments show that the EWO theoretical concept can support technical implementations that enable the reduction of time needed to program welding-robots and the amount of skilled human resources required.

Further tests are required to determine how the complexity of the product affects the performance of the implementation, and how advanced geometries can be handled (e.g. multi-pass welding). The EWO shows promising potential as a platform for welding process-data. It is, therefore relevant to investigate further if the data structure presented is sufficient to support the parameters required to perform other welding procedures besides the GMAW procedure considered in the experiments conducted in this paper. Transferability to other applications where geometric discretization is possible (e.g., 3D printing) is also relevant to be investigated, especially where the orientation of the tool is relevant.

References

1. Sprovieri, J.: New technology for robotic welding. Assembly magazine (2016). https://www.assemblymag.com/articles/93555-new-technology-for-robotic-welding. Accessed 9 May 2022
2. The Benefits of Robotic Welding in Small Business. Robotics and Automation News (2020). https://roboticsandautomationnews.com/2020/02/20/the-benefits-of-robotic-welding-in-small-businesses/30181/. Accessed 9 May 2022
3. Sarivan, IM., Madsen, O., Waehrens, B.V.: Towards automatic welding-robot programming based on product model. In: Towards Sustainable Customization: Bridging Smart Products and Manufacturing Systems. CARV MCPC 2021 2021. Lecture Notes in Mechanical Engineering, pp. 174–181. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-90700-6_19
4. Sivakumar, S., Dhanalakshmi, V.: An approach towards the integration of CAD/CAM/CAI through STEP file using feature extraction for cylindrical parts. Int. J. Comput. Integr. Manuf. **26**(6), 561–570l (2013)
5. Sigrid, B.C., Johannes, B.: Parametric robot control: integrated CAD/CAM for architectural design. In: ACADIA Proceedings (2011)
6. Kuipers, T., et al.: A framework for adaptive width control of dense contour-parallel toolpaths in fused deposition modeling. Comput. Aided Des. **128**, 102907 (2020)
7. Feucht, T., Lange, J., Erven, M., et al.: Additive manufacturing by means of parametric robot programming. Constr. Robot. **4**, 31–48 (2020)
8. Lobov, A., Tran, T.A.: Object-oriented approach to product design using extended NX open API. Procedia Manuf. **51**, 1014–1020 (2020)
9. Tran, T.A., Lobov, A., Kaasa, T.H., et al.: CAD integrated automatic recognition of weld paths. Int. J. Adv. Manuf. Technol. **115**, 2145–2159 (2021)
10. Prescott, S.A.O., Tran, T.A., Lobov, A.: Automatic weld path definition in CAD. Procedia Manuf. **51**, 478–484 (2020)
11. Schmidt, A., et al.: A generic data structure for the specific domain of robotic arc welding. Procedia CIRP **72**, 322–327 (2018)
12. Mohammed, S.K., Arbo, M.H., Tingelstad, L.: Constraint Identification from STEP AP242 files for automated robotic welding. In: 2021 IEEE 12th International Conference on Mechanical and Intelligent Manufacturing Technologies (ICMIMT), pp. 277–282 (2021)
13. Bedaka, A.K., Lin, C.-Y.: CAD-based offline programming platform for welding applications using 6-DOF and 2-DOF robots. In: 2020 International Conference on Advanced Robotics and Intelligent Systems (ARIS), pp. 1–4 (2020)
14. Jiang, Q.X., Wang, S.H., Wei, X.Q., Hou, R.S., Chang, B.H., Du, D.: Model-driven online correction welding robot automatic programming system. In: IOP Conference Series. Materials Science and Engineering, vol. 504, no. 1 (2019)
15. Shen, W., et al.: A welding task data model for intelligent process planning of robotic welding. Rob. Comput. Integr. Manuf. **64**, 101934 (2020). https://doi.org/10.1016/j.rcim.2020.101934
16. Horn, B.K.P.: Closed-form solution of absolute orientation using unit quaternions. J. Opt. Soc. Am. A **4**, 629–642 (1987). https://doi.org/10.1364/JOSAA.4.000629
17. Matt, J.: Absolute Orientation - Horn's method (2022). https://www.mathworks.com/matlabcentral/fileexchange/26186-absolute-orientation-horn-s-method. MATLAB Central File Exchange. Accessed 16 May 2022
18. https://github.com/MateiSarivan/EWO

19. Bejlegaard, M., Sarivan, I.-M., Waehrens, B.V.: The influence of digital technologies on supply chain coordination strategies. J. Global Oper. Strateg. Sourcing **14**(4), 636–658 (2021)

Influence of the Joining Force on the Nugget Diameter During Resistance Spot Welding of Aluminum Materials

Jonas Pestka[1(✉)] and Stefan Weihe[2]

[1] Mercedes-Benz AG, Benz-Str. 10, 71063 Sindelfingen, Germany
`jonas.pestka@mercedes-benz.com`
[2] Institute for Materials Testing, Materials Science and Strength of Materials (IMWF), Pfaffenwaldring 32, 70569 Stuttgart, Germany

Abstract. Ever new requirements for the environmental compatibility and safety of products do not stop at the automotive industry. This results in ever shorter time windows for the pre-series processes. For this reason, time-consuming preliminary tests must be minimized to ensure an accelerated start-up and production that is as trouble-free as possible. This maxim applies to all parts of the production of an automobile, but especially to the body shop, where just-in-time production has been the norm for years and any downtimes are particularly severe. In order to meet these requirements, it is advantageous to understand or predict the result of a production process as well as possible.

This paper deals with resistance spot welding of aluminum and aims to generate knowledge about the relationship between setting parameters and target value. Since resistance spot welding has a large number of setting parameters and a comprehensive investigation of these would go beyond the scope of this study, the focus is on the joining force and the features to be generated from it. The target parameter is the nugget diameter of the joining points. The data are determined experimentally and evaluated with the help of Machine Learning (ML) approaches. The correlations are presented in graphical form. The quantification of the influence of the joining force on the nugget diameter is expected to lead to a reduction of tests and thus to a gain in time.

Keywords: resistance spot welding · aluminum · welding force · machine learning methods

1 Introduction

This study deals with the influence of the change in joining force during the resistance spot welding (RSW) process of aluminum on the nugget diameter. Since the setting parameters can be recorded throughout the process and evaluation methods for large amounts of data are now established and available, such an investigation is appropriate.

The two main welding parameters in RSW are the joining force and the main current in the process flow. Since the influence of the main current on the joining result has already been well investigated, this paper focuses on the joining force. More precisely, on the change of the joining force during the process.

Since pure resistance spot welding is hardly ever used in automotive body construction, the process variant of resistance spot weld bonding is used to describe a realistic process. The required features and target data are generated in the form welding tests. The basis for the investigation is a full-factorial test plan of two Material-Thickness-Combinations (MTC). These consist of the same material and the same thickness in order to exclude or minimize influences of the alloy or the sheet thickness ratio. In the evaluation, different Machine Learning model approaches are tried out to test which model is most suitable. Then a model is chosen and the relevant force related features are determined. The evaluation of the most important features is performed by a second degree regression using Machine Learning again.

2   Hypothesis and Method

The welding process is divided into 3 main components: the squeeze time, the weld time and the forge time. During the squeeze time, the welding force is applied and maintained. This enables a defined surface pressure throughout the process. In the course of the weld time, the spot weld is generated. First, the joint is prepared by means of a preheating current, and while the main current is flowing, the base metal melts. During the forge time, the joining force is still present and the welding lens is given its final shape. Since the material is molten at this time, the force curve changes significantly here. This is due to the type of control (force control active; position of the electrodes variable). However, the alloy, the sheet thickness, the amount of joining force and many other factors have an influence on the change of the force signal at this point in the welding process.

From this fact, it is hypothesized that features related to the force signal can predict the nugget diameter.

3   Literature Review

The joining force is one of the most important parameters in resistance spot welding [1]. It ensures a positional fixation of the joining partners before the start of the weld current and thus also influences the electrical contact and the current flow at the joint [2]. In addition, this parameter also significantly determines the shape and quality of the weld lens during cooling. For example, notches in the joining zone can be minimized or even prevented by a clever choice of welding force [2, 3].

The Automotive Research Centre was able to prove in experiments that a joining force of 3 kN is sufficient to produce spatter-free aluminum joints. Results that meet all quality requirements, however, require 6 kN [4]. In the study by Schmal, an optimum joining force of 8 kN is referred to for a two-sheet joint in order to meet the required quality standards [5]. Al Quran et al. also quantified the influence of the joining force on the nugget diameter and the resulting strength [6].

Machine Learning algorithms have already been successfully applied in the field of joining technology. For example, Kim et al. successfully used regression to optimize parameters for laser welding, which resulted in a significant improvement of the joining result [7]. Yang et al. also succeeded in determining the influence of various parameters on the joining result and thus achieving more optimal results. However, his results are related to arc welding [8].

However, ML methods have also been used for resistance spot welding. For example, Yu used the logistic regression of the power signal in 2015 to estimate the quality of joining points [9]. Zhou et al. went one step further and applied 3 ML methods to industrial data in order to predict the quality of the expected joining points [10].

The following investigations build on and extend all the findings just mentioned. This is done by focusing on the force signal and evaluating whether it is suitable as a predictor.

4   Setup

4.1   Material

In modern car body-in-white construction, natural ageing AlMg alloys (5000 series) and artificial ageing AlMgSi alloys (6000 series) are primarily used. Series 5000 alloys are characterized above all by good corrosion resistance. However, intergranular corrosion may occur at higher service temperatures [11]. The 6000 series alloys offer a good alternative to this and their technological properties can also be adjusted via the artificial ageing process.

In the course of this investigation, two Material-Thickness-Combinations (MTCs) are considered. Both use the alloy AL6-HDI-TZ-U [12] and are therefore homogeneous. In the first one, sheets of thickness 1.5 mm are used whereas the second one consists of samples of thickness 3.0 mm. The chemical composition is shown in Table 1. The suffix TZ describes the coil treatment condition (Ti & Zr with a coating weight of 2.0–8.0 mg/m^2) and the suffix U refers to the intended use of the components made from it (Unexposed). The material is in the solution heat treated and naturally aged condition (T4) [13]. The samples were available in 88×500 mm format.

Table 1. Chemical composition of the alloy AL6-HDI [12]

Si [%]	Fe [%]	Cu [%]	Mn [%]	Mg [%]	Cr [%]	Zn [%]	Ti [%]	V [%]
≤0,15	≤0,35	≤0,25	≤0,30	≤0,90	≤0,15	≤0,25	≤0,15	≤0,10
Other admixtures individually ≤0.05%, total ≤0.15%								

4.2   Equipment and Execution

The following equipment was available for the welding tests (Table 2):

Table 2. Investigation equipment

Equipment	Product name
Adhesive	DUPONT Betamate 1640
Application robot	KUKA LBR iiwa 7 800
Processing equipment for adhesive	BEYER & OTTO TS-215/HN
Joining robot	KUKA KR 270 R2700 ultra
Welding gun	NIMAK Robot C-Welding Gun
Robot cabinet	BOSCH WSC70RC-0404.25-L1-2-250-1450-11N-WT
Welding controller	BOSCH PRC 7000
Caliper	MAHR MarCal 16 EWRi

The production of the welding samples was done in several steps. The sequence is as follows (Fig. 1):
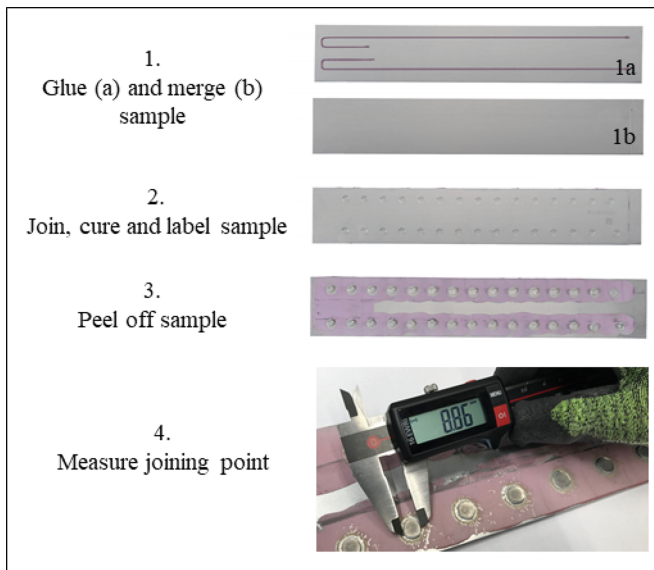
Fig. 1. Sequence of joining tests and data acquisition

It should be noted that the two sheets of sample overlap by 10 mm and 30 points have been joined on them in two rows. To prevent shunts, the points are spaced 30 mm apart and 55 mm between the rows (see Fig. 2).
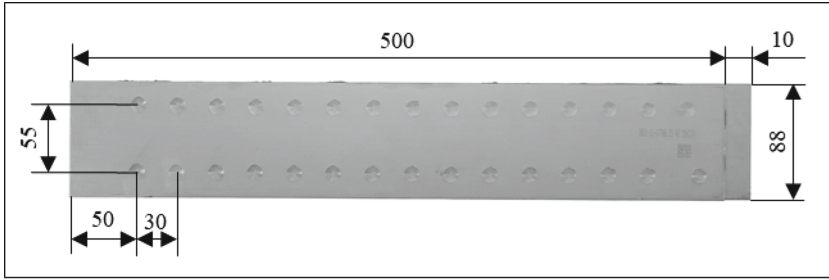
Fig. 2. Sample geometry and joining point position (all dimensions in mm)

4.3 Meta Data and Design of Experiments

The welding program involves a fixed pre-current of 12 kA for 400 ms, followed by an upslope to the main current (70 ms) and then the main current (90 ms). The main current is variable and defined by the design of experiments (DOE). So is the joining force. The Remaining Electrode Thickness (RET) is also listed as a setting parameter. It describes the maximum distance between the front surface and the conical cavity on the rear side of the electrode (see Fig. 3). This reduction represents an electrode that is already about 50% worn. The electrodes used conform to the ISO 5821 - A0 – 20 – 22 - 50 standard [14]. However, the radius of all electrodes is increased to 100 mm by means of a tip dresser. This ensures a larger contact area on the samples and therefore a more uniform current flow in the process.

After the samples have been joined, the adhesive is cured in an oven (30 min at 180 °C) and then they undergo a roll-off test in accordance with DVS 2916-1 [15].
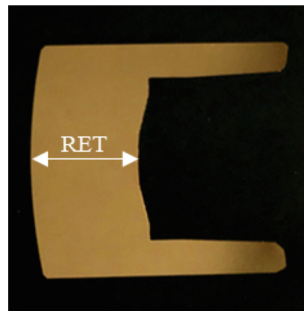
Fig. 3. Remaining Electrode Thickness (RET) of an electrode (here seen in a cut)

Each MTC consists of 5280 points. The DOE was structured in a randomized block design. Since every possible combination of setting parameters occurs exactly once in the experimental plan, this is also referred to as a completely randomized block design [16]. The necessity for this resulted from the large number of points to be joined and the system-specific peculiarities. Table 3 shows how the individual MTCs were subdivided.

The columns of the table represent the setting parameters to be investigated and the rows represent the values of these parameters.

Table 3. DOE of these investigations

Block 1 MTC (5280 spots)	Block 2 RET (2640 spots)	Block 3 F (660 spots)	Randomised I (30 spots)
Top sheet & Bottom sheet	9 mm	5 kN	26 – 47 kA
		6 kN	26 – 47 kA
		7 kN	26 – 47 kA
		8 kN	26 – 47 kA
	7 mm	5 kN	26 – 47 kA
		6 kN	26 – 47 kA
		7 kN	26 – 47 kA
		8 kN	26 – 47 kA

5   Target Data and Feature Engineering

5.1   Target Data

The target data of this investigation is the average nugget diameter of the individual joining points. This is calculated analogously to DVS data sheet 2916-1 as follows [15]:

$$d_a = \frac{d_1 + d_2}{2} \tag{1}$$

d_1: Maximum nugget diameter
d_2: Minimum nugget diameter

5.2   Feature Engineering

Furthermore, the time-dependent parameters are recorded and evaluated for all joining points. The sampling rate is 1 kHz. Figure 4 shows an example of the recorded time series of a weld.

These time series are divided into sections and within each section defined features are calculated and written out (feature engineered). These features are e.g. the high and low points, the mean value and the standard deviation or the slope of the regression degrees within the sections. Some of these features are also marked in the Fig. 4.
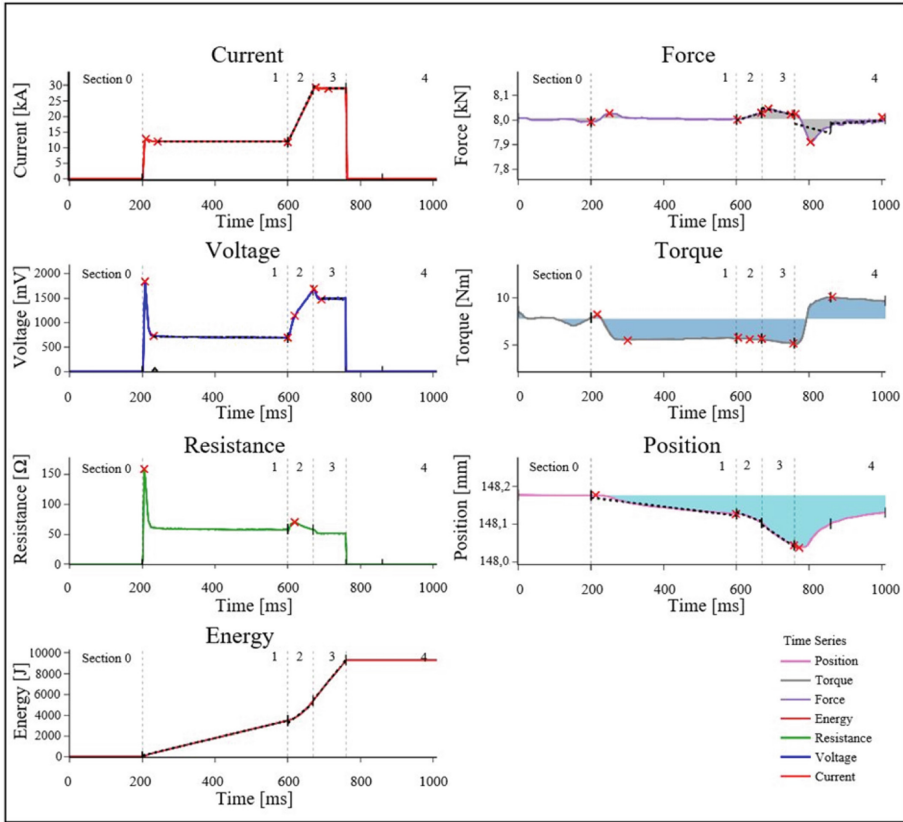
Fig. 4. Time series of a joining point (MTC: 2x 3.0 mm AL6-HDI; Current: 36 kA; Force: 7 kN)

However, only force related features are relevant to this study. For example the parameters p and σ of the force signal of Sect. 4. p is the momentum [Ns] and σ the standard deviation [N] of the force signal during the forge time (Sect. 4). σ is calculated as follows:

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(f_k - \overline{X})} \tag{2}$$

n: Number of points measured in Sect. 4
k: Summation index
$\overline{X}$: Mean value of all values in Sect. 4
f_k: Force value

The momentum p (more precisely the momentum change Δp) is calculated this way:

$$\Delta p = \int_{t_1}^{t_2} F(t)dt \tag{3}$$

Δp: Momentum (momentum change)

F(t): Force progression in Sect. 4

t_1: Start time Sect. 4

t_1: End time Sect. 4

These features are merged with the metadata of a joining point to produce a data set for a welding point. The data sets are analyzed in the further process of the investigation.

6   Results

6.1   Machine Learning Approaches

As reported in part 5.2, a large number of features is generated from each spot weld. In a first step, all these features were used train (60% of the data) and test (20% of the data) different regression models. This is followed by a 10-fold cross-validation for each model using the remaining 20% of the data set. Table 4 shows the R^2 values for each model. This is based on the validation data. It is easy to see that Gradient Boosting and Random Forrest have much better coefficients of determination than the other models. Therefore, only these models can be considered for further investigations.

Table 4. Coefficients of determination for different Machine Learning approaches

	Gradient Boosting	K-Nearest Neighbors (K = 5)	Lasso Regression ($\alpha = 0{,}1$)	Support Vector Regression ($\varepsilon = 1{,}0$; C = 1,0)	Random Forrest	Ridge Regression ($\alpha = 1{,}0$)
R^2	0,84	0,69	0,73	0,74	0,86	0,80

6.2   Permutation Feature Importance

After tuning the hyperparameters, the Gradient Boosting method was found to be the best performing of the models investigated here. Thus, the investigation regarding the relevant features is carried out on this model. To evaluate which features are relevant and which are not, Permutation Feature Importance (PFI) is applied to the results of the Gradient Boosting model. The result of this investigation (boxplots) can be seen in Fig. 5. It should be noted that the focus was deliberately placed only on features that are related to the force signal. It is shown that two features have a relatively large influence on the nugget diameter: The parameters p (feature name: "force_area_section_1_4_neg") and σ (feature name: "force_standard_metrics_section_3_4_4") of the force signal of Sect. 4. The influence of these parameters will be further investigated below.
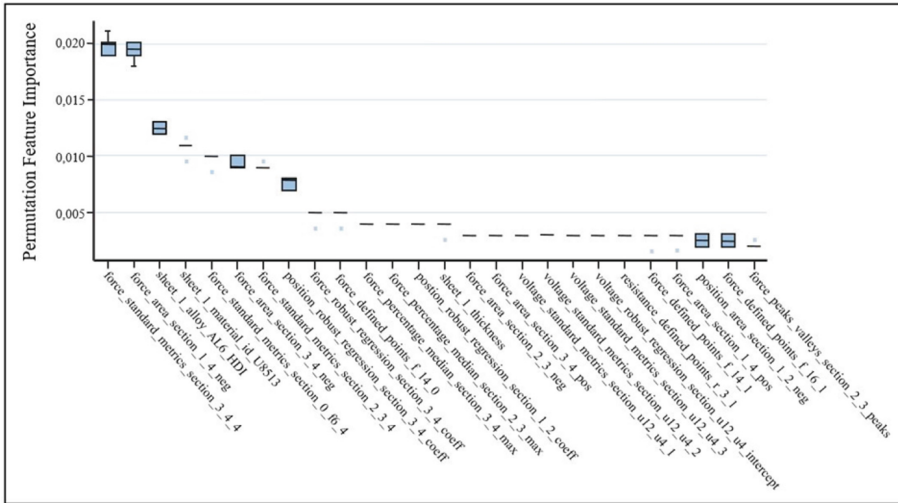
Fig. 5. Top 25 features (all sorts of signals)

6.3 Visualization of Datasets

In order to better assess the influences of the two features p and σ, they are visualized. These features are shown in a 3D representation in relation to the nugget diameter (see Fig. 6). It can be seen that the two features do not have a linear effect on the nugget diameter but run in the shape of a horn. This suggests a quadratic relationship. Furthermore, it can be seen that the MTC, which consists of 1.5 m thick plates, expands further in all spatial directions and also has many more joining points with a nugget diameter of 0 mm than the comparative experiment with 3.0 mm thick plates. The reason for this is the lower stiffness of these samples. This allows a larger deflection of the samples during welding and therefore also larger values for p and σ.

However, it also remains to be noted that no points larger than 11 mm can be obtained with either MTC. It is also noteworthy that when using plates with 1.5 mm thickness, the value of the standard deviation seems to have much less influence on the nugget diameter than with 3.0 mm thick plates.

6.4 Second Degree Regression

Figure 6 shows that the momentum has a much greater influence on the nugget diameter than the standard deviation. Therefore, only the momentum will be considered in the following investigations.

In the second step of this investigation, it should be clarified to what extent the momentum during the forge time is suitable as a predictor in order to be able to make a prediction regarding the nugget diameter to be expected. The method of analysis chosen here is regression. As already mentioned, the influence of p on the nugget diameter is not linear. For this reason, the analysis is done using the second degree regression.
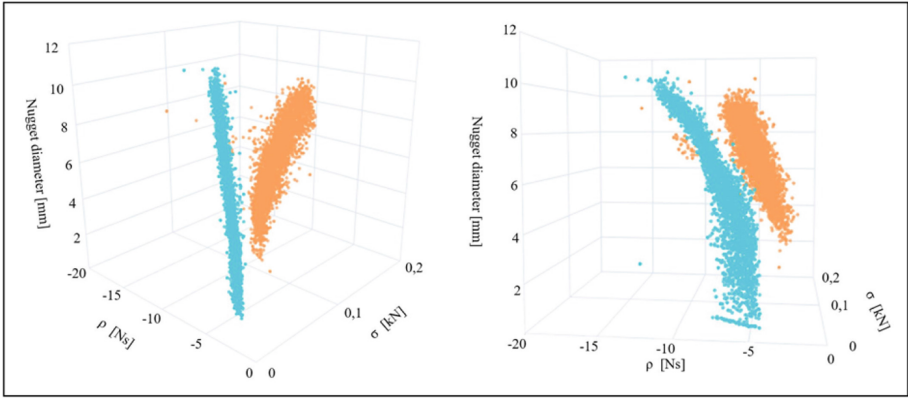
Fig. 6. Graphical visualization of the impulse p and the standard deviation σ over the nugget diameter (blue: samples with 1.5 mm thickness orange: samples with 3.0 mm thickness)
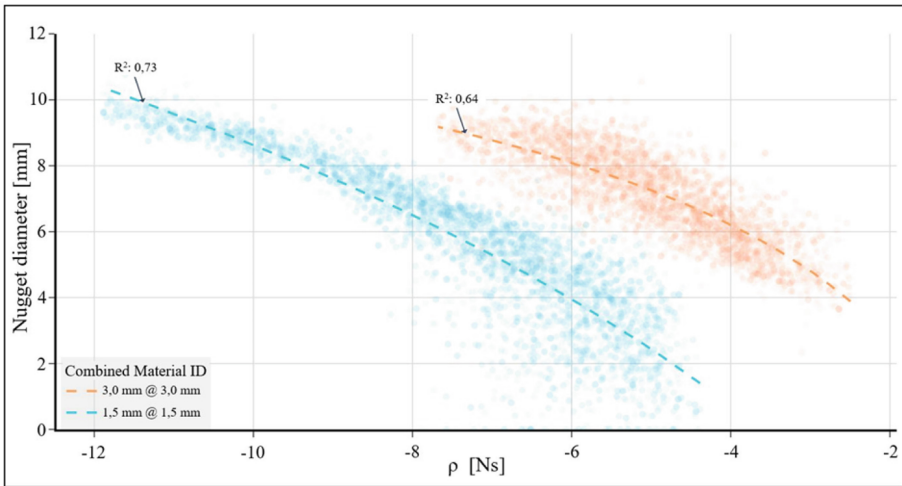
Fig. 7. Regression degrees and R^2 value across all main current values and MTCs

As can be seen in Fig. 7, R^2 values of 0.64 (3.0 mm sheet thickness) and 0.73 (1.5 mm sheet thickness) are obtained.

This result is surprising in that only a single feature forms the basis for this regression. It also shows that the larger range of p values has a positive effect on the R^2 value when using the 1.5 mm thick plates. In addition, it is believed that the lower strength of these specimens has a positive effect on the expression of the momentum, which in turn results in a larger R^2 value across the experimental space, making a better prediction possible.

7 Conclusion and Outlook

In the course of this investigation, it can be shown that the change in the force signal during the forge time in spot weld bonding correlates with the nugget diameter. It is shown that, depending on the sheet thickness used, the standard deviation of the force signal during the forge time can have more or less influence on the nugget diameter. Furthermore, with the help of feature engineering, the momentum during the forge time can be determined and it is also suitable up to a certain degree as a predictor for the nugget diameter. In the context of this investigation R^2 values of 0.64–0.73 can be proven. In order to increase this value, the experimental data obtained here can be used and further features can be applied to the regression analysis. However, this may also lead to overfitting.

Alternatively, the data set presented here can also be examined with the help of other methods, various statistical approaches or additional Machine Learning models (e.g. ANN etc.).

References

1. Kirchheim, A., Schaffner, G., Staub, R., Jeck, N.: Elektrodenkraft als wichtige Prozessgröße beim Widerstandspunktschweißen. Z. Schweißen und Schneiden **53**, 635–637 (2001)
2. Singh, S.: Beitrag zur Verbesserung und Sicherung des Tragverhal-tens von Widerstandspunktschweißverbindungen an Aluminium- und Stahlwerkstoffen durch technologische Maßnahmen und durch Ent-wicklung einer Regeleinrichtung. Dissertation, RWTH Aachen (1977)
3. Ruge, J.: Einfluss der Stromform auf das Anlegierungsverhalten verschiedener Aluminium-Elektrodenwerkstoffkombinationen beim Widerstandspunktschweißen. Report AiF-Project Nr. 5233 (1983)
4. Developments towards high-volume resistance spot welding of aluminium automotive sheet component (2008)
5. Schmal, C., Meschut, G.: Refill friction stir spot and resistance spot welding of aluminium joints with large total sheet thicknesses (III-1965-19). Weld. World **64**(9), 1471–1480 (2020). https://doi.org/10.1007/s40194-020-00922-2
6. Al Quran, F.M.F., Matarneh, M.I., Belik, A.G.: Influence of main characteristic features of spot welding on welded connection/joint strength. J. Eng. Phys. Thermophys. **87**(2), 394–398 (2014). https://doi.org/10.1007/s10891-014-1024-2
7. Kim, T.W., Park, Y.W.: Parameter optimization using a regression model and fitness function in laser welding of aluminum alloys for car bodies. Int. J. Precis. Eng. Manuf. **12**(2), 313–320 (2011). https://doi.org/10.1007/s12541-011-0041-8
8. Yang, H.-L., Cai, Y., Bao, Y.-F., Zhou, Y.: Analysis and application of partial least square regression in arc welding process. J. Cent. South Univ. Technol. **12**(4), 453–458 (2005). https://doi.org/10.1007/s11771-005-0181-z
9. Yu, J.: Quality estimation of resistance spot weld based on logistic regression analysis of welding power signal. Int. J. Precis. Eng. Manuf. **16**(13), 2655–2663 (2015). https://doi.org/10.1007/s12541-015-0340-6
10. Zhou, B., Pychynski, T., Reischl, M., Kharlamov, E., Mikut, R.: Machine learning with domain knowledge for predictive quality monitoring in resistance spot welding. J. Intell. Manuf. **33**(4), 1139–1163 (2022). https://doi.org/10.1007/s10845-021-01892-y
11. Friedrich, H.E. (Hrsg.): Leichtbau in der Fahrzeugtechnik. Stuttgart (2017)

12. VDA Empfehlung 239-200: Flacherzeugnisse aus Aluminium, Verband der Automobilindustrie e.V. (VDA), Berlin (2017)
13. DIN EN 515: Aluminium and aluminium alloys - Wrought products - Temper designations EN 515, DIN Deutsches Institut für Normung e. V. Berlin (2017)
14. DIN EN ISO 5821: Resistance welding - Spot welding electrode caps, DIN Deutsches Institut für Normung e. V. Berlin (2010)
15. DVS 2916-1: Testing of resistance welded joints - Destructive testing, quasi static, Deutscher Verband für Schweissen und verwandte Verfahren e. V. Düsseldorf (2014)
16. Dodge, Y.: The Concise Encyclopedia of Statistics. Springer, New York (2008). https://doi.org/10.1007/978-0-387-32833-1

Towards an Automated System for Robot Assembly Cell Layout Optimization

Joshua Beck$^{(\boxtimes)}$ (iD)

Fraunhofer Institute of Manufacturing Engineering and Automation (IPA), Nobelstr. 12,
70569 Stuttgart, Germany
joshua.beck@ipa.fraunhofer.de

Abstract. The manufacturing industry is exposed to various challenges: Increasing competition from companies abroad, the growing unpredictability and volatility of markets, and rising customer demands forcing manufacturers to be highly flexible and offer low-cost products. Using simulation tools, productions systems can be planned and analyzed before the deployment in the field, reducing the commissioning of named systems and avoiding cost-intensive errors. However, manual generation and validation of simulation models is a tedious task and demands expert knowledge of the planners about the simulations tools used. Automated generation of simulation models promises to reduce further commissioning time, making simulation tools more cost-effective. This paper compares existing approaches for 3D simulation generation and/or validation and further presents a concept for automated generation of a 3D model of a robot cell, creating a simulation and validating the processes taking place with minimum user intersection to accelerate robot cell commissioning. This is done by combining layout information with assembly related information extracted directly from 3D product model.

Keywords: 3D Simulation · Robot Assembly Cell · Layout Optimization

1 Introduction

On the way to a fluid and dynamic production, conventional planning methods must be overhauled. The planning and validation of production cells must be realized in the shortest possible time to react to extreme events or unexpected changes in production. Today, robotic assembly cells are often still planned as 2D layouts or inflexible CAD models. A robot assembly cell's layout can be tested and optimized before it is built using simulation tools. Validation and optimization is faster than in real-time, parallelizable and does not need to be physically tended for an environment to be reset. [1] However, simulative validation of assembly cells still has flaws. The creation and analysis of a simulation model involve many manual, time-consuming activities that require expert knowledge of the simulation software from the users, which is contrary to the stated goal of simulation studies. The aim should therefore be to automate the modelling and the analysis of robot assembly cells in the simulation by directly obtaining existing

information, e.g. from exploiting Enterprise Resource Planning (ERP) Systems or from the product model itself. Furthermore, to use the optimized layout and robot programs generated in the simulation environment for a faster ramp-up time of a robot cell.

This papers goal is to evaluate existing (semi-)automated approaches for 3D simulation generation, validation and layout optimization of robot cells. It further presents a concept for an automated 3D layout optimization for robot assembly cells. The rest of the paper is structured as follows: The next chapter gives an overview of criteria to analyze the state of the art. Section 3 reviews the state of the art for automated simulation generation and layout optimization. The reviewed approaches are discussed in Sect. 4. In Sect. 5, the concept of the planned approach is presented in general together with assumptions made. Finally, Sect. 6 concludes the paper and states future prospects.

2 Aspects of Automated Layout Optimization

Four individual problems must be overcome during simulation studies after the goal has been specified: Data collection, model implementation, validity check, experiments and analysis. [2] To optimize a layout for a robotic cell automatically through 3D simulation studies, firstly, necessary information must be gathered. Once this information has been structured, it can be used to create a model in a simulation environment. Afterwards, processes can be modelled using assembly-relevant information and assigned to the given robot. Subsequently, the processes and robot motions are validated for the given layout. Measured by defined key performance indicators (KPI's), both the processes and the arrangement of the components can be analyzed and optimized.

2.1 Definition of Criteria for 3D Simulation Generation and Optimization

To evaluate existing approaches and to identify gaps in research, a list of criteria has been defined. They are separated in four categories, each one for one of the declared problems: Data preparation, model generation, simulation and optimization. Furthermore, two more questions are defined: Is the focus on robot assembly cells, other kind of robot cells or models, e.g. assembly lines, or does the work has another focus than robot assembly cells? The other question is whether the simulation can be used for robot offline programming so that the results can be used beyond the simulation other than the layout.

For data preparation, the source of data and whether it is manually edited or prepared is of interest. In addition, the approach is verified to be able to deal with general conditions and constraints.

With respect to model creation, existing approaches are evaluated as to whether they actually create a 3D model of a system automatically based on the input data or whether the model is assumed or if a different type of model is applied. Additionally, the number of robots that can be considered in a cell or in a layout is examined.

After building the model, the simulation generation is assessed. Special attention is paid to an actual simulation of the assembly processes. They are differentiated, whether this is done in detail, only the movements of the robot are considered, the movements are calculated but not simulated, or if the simulation is outside the scope of interest. For the optimization of the system, the algorithm used is interesting and the focus of the

optimization. The state of the art is distinguished whether only the position of the robot is optimized, the positioning of components around the robot, whether the entire system, if only specific kind of components are considered or no optimization takes place at all. These criteria are listed in Table 1.

Table 1. Criteria for evaluation of existing approaches

Criterium	Option 1	Option 2	Option 3	Option 4
General				
Focus robot assembly cells	No ○	Other type of robot cell ◑	Yes ●	
Robot program export	None ○		Yes ●	
Data Collection				
Data preparation	Manual ○	Manually edited ◐	Automatic / included ●	
Brownfield and greenfield	Only one ○		Yes ●	
Model generation				
3D-Model	No, other model type, prerequisite ○		Yes ●	
Amount of robots	None ○	1 robot ◑	>1 robot ●	
Simulation				
Assembly task simulation	No ○	Calculated, simplified ◑	Extensive, collision analysis ●	
Additional path planner	None ○	Simplified ◑	Yes ●	
Layout Optimization				
Applied method				
Considered components	No optimization, other type ○	robots/one type of components ◕	With Exceptions ◑	Whole layout ●

3 Approaches for 3D Modelling and Optimization

The idea of automated modelling and simulation of operations and processes is not new and has its origins in the 1990s. [3] As one of the first approaches, Luedth presented a method in 1992 to place components of a robot cell around a defined robot base position. The space was divided into discrete voxels, and the voxels were assigned to the component, which are located within the component. Subsequently, objects are placed around a robot's base such that they only occupy free voxels to ensure that the arrangement of components remains collision-free. Shortest robot trajectories were considered as optimization criterion. [4] Similar approaches were published by Tubaileh [5] and Filipovic [6], where components are placed around a central robot base. They used simplified geometries in 2D space. In both cases minimal joint motions of the robot are targeted. Simulation has not been applied. In general, discretized workspaces (WS) are used due to their ease of use and faster computation time. This simplification was also applied in Hwang et al. [7], whose focus was to find an optimized robot base position for existing working poses, which they could use for spot welding on car bodies. The 3D model however, must be given beforehand.

A similar goal had Spensieri [8] using brute force technique to find a valid robot position and then solving a classical travelling salesman problem to find an optimized sequence of welding spots using IPS [9] as simulation tool. The 3D model must be given beforehand as well. Similar approaches to place robots come from the field of mobile robot applications. Vahrenkamp for example published an article about optimal mobile robot placement for task execution in given environments. [10] A reachability map was generated for the specific robot to evaluate different base positions for predefined task executions.

Focusing on robot assembly cells in his dissertation, Rossgoederer [11] took a hierarchical approach for the cell layout optimization. In contrast to previous publicized works, not only one type of component, such as a robot or peripheral devices around the manipulator, were considered. Also, the optimized placement of a desk and then the placement of work piece carriers on top of it was possible using a hierarchical genetic algorithm. The work aimed to optimize the layout of hybrid assembly stations. For this purpose, starting from a Bill of Materials (BOM) and a work description, a layout structure was first developed in which the hierarchy levels were defined. The CAD models of the components had been manually enriched with additional information, for example, defining the robot's footprint and baseplate. From this additional information, the restrictions for the optimization are determined. All components were first loaded into the environment and positioned collision-free for the model. The initial arrangement was manually created. The robot processes were designed using defined program blocks for gripping, transporting and joining. In case of collisions during movement, a sliding algorithm was implemented to generate collision-free trajectories. The constraints chosen were the reachability of the target positions for the robot. Free degrees of freedom not fixed by the CAD features were used to discretely position fixtures on a table. A virtual reality system was developed for the manual assembly stations since these could not be processed automatically.

Lietaert et al. [12] developed a system to optimize collaborative production cells with an extensive focus on ergonomics for human workers in the cell. As output a feasible work cell layout is given with its dedicated ergonomic score. As an input the system needs geometric information of all components and the task allocation. For optimization a non linear programming solver was implemented. Another work was published by Sharma et al. [13] They use point clouds generated from CAD models of the components to optimize a robot cells layout. Evaluating joint angles, simulated annealing was chosen as optimization algorithm.

Leiber et al. [14] presented an approach for semi-automated layout planning for assembly lines. Resources matching the assembly processes are selected from a database based on user input and arranged around stations around a conveyor belt using collision and reachability analysis. An evolutionary algorithm is used for optimization and, as a fitness function, the position quality with respect to the product as well as the assembly cost of the resources. The layouts are analyzed in a rigid body simulation. However, many input information are needed from the operator, such as the type of processes, the sequence, the region of interest at which the operations should be performed.

Bachmann et al. [15] published an robot cell optimizer for an application with two robots using a genetic algorithm. As input, CAD-data as well as the task sequence

and the location of the task on the associated part. For each component, restrictions are defined for their placement. In summary, these restrictions span the solution space. Although using 3D models, the solution space is reduced to a 2D problem with fixed robot base positions generating solutions for the placement of workpieces on a table. Simulation is used only for feasible, optimized layouts. For evaluation, path length as well as manipulability are considered.

Seeber et al. [16] published another work for assembly line layout generation using physical modelling of relationships between the components. Generated information is defined in AML [17]. Among them are 2D coordinates, the structure and hierarchy of individual assembly stations, their dependencies, peripheral interference contours, and process flows.

Michniewicz [18] has developed a complete system to develop automated work planning based on simulation models for assembly stations. Information is extracted directly from the product's CAD data as a data basis. In a first step, the algorithm evaluates the individual stations' feasibility to perform given tasks and assigns them accordingly. Then, first local simulations are analyzed to validate sub-process specific capabilities such as the correct gripping of components with a gripper, followed by a global simulation for reachability analysis. Finally, the system is analyzed economically through cycle time analysis and machine hourly rate calculation. The layout of the cells are untouched, processes are only assigned to given stations. Therefore, 3D models are not adapted, but only used.

Lämmle et al. proved in [19, 20] that AML can be used to build and simulate production cells and to describe the dependencies between individual resources and processes using a product-process-resource model (PPR) in VisualComponents [21]. In this context, program modules for handling and gripping objects as well as a placeholder process for screwing applications were developed. However, the model was neither analyzed nor optimized. The approach was limited to component models already existing in the software internal library.

Preparing or creating necessary input data is essential. User enters often information manually. To look at the problem of data acquisition, Weigert [22] developed an automated exchange system as middleware and link between existing software tools. The basic idea was to create a uniform data format based on AML, which can be used utilizing parsers and importers in different software tools, for example, to develop 3D simulations or discrete models. For the examples used, the origins of the information were not mentioned.

4 State of the Art Discussion

To each of the named approaches the criteria defined in Sect. 2 were applied. The discussed approaches and their corresponding classification are shown in Table 2.

The approaches can be divided into four categories, as they are sorted in Table 2 and depicted in Fig. 1: The first ones are *task-centered* solutions, where robots are placed into a given environment to fulfill their tasks. Examples are placing robots on an assembly line or mobile robotics. The second group are *robot-centered*: Arranging components around fixed robot bases. The third category are approaches, where all components are considered for layout generation and optimization.

Table 2. Comparison of existing approach for robot assembly cell

Authors	General		Data Collection		Model Generation		Simulation Generation		Validation/Optimization	
	Focus robot assembly cell	Robot program export	Data preparation	Green and Brownfield	3D Model	Amount of robots	Assembly task simulation	Path planer	Applied method	Considered components
Leiber et al.	◐	○	●	●	●	●	●	●	GA	◐
Vahrenkamp et al.	○	●	○	●	○	◐	●	●	Inverse Kinematic	◐
Hwang et al.	○	●	○	○	○	◐	◐	◐	TSP	○
Spensieri et al.	○	●	○	○	○	◐	◐	◐	TSP	○
Bachmann et al.	●	○	○	●	●	●	○	○	GA	◐
Sharma et al.	●	○	○	●	○	◐	◐	○	SA	●
Tubaileh et al.	●	○	○	○	○	◐	◐	○	SQP	◐
Filipovic et al.	●	○	○	○	○	◐	◐	○	SQP	◐
Luedth	●	○	○	○	○	◐	◐	◐	Seq.	○
Rossgöderer	●	●	◐	●	●	◐	◐	◐	GA	●
Lietaert et al.	◐	○	○	○	○	●	○	○	Ergonomy score	●
Lämmle et al.	●	●	○	●	●	●	●	○	None	○
Seeber et al.	◐	○	○	●	○	●	○	○	Physical rlt. modelling	●
Michniewicz	◐	○	●	●	○	○	○	○	None	○
Weigert et al.	○	○	●	●	●	○	○	○	None	○

The last group has neither the focus on simulation generation nor optimization of layouts, however, they are examples for how data management and integration in such a system can be applied.

Comparing the goal with the reviewed approaches, none of them shows the specific focus or covers all questions stated in Sect. 2. Many papers address sub-aspects that need to be considered but represent isolated solutions to specific problems. Arrangement planning tools that only want to either place the robot or place other components around a robot are not sufficient on their own to be able to look at different types of cells. Especially the task-centered approaches, except Leiber, focus on the task sequence and are not suitable for assembly lines, but not for layout optimization of modular robot assembly cells. Also disadvantageous is that many approaches need manually added information or existing simulation models in general. An approach is needed that can place both handling devices and peripheral components while considering given constraints.

Most interesting are the approaches of Rossgöderer, Leiber, and Bachmann et al. Rossgoederer already focuses on robot assembly cells with an extension to hybrid cells. However, only one robot is considered in a cell. Drawbacks of the approach are the concentration of only one robot in a cell and long processing times. Additionally, a simple path planning algorithm is implemented. Especially when using process time as KPI, an extensive path planner is crucial to find a nearly optimal layout. Leiber can evaluate the position and simulation of multiple robots, however the focus is on assembly lines, not on single robot cells. Placement of separation systems for the robots are not considered. Furthermore, collision detection only works between moving and static objects, not between the robots. Bachmann reduced the problem into a 2D problem.
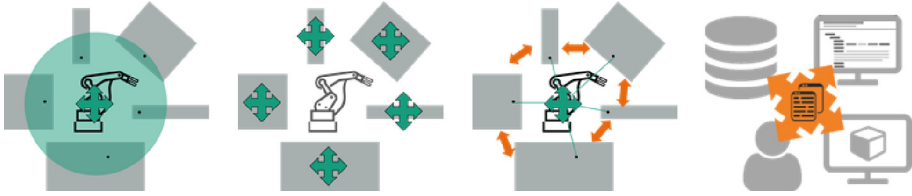
Fig. 1. Categories of discussed state of the art papers from left to right: task-centered, robot-centered, holistic, data management systems

This examination shows that none of the given approaches are sufficient for the simulation-based layout evaluation of robot assembly cells for multiple robots. In the following, a concept for the specified goal is proclaimed.

5   Concept Formulation

The overall goal of the planned system is to reduce manual effort for setting up an optimized robot assembly cell by using 3D process simulations and available information from layout as well as from product planning and design. The concept is depictured in Fig. 2. Robot cells with multiple manipulators shall be considered as only Bachmann et al. include this for robot cells. The necessary input for the system is to be created automatically with minimal manual input and is divided into layout information and assembly-related product information. The layout information given is assumed to be a rough layout for an assembly cell from an upstream 2D layout planning in AML as given in [16]. This planning shows where resources are located and which processes there are to perform. However, this information is insufficient to create the 3D model in detail. In particular, height information is not given in 2D. Rules must be defined to extrapolate 2D to 3D relationships. Furthermore, assembly-relevant details on process sequences and the region of interest for the process are not mapped. This information can be generated from the product model itself, for which preliminary work in the field of computer-aided assembly planning (CAAP) also already exists. [23, 24] These data serve as a starting position for the 3D layout planning. Visual Components is used assimulation software. Lämmle et al. [19, 20] have shown that this software is suitable for the automatic generation of cell layouts and the automated generation of simulations and the software was used in previous work before [25]. However, not only existing models from the software's catalogue will be used, but a hybrid approach is followed: Resource models from an external database shall be used models as well as predefined component models from the software's library, especially for kinematic chains such as industrial robots.

A set of rules must be defined for the placement of components. An example could be that the bounding box of a component must not overreach the bounding box of the parent component to ensure for example that a work piece carrier is placed correctly on a table. As robot cells are portrayed only, access points, might it be for AGV's or conveyors, are excluded or considered as fixed points in the constraints of the layout.

Program modules are to be defined and used for the automated generation of the processes, as already published in [25]. These program modules portray process skills
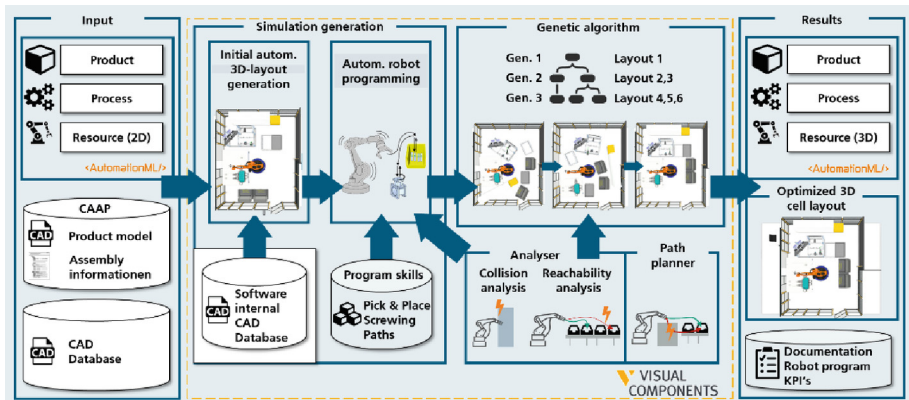
Fig. 2. Concept for an automated approach for robot cell layout optimization

for the handling devices to perform. In the software, similar to the approach of Spensieri et al. [8], internal functions shall be used to develop analysis tools to check the processes on the given layout automatically. Of special interest are tools to analyze reachability and collision clearance. For optimization, an evolutionary algorithm will be used in a hierarchical manner as in [11] to reduce the complexity of the optimization problem. Although the idea of physically modelling the dependencies between components is an attractive solution, the modelling in three-dimensional space is much more complex compared to the two-dimensional one. The layout and processes will be analyzed for the following KPI's: The cycle time is to be considered to examine the arrangement for its efficiency. It is not necessary to simplify and utilize the axis angles or path length of the robot as the motion simulation can be performed. As a result, an optimized robot assembly cell layout, program templates for the robot's execution and the KPI's as well as a documentation of the given layout Next to it, a minimal footprint is desired. Research question to be answered in this work are:

1. How can 2D layout data be extrapolated to 3D in a meaningful manner?
2. Can automatically generated layouts be compared or are even better than manually generated ones?
3. How can an evolutionary algorithm be parametrized to create new layouts in a given amount of time, so a human operator can to other tasks in the mean time?

6   Conclusion and Next Steps

Simulation is a key tool to improve and accelerate layout optimization of robot cells. Different solutions and approaches already exist in the literature to automate layout optimization or to generate simulation models. However, there is no approach yet for robot assembly cells with multiple robots. A concept was proposed, which shows a solution to the given problem. In a first step, a tool must be developed to combine and merge the existing data from the AML of the layout planning with the information from the CAAP. Particularly critical here is the classification of the processes and their positions and, for the analysis, the division of the components into movable and static

objects and the definition of boundary conditions. Logically, a conveyor belt that delivers partially assembled semi-finished products to a station must not change position. In addition, the database must be built, which contains the CAD model of components and other information, such as the name. It would also be possible to add the capabilities of the models to enable a unified database for all tools from the toolchain. An Add In for Visual Components must be written afterwards to build the model. First program modules and collision analysis tools for the simulation analysis of the processes as well as the automatic modelling of the products to individually manageable objects, have already been developed by the author in [25] and can be extended. Last, the optimization of the layout and the robot trajectories must be considered and implemented.

References

1. Collins, J., Chand, S., Vanderkop, A., Howard, D.: A review of physics simulators for robotic applications. IEEE Access **9**, 51416–51431 (2021)
2. Lütjen, M.: Modellierungskonzept zur integrierten Planung und Simulation von Produktszenarien entwickelt am Beispiel der CFK-Serienfertigung, Dissertation, Bremen, Universität Bremen, Fachbereich Produktionstechnik (2014)
3. Kuhlenkoetter, B., Schyja, A., Hypki, A., Miegel, V.: Robot workcell simulation with automation ML support - an element of the CAx-tool chain in industrial automation. In: Robotics (ISR), 41st International Symposium on and 2010 6th German Conference on Robotics (ROBOTIK), Date, 7–9 June 2010, vol. 41, pp. 1076–1082 (2010)
4. Lueth, T.C.: Automated planning of robot workcell layouts. In: Proceedings/1992 IEEE International Conference on Robotics and Automation, May 12–14, 1992, Nice, France, pp. 1103–1108. IEEE Computer Society Press. Los Alamitos, Calif. (1992)
5. Tubaileh, A.S.: Layout of robot cells based on kinematic constraints. Int. J. Comput. Integr. Manuf. 1–13 (2014)
6. Filipović, M., Bogdan, S., Petrović, T.: Robot workstation failure recovery based on a layout optimization. Trans. Inst. Meas. Control. **40**(2), 375–388 (2018)
7. Hwang, Y., Yun Kyu, Y.: Performance based architectural design optimization: automated 3D Space Layout using Simulated Annealing. In: ASHRAE, pp. 292–299 (2014)
8. Spensieri, D., Carlson, J.S., Bohlin, R., Kressin, J., Shi, J.: Optimal robot placement for tasks execution. Procedia CIRP **44**, 395–400 (2016)
9. Industrial Path Solutions - IPS. Homepage. https://industrialpathsolutions.com/. Accessed 05 Jun 2022
10. Vahrenkamp, N., Asfour, T., Dillmann, R.: Robot placement based on reachability inversion. In: 2013 IEEE International Conference on Robotics and Automation (ICRA 2013): Karlsruhe, Germany, 6–10 May 2013, pp. 1970–1975. IEEE. Piscataway, NJ (2013)
11. Rossgoderer, U.: System zur effizienten Layout- und Prozessplanung von hybriden Montageanlagen, Dissertation, München, Technische Universität München, Lehrstuhl für Montagesystemtechnik und Betriebswissenschaften (2002)
12. Lietaert, P., Billen, N., Burggraeve, S.: Model-based Multi-Attribute Collaborative Production Cell Layout Optimization. In: 2019 20th International Conference on Research and Education in Mechatronics: 23–24 May 2019, Wels, Austria, pp. 1–7. Institute of Electrical and Electronics Engineers. Piscataway, New Jersey (2019)
13. Sharma, A., Jha, A.K., Halder, A.: Layout optimization of a robotic cell for foundry application by CAD based point cloud modeling – a case study. Ind. Robot. **44**(6), 788–797 (2017)
14. Leiber, D., Hammerstingl, V., Weiß, F., Reinhart, G.: Automated design of multi-station assembly lines. Procedia CIRP **79**, 137–142 (2019)

15. Bachmann, T., Nottensteiner, K., Roa, M.A.: Automated planning of Workcell layouts considering task sequences. In: 2021 IEEE International Conference on Robotics and Automation: May 30-June 5, 2021, Xi'an, China, pp. 12662–12668. IEEE. Piscataway, NJ (2021)
16. Seeber, C., Albus, M., Fechter, M., Neb, A., Yoshida, S.: Automated 2D layout design of assembly line workstations through physical principles. Procedia CIRP **104**, 1197–1202 (2021)
17. Drath, R., Luder, A., Peschke, J., Hundt, L.: AutomationML - the glue for seamless automation engineering. In: IEEE International Conference on Emerging Technologies and Factory Automation, 2008: ETFA 2008; 15–18 Sept., 2008, Hamburg, Germany, pp. 616–623. IEEE. Piscataway, NJ (2008)
18. Michniewicz, J.: Automatische simulationsgestützte Arbeitsplanung in der Montage, Dissertation, München, Technische Universität München, Lehrstuhl für Betriebswissenschaften und Montagetechnik (2018)
19. Laemmle, A., Gust, S.: Automatic layout generation of robotic production cells in a 3D manufacturing simulation environment. Procedia CIRP **84**, 316–321 (2019)
20. Lämmle, A., Seeber, C., Kogan, E.: Automatic simulation model implementation of robotic production cells in a 3D manufacturing simulation environment. Procedia CIRP **91**, 336–341 (2020)
21. Visual Components 4.4. Homepage. https://www.visualcomponents.com/. Accessed 05 Jun 2022
22. Weigert, D., Aurich, P., Reggelin, T.: Durchgehende Modellerstellung zwischen Simulation-, Visualisierungs- und Konstruktionswerkzeugen für die gesamtheitliche Planung von Produktions- und Intralogistiksystemen. Simulation in Produktion und Logistik (2017)
23. Neb, A., Brandt, D., Rauhöft, G., Awad, R., Scholz, J., Bauernhansl, T.: A novel approach to generate augmented reality assembly assistance automatically from CAD models. Procedia CIRP **104**, 68–73 (2021)
24. Neb, A., Göke, J.: Generation of assembly restrictions and evaluation of assembly criteria from 3D assembly models by collision analysis. Procedia CIRP **97**, 33–38 (2021)
25. Beck, J., Neb, A., Barbu, K.: Towards a CAD-based automated robot offline-programming approach for disassembly. Procedia CIRP **104**, 1280–1285 (2021)

Adaptive Intralogistics with Low-Cost AGVs for a Modular Production System

Javier Stillig[1](✉), Carolin Brenner[2], and André Colomb[2]

[1] Institute of Electrical Energy Conversion, Stuttgart, Germany
`javier.stillig@iew.uni-stuttgart.de`
[2] Institute of Mechanical Handling and Logistics, Stuttgart, Germany
`http://www.ift.uni-stuttgart.de`
`https://www.iew.uni-stuttgart.de`

Abstract. Due to ever faster changing market requirements, industrial production equipment needs to become much more flexible. For this reason, the Institute of Mechanical Handling and Logistics and the Institute of Electrical Energy Conversion are developing versatile, automated, and easily adaptable solutions to increase the flexibility of future intralogistics systems. As part of the ANTS 4.0 research project, a modular low-cost automated guided vehicle has been created, which breaks down the flow of goods into its smallest units: A small load carrier. The vehicle is prepared to be charged inductively and guided by color coded LED strips inside the floor, controlled from a superordinated artificial intelligence algorithm. In case of finding an obstacle by the object detection integrated in the floor, the route is recalculated and adapted in real-time.

Keywords: Modular Production System · Intralogistics · AGV · Industry 4.0 · Intelligent Floor

1 Introduction

The economic success of a product largely determines the way it is manufactured. As market influences become increasingly volatile and unpredictable across industries [1–4], factories have to reconfigure their manufacturing even more to current market trends in the future. This form of reconfiguration is also referred to as the adaptability of a manufacturing process. A higher degree of adaptability is achieved, among other things, through mobile and interoperable production equipment [5–8], which is able to be set up at any location in the shop floor environment and interact with other production equipment, forming a production network. All machines and systems are movable and compatible with each other. The flow of goods from and to the lines is fully automated by Automated Guided Vehicles (AGV). A control system is superordinated to the production and transport equipment, which synchronizes all activities on the production and logistics areas based on the current production orders.

Current research and development activities are therefore focusing on decentralized managed AGVs that can react flexibly to new requirements [9]. Among the systems already in operation is the KIVA system used by the U.S. company

Staples, in which around 500 robotic units transport shelves from a warehouse to picking stations [10]. Further, logistics groups such as Amazon already use over 200,000 robots in their factories today [11].

Flexibility is achieved by dividing the transport volumes into ever smaller quantities as the number of transport units grows. This granularization simultaneously requires the development of new, cost-effective AGVs. *KATE*, a small-size AGV from the University of Stuttgart and the Götting company, the low-cost AGV *Locative* and the *Multishuttle* from Dematic GmbH which bases on the Fraunhofer IML's project, are just some of the examples that have resulted from this need. Autonomous vehicles that can be linked together to increase payloads have also been developed, for example, in the KARIS project of Karlsruher Institute of Technology [12]. In addition, several companies sell transport systems based on swarm intelligence, such as the companies Knapp with the *Open Shuttle* or *Agilox*. In ARENA2036, a first functional prototype, called *BoxAGV*, was developed in 2020, which can follow dynamically lighted LED tracks using a simple camera. Based on this functional model, the AGV SCOOTY presented in this paper followed, extended to decode meta-information from the LED patterns along the way.
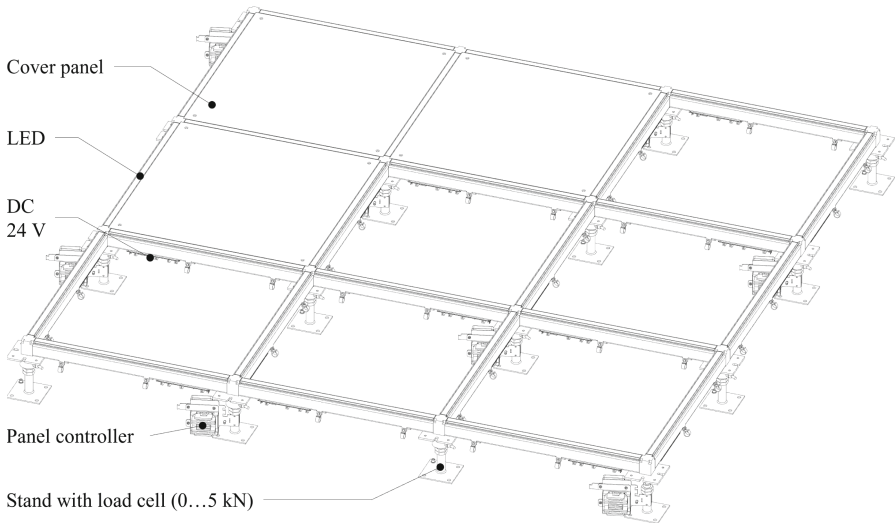
Fig. 1. Schematic of a 3 × 3 panel prototype *v0.8* of the Intelligent Floor (IF) with a height of 180 mm, which mainly consists of the components: stand with load cell, panel controller, crossbeams including 24 V distribution and LED system as well as basic floor panels. (Drawing: Bosch)

For the step-by-step realization of such a future production scenario, the Intelligent Floor (IF) contributes to the solution. The IF is a patented raised floor system [13] designed as a universal infrastructure platform for factory buildings. The open platform[1] supports the goals of the modular production and enables production equipment to be more mobile and compatible to each other.

The modular IF design, as shown in Fig. 1, consists of individual $0.36\,\mathrm{m^2}$ sized square panel elements, which can be equipped with different sensor and actuator functionalities chosen by the user. The floor can be adjusted in height via stands. The load capacity of the prototype *v0.8* is nominally $7\,\mathrm{kN}$ (breaking load $14\,\mathrm{kN}$) on a test area of $100 \times 100\,\mathrm{mm}$. It is therefore suited for light to medium-duty transport operations, like AGVs and light-weight fork lifters. Even in the basic version of the IF, sensor and actuator functionality is built into the substructure. For instance, the depicted load cell, integrated in the stand, can nominally measure loads up to a magnitude of $5\,\mathrm{kN}$. Furthermore, LED strips are mounted as visual actuators on the crossbars of the substructure, whose LEDs can be controlled individually and in any color.

2   Current Situation

According to the research study [14], the number of registered current vehicle models increased by 349% in Germany between 1990 and 2014. In addition, the number of possible equipment options has also increased along with the model variants, so that today almost every new car allows several million permutations of configuration options. This increase in product variance represents the first market development.
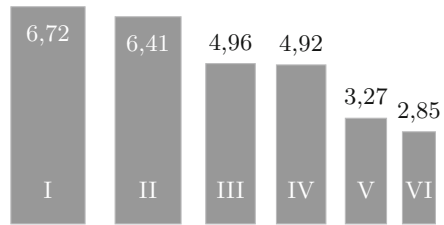
Fig. 2. Product life cycles of VW Golf generations I to VI based on million units sold in the period from 1974 to 2012.

The second market development is the shortening of the product life cycle according to Fig. 2. If the life cycle of a Volkswagen Golf I was nine years in 1983, it has been shortened to only four years in 2012 [15]. These two market developments present major challenges not only to vehicle manufacturers, but also to all their suppliers. Their products also have to be redesigned and produced at ever shorter intervals and in ever more diverse variations, and the corresponding production equipment has to be adapted to market developments. The driving force behind the increase in the number of variants and the shortening of product life cycles is the interaction between consumers and producers. The customer's

---

[1] The mechanical structure and parts of the operating software of the IF are freely available on request, so that everyone has the possibility to develop and sell own elements for the IF. Participation in the expansion of the IF ecosystem is explicitly requested and it is not limited to hardware and software products, but also includes services, such as assembly or planning services.

desire for choice and product individualization is contrasted by a large number of competing manufacturers who expect to gain sales advantages from shorter product cycles.

With increasing digitalization, consumers and producers are becoming better connected than before via the Internet. For the consumer, this means an increase in choice. For producers, the result is more global competition and thus increased pressure to develop their unique selling propositions. This pressure reinforces the market developments outlined above, so that a reduction in the number of variants and an extension of the life cycle time are not to be expected in the future. The change of a product often causes profound changes in the business processes of a company. This means that not only production processes are affected by the change, but also, for example, logistical and commercial processes. To maintain their competitiveness, manufacturing companies are therefore called upon to find ways of adapting quickly to volatile markets [16].

One approach to this adaptation is to break up inflexible logistics chains. These are found not only in the form of conveyor belt systems within production lines, but also, for example, in warehouses with dedicated storage locations. The ability to adapt is key to the modular production and an important research topic that scientists, especially the production-related institutes of the University of Stuttgart, have been working on for more than 25 years [17–22]. One of their central findings is that production equipment must become more modular, mobile, and interoperable in response to the rapidly changing demands of the global market.

3   Novel Platform for Modular Production

The IF offers various possibilities to implement the requirements of the modular production in the daily manufacturing practice. As shown in Fig. 3, the Customer and Innovation Center of Bosch Rexroth AG at the Ulm site, demonstrates the possibilities of the IF in automation on a small scale. For example, the LED-based indication system can be used to immediately adapt the space and route markings on the shop floor to changes in the production configuration. If an assembly station is moved to a new location within the production facility which is equipped with an

Fig. 3. Model factory of Bosch Rexroth AG in Ulm, Germany for the demonstration of products and solutions for the future modular production. (Photo: Bosch)

IF, the walkways, transport routes and logistics areas associated with the assembly station can be moved automatically as well. By signaling this information immediately visible on the floor, the operating staff is supported best.

The power supply to the assembly station is ensured by the wireless power transfer (WPT) devices currently capable of 240 W up to 3.7 kW installed in the floor. They enable the station to be commissioned installation-free. In addition, the stationary load cells, which are installed in the 620 mm grid below the floor, can be used for object detection. They confirm whether the weight of incoming goods corresponds to the stocks in the resource planning system, as described in [23]. Furthermore, they are used to trigger automatic incoming goods bookings or detect obstacles on walkways and transport routes. Due to the cyclic storage of measured values with an accuracy in the kilogram range, movement patterns on the floor can be detected and analyzed. The results of the analysis are used, among other things, to optimize production with respect to the walkways of workers or transport routes of AGVs.

The combination of LED display and detection system is particularly useful for the operation of AGVs. Thus, the hazardous area of an AGV can also be displayed dynamically, in the form of a "traveling" light frame around the AGV on the floor. This serves to reduce the number of unexpected and wasteful stops the AGV has to make, to avoid collisions with an unsuspecting person violating its safety zone. Also useful is the display of the AGV track to show humans the planned route in advance. In stationary robot applications, the IF, in combination with the optional safety panel, can indicate where dangerous movements occur. With this envisioned solution, an approaching worker would cause the movements to be slowed down or safely stopped in time[2]. Another significant use case of the IF for intralogistics lies in the flexible navigation of AGVs with a physical track, which is described from a more detailed technical and economical point of view in [23,24]. Section 5 shows in detail that the obvious contradiction between flexible navigation and track guided AGVs is in fact not one.

4   Scalable Low-Cost AGV Construction Kit

For the numerous transportation tasks in logistics, modern AGVs are evolving into autonomous mobile robots with a multitude of sensor equipment and intelligence to perceive and safely navigate their environment. On the other hand, this trend has a price in terms of AGV hardware cost and system complexity, where the IF allows exploring a different distribution of responsibilities and therefore a novel system architecture. For a granular and versatile material transport solution, the number of AGVs needs to scale cost-effectively, and the model variety should be kept small to create a greater redundancy of interchangeable transport vehicles. However, sometimes different AGVs are required depending on the purpose, especially the dimension and weight of the cargo (Fig. 4).

These design considerations led to the development of the Scooty concept as a scalable AGV construction kit. Toward a low-cost design, its functions are

---

[2] The current regulatory situation regarding safety requirements does not support such completely flexible applications. This topic needs to be addressed in future research, not limited to the presented approach.

Fig. 4. The AGV Scooty in operation on the IF located at the ARENA2036 in Stuttgart, Germany. (Photo: University of Stuttgart/Uli Regenscheit)

reduced to the required minimum, shifting as much as possible into the Intelligent Floor, such that the two can be seen as a collaborating overall system. While the IF includes many features concerning *sensing*, *signaling* and *power distribution* tasks, generating *motion* is not part of its purpose. Thus the AGV design focuses on this aspect, while taking advantage of the IF's functions in a synergetic manner instead of replicating them. As for the interchangeability of AGVs, the concept can be adapted to many different vehicle shapes because of its modular design, and is also based on a universal control architecture which allows common treatment for virtually any type of vehicle chassis. That accommodates the needed AGV variety even from third-party suppliers while keeping a consistent interface to control them.

In its simplest form, Scooty constitutes a chassis matching the 600×400 mm base size of small load carriers, four driven steered wheels and a small, easily replaceable battery. The control electronics are kept simple, avoiding complex sensor arrangements. Within this size constraint, most existing AGVs use a very simple drive concept with only two driven wheels, or a design utilizing Omni-Wheels (e.g. Mecanum). Regarding the control system of AGVs, these different chassis configurations are an important distinction, resulting in different Degrees of Freedom (DoF). AGVs typically travel with the front centered on some kind of (possibly virtual) track. Those with 2 DoF are restricted regarding their orientation, always pointing "forward" along the path. Take for example the well-known single-track kinematics model. Many vehicles can maneuver like a car or bicycle: controlled by one steering angle and one speed of the driven wheel, resulting in two control variables. The steering angle represents configuration, affecting the direction of the velocity vector (not the pose of the vehicle,

non-holonomic constraint). The second control variable scales the vector and thus the speed of the motion [25].

Omnidirectional (3 DoF) vehicles provide more flexibility in moving through production and logistics layouts, even between the IF's fixed grid of tracks. They are capable of reaching any destination with arbitrary orientation, and can e.g. switch to another parallel lane without changing their heading. A prominent example developed at the IFT is the Independent Fork System [26], designed to transport a pallet in any possible direction. Such omnidirectional motion enables, among other things, smaller radii, which can optimize space usage in production. The maneuvering time of 3 DoF vehicles is also shorter, since they can travel in any desired direction without changing their orientation beforehand. This enables quicker evasive maneuvers and faster cycle times. As the production layout becomes fully flexible, stock and storage areas need to be redefined and repositioned as well, leading to ad-hoc block storage areas as an appropriate solution. These can be operated most space-efficiently when the individual carriers (still residing on the AGV) can move in any direction, allowing each one of them to reach the block storage boundary quickly.

Another reason to prefer omnidirectional AGVs lies in the nature of the transported goods/carriers. By definition, there is no predetermined "front"or "back" for a pallet or small load carrier. If at all, the orientation of the cargo defines some preferred direction how the carrier should be picked up or delivered. But the *vehicle's* orientation is a mere technical detail, irrelevant to the transportation task. Thus, a vehicle which can freely move in any direction serves the task in the most flexible manner. When the size of the transported goods increases, several AGVs may even collaborate to act like a single vehicle, as in the case of the Independent Fork System. But if any one (or worse, several) of them is restricted to 2 DoF, the motion constraints can conflict with each other, making the whole compound impossible to maneuver. Standardizing on omnidirectional vehicles from the start avoids this problem and allows maximum transport flexibility, even when combining them to collaborate.

The SCOOTY concept is such an omnidirectional AGV with 3 DoF, capable of reaching any destination regardless of its orientation. This is achieved by using four independent steered wheel modules, as sketched in Fig. 5. This results in more stable and smoother motion characteristics, as well as higher load-carrying capacity in relation to the same design size, than possible with Mecanum wheels[3]. The modules can even be arranged differently to construct a vehicle form factor with an arbitrary number and positioning of the wheels. SCOOTY's universal control structure facilitates this flexibility on the software level accordingly.

In the state of the art a model for every type of AGV needs to be set up and an individual controller is designed. During the development of the different types of AGVs at the IFT the vision of such a universal vehicle model emerged. As a basis, the concept for controlling a vehicle with an arbitrary number of
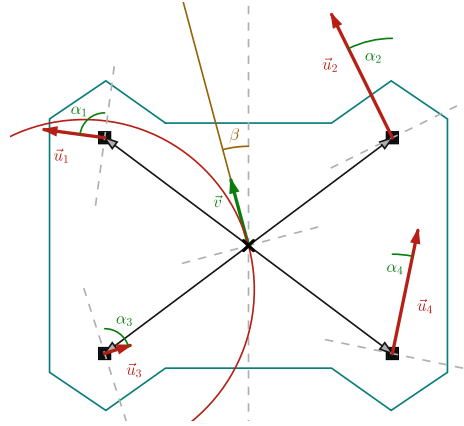
---

[3] Mecanum wheels have a discontinuous and small contact area toward the floor, causing vibrations and poorly defined intermediate states. Irregular friction characteristics lead to additional, hard to control disturbances.

wheels is introduced in [25]. This mathematical model describes the coordinated motion of a vehicle using three physically independent, dimensionless parameters: nominal velocity v_n, nominal curvature κ_n and slip angle β. With these Omni-Curve-Parameters (OCP), any vehicle with arbitrary number and geometrical arrangement of wheels can be controlled, respecting and utilizing its available degrees of freedom. Two parameters describe the motion mode and only influence the steering configuration of the vehicle wheels. The third parameter, the nominal velocity, changes its physical interpretation depending on the prevailing configuration.

With the aid of the mathematical model set up, these parameters can be used to provide each steered wheel drive with its target speed $\mathbf{u}_i$ and target steering angle α_i for any allowed motion. It assures that the pole rays of all wheels intersect at a common Instant Center of Rotation (ICR), as in Fig. 5, and avoids singularities in the calculated terms. The over-determined system of a vehicle with arbitrary steered wheels becomes controllable.

Fig. 5. Kinematics model of the four steered wheels used in SCOOTY. This configuration shows a nonzero slip angle β, so the velocity $\mathbf{v}$ diverts from the vehicle main axis, while following a curved path.

Furthermore, the choice of control variables scales naturally with any vehicle's degrees of freedom. From track-bound 1 DoF systems, where only the nominal velocity v_n is relevant, 2 DoF systems evolve gradually by adding the nominal curvature κ_n, so any given curvature-continuous path can be followed. This basis of two control variables is kept even for 3 DoF vehicles, because these types of motions are intuitively comprehended by humans in the vicinity or acting as operators. Accordingly, the third control variable β corresponds to the added ability of an omnidirectional vehicle to move in any direction relative to its main axis, as relevant from the perspective of the logistics context. A standardized, machine-usable description of different vehicle types' capabilities, expressed as limits in these OCP control variables, is presented in [27].

Besides the main function of actual transport motion, SCOOTY still requires components for a suitable energy supply. The developed hybrid approach utilizes the IF's wireless power transfer capability if available. In case the floor is equipped with the optional whole-area WPT solution (still in development), local navigation can then also be facilitated using the measured electromagnetic field intensity [28]. As a buffer between available WPT spots, and for use-cases where WPT is not widely available, SCOOTY contains a small Lithium battery. The solution is designed with a minimal cost goal, thus it uses a standard, quickly

pluggable power tools battery pack. Because of its high production volume and good international availability, this can be sourced for a competitive price. Further, the battery replacement is left as a task for the human workers, refraining from complicated automatic charging or battery replacement procedures. For example, when loading or unloading the cargo at manual work stations, sliding in a new battery and putting the empty one on a charger is a small extra chore to be done on demand. Fixed charging stations in contrast cause longer standstill times for the AGV, reducing its availability. Collaboration with human workers is embraced in the production concept on multiple levels and a simple, pragmatic approach preferred over elaborate automation. Regarding the nature of SCOOTY as the basis for a flexible AGV construction kit, this preference is however not set in stone, but can be adapted as different requirements arise.

5   Communication Using Dynamic LED Tracks

As the material flow through the production becomes more granular, the number of AGVs in operation increases. That leads to challenges regarding the communication paths between all AGVs and other, possibly centralized, coordination entities. Especially if real-time control information is to be exchanged, reliably low communication latency and possibly high data throughput may be required. But these requirements in general do not correspond well with wireless technologies such as Wireless LAN (Wi-Fi) or 5G cellular where the available spectrum always represents a shared resource among all participants.

At this point, the unique technical features of the IF enable some new approaches to solve these challenges. First and foremost, the AGVs can use the dynamically controlled LED strips as guidance tracks to exactly follow their intended pathways. These can however be adapted very quickly in response to routing changes, obstacles, or to avoid traffic jams especially at intersections, while still sticking to the basic rectangular grid. Only in order to cut corners, switch between lanes, or reach a WPT charging spot would an AGV temporarily need to leave the guiding track. The tight 620 mm grid limits the possible error in accuracy when positioning freely between the tracks in these special cases.

The second advantage is the active lighting, which allows much more robust detection than the classical contrast-based optical tracks which are harder to distinguish from random floor markings and stains, and require separate elaborate illumination. Especially the motion blurring effect of a moving camera affects this aspect, which is best mitigated by a minimal shutter time, thus requiring high light intensities in return.

The third and most substantial new solution derives from the possibility to fully control every single LED with an individual color. This enables the floor to signal a small amount of meta-information along with the geometrical information of where the track lies. Rather than distinguishing only between e.g. blue, green and yellow tracks, the exact sequence of LED colors allows to encode a few bits of information tied to an exact location. The design goals for the implemented approach encompass the following:

1. Give each track an orientation, marking where it comes from and which way to go. This is especially useful for omnidirectional AGVs, which can decide to start in either direction without first changing their orientation.
2. Announce path changes well ahead of reaching the point where the AGV needs to take action. In the perpendicular floor grid, slowing down and initiating an intermediating turn *before* reaching an intersection point with a hard 90° corner makes sense for fluid and quick motion.
3. Mark exact geometrical locations such as stopping points and start of a defined curve radius to cut corners, hitting the perpendicularly adjoining next track segment again exactly.
4. Convey additional command and context information while an AGV is standing still, e.g. for state reconfiguration or to initiate a charging process.

The chosen basic pattern for information encoding respects the physical properties of the LEDs and cameras. Each light point on the strips contains three individual LED components in red, green and blue color. Each pixel in commonly used digital image sensors actually consists of a matrix of three sensors with filters for the same three basic colors. Thus, the atomic unit of information is whether each of the three colors are lit or not, leading to $2^3 = 8$ combinations. All three off corresponds to the "black" state, which cannot be distinguished from the surroundings. Similarly, the "white" state is reserved as the only one with all three basic colors lit, and exempted from information encoding, which also limits the needed electrical current. The remaining combinations of one or two LED colors are labeled **R**ed, **G**reen, **B**lue, **Y**ellow, **P**ink and **T**urquoise, totaling six basic code points.

These code points are grouped into symbols of size three, with $6^3 = 216$ possible permutations. This number is chosen deliberately, as it represents the smallest group size where an orientation can be distinguished from any subsequence within a stream of repeated symbols. For example, a sequence of $...\mathrm{R\,G\,R\,G\,R\,G}\,...$ is equivalent to its reversed version $...\,\mathrm{G\,R\,G\,R\,G}\,\mathrm{R}\,...$, thus losing the direction. With three code points per symbol, this becomes unambiguous, as each symbol of three different code points has a defined reading direction, leaving only possible $\frac{6!}{(6-3)!} = 120$ partial permutations. These permutations contain three rotated and another three matching mirrored symbols, so in total only $\frac{120}{6} = 20$ unambiguous, repeatable and oriented symbols remain. They can be used to convey varying state while following a track, such as the chosen pattern $\boxed{\mathrm{B\,G\,R}}$ denoting a standard track to be followed with straight alignment in the direction from blue over green toward red, as illustrated in Fig. 6.

Other defined patterns may include announcement of upcoming corners, e.g. a 90° left or right turn. The corner radius must be chosen carefully respecting the 620 mm grid size. When seeing such a code, SCOOTY deliberately departs from the track to perform a hard-coded turn with constant radius of approx. 300 mm, trying to find the new track again after a predefined turning angle with respect to the previous track. Other information may include the AGV orientation, since omnidirectional vehicles might also follow the track with their
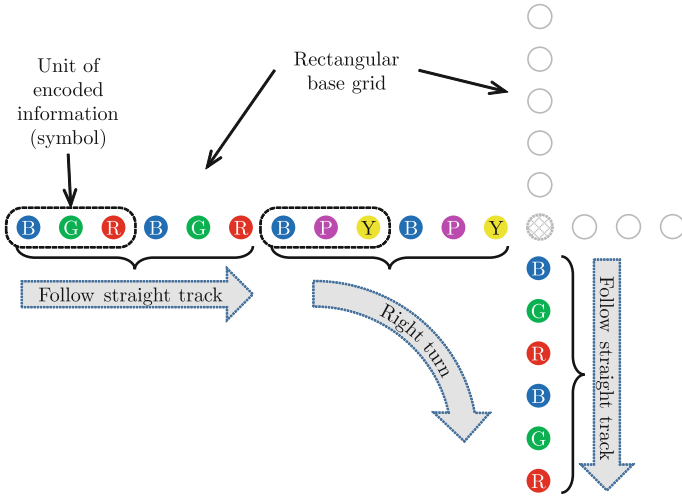
Fig. 6. Example showing a right-turn with appropriate encoding before and after the intersection.

main axis perpendicular to the track for example. Switching to lower or higher speeds in light of an upcoming corner or long straight run may be another possibly conveyed information. Such signals could be encoded as a single three LED symbol within the track, to make it clear where exactly the indicated state applies.

Some of the remaining symbols, which lack an orientation, may be defined as an alphabet to signal arbitrary information when the AGV is standing still. As of yet, the only and most important symbol defined here is the all-red R R R sequence commanding the AGV to stop, which is also easy to recognize intuitively for humans. An encoding for additional commands still needs to be defined, but a time-multiplexed serialization of symbols as used in wired serial port links or other wireless transport protocols appears as an obvious choice. The limited field of view of the track sensor camera can be mitigated easily in this situation, as the approximate location of the track is clear with only a single LED in sight, which is sufficient for the vehicle to re-align its camera to cover as much track as possible.

An important concern is that there will be no return path from the AGV back to the IF. This property certainly limits the possible data throughput because of a lack of explicit synchronization or acknowledgments. The collaboration concept however embraces this limitation, as the AGVs should have very little valuable information to return to a central control entity. They basically follow what the floor indicates through its lighted tracks, and only give occasional status updates when required. For this usage pattern, the scalability problems of e.g. Wi-Fi as a second out-of-band communication channel are much less relevant, mitigated by first avoiding high-bandwidth real-time communication and shifting

the remainder to the LED channel. That is one aspect where the Intelligent Floor offers a novel approach, as most of the complexity is moved from the AGV into the floor controllers, and the vehicles themselves are purpose-built to concentrate on the actual motion. This allows SCOOTY to even eliminate most safety sensor components, based on the assumptions that the IF will signal whether the path is clear of obstacles, and the legally required safety precautions being much lower for a vehicle of very limited hazard potential because of its low mass and speed.

In the envisioned production environment, showing complete tracks from start to finish for each AGV will very quickly lead to a wild maze of different patterns being displayed, with overlapping sections of different AGVs' tracks possibly being ambiguous. Thus, the tracks for each AGV are meant to be shown only in its immediate surroundings, only few segments ahead. This provides the required predictability for humans in the vicinity to see where the AGV is heading. On the technological level, the IF's load sensing capability can be combined with other location sources to determine the vehicle position and deduce the area where the track should be displayed.

6   Conclusion

The realization of a modular production requires among other things the flexibilization of the material flow, whereby the adaptation to newly defined production specifications has to be carried out as quickly as possible without time-consuming programming of individual production and logistics elements. The goal is to break down all production equipment into location-flexible modules in order to be able to dynamically configure and resolve machine systems as well as to end the separation between value creation and logistics which is common today. To enable this flexibilization, a dynamic and automated real-time adaptation of the intralogistics flows of goods is necessary.

Classical logistics systems with their rigid material flow planning have reached their limits. For this reason, new logistics systems must be able to transport the goods directly from the storage location to the point of use automatically, in the appropriate quantity and at the right time. The presented AGV SCOOTY in cooperation with the Intelligent Floor fulfills the requirements for future logistics systems. The IF also offers other interesting possibilities for manufacturing automation and human/machine interaction.

References

1. Wittenbrink, P.: Handlungsbedarf Transportmanagement. In: Transportmanagement, pp. 1–20. Springer, Wiesbaden (2014). https://doi.org/10.1007/978-3-8349-3825-1_1
2. Ramsauer, C., Rabitsch, C.: Agile Produktion - Ein Produktionskonzept für gesteigerten Unternehmenserfolg in volatilen Zeiten. In: Industrial Engineering und Management. TFP, pp. 63–81. Springer, Wiesbaden (2016). https://doi.org/10.1007/978-3-658-12097-9_4

3. Hämmerle, M.: Personal-Flexibilisierungsinstrumente in Produktionsunternehmen. In: Spath, D., et al. (eds) Neue Entwicklungen in der Unternehmensorganisation, pp. 545–565. Springer Berlin (2017). ISBN: 978-3-662-55426-5. https://doi.org/10.1007/978-3-662-55426-5_66

4. Mayr, A.: Veränderungen im Kostenmanagement durch die Digitalisierung. In: Controlling – Aktuelle Entwicklungen und Herausforderungen, pp. 137–161. Springer, Wiesbaden (2019). https://doi.org/10.1007/978-3-658-27723-9_6

5. Gogoll, V.: Erarbeitung des aktuellen Forschungsstandes Industrie 4.0 im Bereich Maschinenbau: Geschäftsmodell-Innovationen in Industrie 4.0 (2017)

6. Trautner, T.F.: Agile Automatisierung von Fertigungszellen- Semiotische Interoperabilität zustandsbehafteter, deterministischer Anlagenkomponenten. PhD thesis. Technische Universität Wien, pp. 1–128 (2021). https://doi.org/10.34726/hss.2021.76044

7. Trierweiler, M., Bauernhansl, T.: Reconfiguration of production equipment of matrix manufacturing systems. In: Weißgraeber, P., Heieck, F., Ackermann, C. (eds.) Advances in Automotive Production Technology – Theory and Application. A, pp. 20–27. Springer, Heidelberg (2021). https://doi.org/10.1007/978-3-662-62962-8_3

8. Stillig, J., Parspour, N.: Novel autonomous guided vehicle system for the use in logistics applications. In: Weißgraeber, P., Heieck, F., Ackermann, C. (eds.) Advances in Automotive Production Technology – Theory and Application. A, pp. 424–432. Springer, Heidelberg (2021). https://doi.org/10.1007/978-3-662-62962-8_49

9. Kirks, T., et al.: Zellulare Transportfahrzeuge für flexible und wandelbare Intralogistiksysteme. In: Logistics Journal 2012 (2012). ISSN: 2192-9084. https://doi.org/10.2195/lj_Proc_kirks_de_01210_01. http://nbnresolving.de/urn:nbn:de:0009-14-34543

10. Wurman, P.R., et al.: Coordinating hundreds of cooperative, autonomous vehicles in warehouses. AI Mag. **29**(1), 9 (2008) https://doi.org/10.1609/aimag.v29i1.2082. https://ojs.aaai.org/index.php/aimagazine/article/view/2082

11. O'Brien, M.: As robots take over warehousing, workers pushed to adapt. AP News (2019). https://apnews.com. Accessed 20 June 2022

12. Hippenmeyer, H., et al.: Ein neuartiges Element für zukünftige Materialflusssysteme: KARIS - dezentral gesteuert. In: Hebezeuge Fördermitte **49**(6), 304–30 (2009). https://www.intralogistik-bw.de/wpcontent/uploads/2018/06/Artikel_KARIS_HF0609_304-309.pdf

13. Stillig, J.: Doppelbodenelement für einen Doppelboden. U.S. Patent 20200173177 (2020)

14. Mandat, M.: Produktportfolios der fünf großen deutschen Autohersteller in Deutschland: 1990 bis 2014: Das Dilemma mit der Vielfalt. PROGENIUM GmbH & Co.KG. (2015). https://docplayer.org/15484141-Produktportfolios-der-fuenf-grossen-deutschen-autoherstellerin-deutschland-1990-bis-2014-das-dilemma-mit-der-vielfalt-muenchen-24.html. Accessed 04 Nov 2019

15. Volkswagen. Absatz des VW Golf im Zeitraum der Jahre 1974 bis 2012 nach Modell. Handelsblatt (2012). https://de.statista.com/statistik/daten/studie/240184/umfrage/absatz-des-vw-golf-nach-modell. Accessed 04 Nov 2019

16. Specht, D., Stefanska, R.: Wandlungsfähige Fabrikstrukturen als Strategie: Zur Steigerung der Wettbewerbsfähigkeit von produzierenden Unternehmen. Zeitschrift für wirtschaftlichen Fabrikbetrieb **102**(5), 286–290 (2007). https://doi.org/10.3139/104.101139

17. Kühnle, H.: Paradigmenwechsel im Produktionsbetrieb - Die Fraktale Fabrik. In: Warnecke, H.J., Bullinger, H.J. (eds.) Produktionsstrategie für das 21. Jahrhundert, vol. 44, pp. 9–43. Springer, Berlin (1994). https://doi.org/10.1007/978-3-642-52359-5_1

18. Westkämper, E.: Marktorientiertes Produzieren in dynamischen Strukturen. In: Warnecke, H.J. (ed.) Fabrikstrukturen im Zeitalter des Wandels - Welcher Weg führt zum Erfolg?, vol. 47, pp. 9–22. Springer, Berlin (1995). https://doi.org/10.1007/978-3-662-07163-2_1

19. Westkämper, E.: Produktion in Netzwerken. In: Schuh, G., Wiendahl, H.P., Komplexität und Agilität - Steckt die Produktion in der Sackgasse?, pp. 275–291. Springer, Berlin (1997) https://doi.org/10.1007/978-3-642-60841-4_19

20. Westkämper, E.: Stuttgart: Institut für Industrielle Fertigung und Fabrikbetrieb (IFF), Stuttgart, Technical report (2006). https://gepris.dfg.de/gepris/projekt/5480668

21. Günthner, W., et al.: Intralogistik im Dialog mit Forschung und Lehre. In: Arnold, D., (eds) Intralogistik - Potentiale, Perspektiven, Prognosen, pp. 239–276. Springer, Berlin (2006). https://doi.org/10.1007/978-3-540-29658-4_6

22. Matthias Hofmann and David Korte. "Neuartiges Logistikkonzept für die automobile Endmontage ohne Band und Takt". In: Bauernhansl, T., et al. (eds) Entwicklung, Aufbau und Demonstration einer wandlungsfähigen (Fahrzeug-) Forschungsproduktion, pp. 85–100. Springer, Berlin (2020). https://doi.org/10.1007/978-3-662-60491-5_8

23. Stillig, J., Parspour, N.: Novel infrastructure platform for a flexible and convertible manufacturing. Adv. Sci. Technol. Engi. Syst. J. **6**(1), 356–368 (2021). https://doi.org/10.25046/aj060141

24. Stocker, J.: Flexible Intralogistiksysteme für die fahrerlose Transportversorgung - Eine Lebenszykluskostenanalyse. Master Thesis. Hochschule Albstadt-Sigmaringen (2022)

25. Brenner, C., Colomb, A.: Konzept zur intuitiven Steuerung omnidirektionaler Flurförderzeuge mit beliebiger Radkon guration. Logistics J.: Proc. 2020 (2020). https://doi.org/10.2195/lj_Proc_colomb_de_2020120_1. http://nbn-resolving.de/urn:nbn:de:0009-14-51330

26. Schröppel, M., et al.: Monofunktionale autonome Transporteinheiten. Logistics J.: Proc. **7**(1) (2011). https://doi.org/10.2195/LJ_proc_schroeppel_de_201108_01. http://nbn-resolving.de/urn:nbn:de:0009-14-31016

27. Brenner, C., Colomb, A.: Anwendung der Omni-Kurven-Parameter zur Bestimmung der Aktor-Stellgrößen und universellen Bewertung der Bewegungsmöglichkeiten unterschiedlicher Fahrwerke. Logistics J.: Proc. 2022 (2022)

28. Brenner, C., et al.: Navigation of an Omnidirectional AGV with an Inductive Floor. In: Forschung im Ingenieurwesen 2022 (2022)

Topology Planning in Swarm Production System: Framework and Optimization

Akshay Avhad[(✉)], Casper Schou, and Ole Madsen

Department of Materials and Production, Aalborg University, Fibigerstræde 16, 9220 Aalborg East Aalborg, Denmark
akshayra@mp.aau.dk

Abstract. A Swarm Production System (SPS) aims to be an agile and resilient Reconfigurable manufacturing system (RMS) paradigm that incorporates mobile workstations and transport robots on the factory production floor. This paper primarily focuses on SPS's initial but recurring planning stage termed topology planning, which dynamically changes throughout the production runtime with spatially adaptive workstations and transporters handled exclusively by a Topology Manager (TM). TM is essential to multi-variant production with the optimal positioning of the workstations and provides a topology that optimizes the traffic flow for the product carrier robots. TM is a bridge to enable SPS to integrate with general planning and scheduling systems like ERP and MES and is comprised of a Topology Planner (TP) that evaluates the ideal configuration of on factory floor for a batch of product mix and a Reconfiguration Decision System (RDS) that decides on applying the estimated new topology during the batch changeover. The paper proposes a framework for the TM to identify its essential functionalities, responsibilities and working principle in a swarm production system. The paper also describes a grid-based heuristic approach applicable to two-dimensional spatial problems to reduce the complexity of the NP-hard problem. The paper focuses on a framework to estimate a reconfigurable shop floor layout with a Force-directed Graph-theory algorithm. A stochastic statistical model evaluates the performance of the optimal topology for throughput and makespan.

Keywords: Swarm Production System · RMS · Topology Manager · Force-directed · Statistical model

Nomenclature

λ_v	Throughput for vth PI
BT	Makespan for whole batch
$BTcurr$	Makespan for deployed topology
C	Cost function in Topology estimation and optimization
CT_1v	CycleTime for 1st product in vth PI
CT_2v	CycleTime for all product except the 1st in vth PI

DT_v	Dispatch timestamp of last product on 1st WR for a vth PI
ET_v	End timestamp for a vth PI
I	unit time loss due to congestion on crossings
n	number of WRs in a PI
P_i	Process time of a ith WR, where i ϵ [1,n]
Q_v	Quantity to be produced for a PI in production
RT	Reconfiguration span for topology
S_v	Speed of TR for a PV in production
ST_v	Start timestamp for a vth PI
T_v	Total travel length of PV subgraph
t_v	Makespan for a vth PI
TT_v	Expected Takt-time for a vth PI
V	maximum number of PV in a batch
v	PV number in a batch, where v ϵ [1,V]
X	Stochastic variable with Uniform distribution where X ϵ [1,x]
x	Maximum number of crossings on PV subgraph
BPM	Batch Processing Module
DES	Discrete Event Simulation
DLP	Dynamic Plant Layout
ERP	Enterprise Resource Planning
FDP	Force-Directed Placement
FLP	Factory location Problem
GA	Genetic Algorithm
LMAS	Line-less mobile assembly system
MES	Manufacturing Execution System
MILP	Mixed Integer Linear Programming
PI	Product Instance
PV	Product Variant
QAP	Quadratic assignment problem
RDM	Reconfiguration Decision Module
SPS	Swarm Production System
TEOM	Topology Estimation and Optimization Module
TM	Topology Manager
TR	Transfer Robot
WR	Workstation Robot

1   Introduction

The concept of Swarm Production System (SPS) proposed in [1] is a more flexible production flow concept compared to the known manufacturing paradigms like Assembly line and Matrix production. In an SPS, the workstations and product conveyances between stations are mobile entities, which can be placed in any location suitably. The main objective is to improve responsiveness to the market's need for producing batches of different of product variants (PV). Each PV has an optimal workstation layout on the shop floor, allowing cost-efficient

volume production. The cost for producing a PV largely depends on resource allocation of workstations and part carrying robots and the cumulative inter-workstations travel length. The efficient SPS planning is pivoted on optimal travel cost in a shop floor layout as resource allocation is a task for ERP and MES enterprise systems. The scope of any production system spans Planning, Scheduling and Control; it applies to SPS too.

The planning stage in a production system starts with identifying resources such as machines, actuators, sensors, and workforce for the assembly operation. An assembly or process workstation is an entity that hosts most of these resources as a unit, termed a Workstation robot [WR] in SPS. Every WR has links to other WRs; a chain of these links forms sequences that enables a PV production. A link indicates a direction of material flow carried by a product carrying Transfer Robot [TR]. In the scope of an SPS, a WR can have multiple linkages depending on the number of PV in production, forming a graph structure with x and y positions for the node WR and edges representing the linkages between WRs. These graphs are a topological structure that enables SPS to produce a given set of PVs in a batch with efficiency in process execution. Each batch has an optimal topology. The optimal topology mandates time efficiency in batch production with better throughput and cycle time. The topology also lays a foundation for subsequent scheduling and control activity in SPS with enumeration and task allocations for TRs in sequential assembly operation for a PV. Identifying optimal positions for every WR in the topology is a combinatorial NP-hard problem with factorial time complexity. Therefore, a multi-step optimisation is proposed in the topology estimation problem in SPS. As shown in the Fig. 1, a Topology Manager (TM) handles the planning stage of identifying, estimating, and optimising the topology in SPS. External ERP and MES systems provide high-level planning and scheduling information essential to the TM's initialisation. Furthermore, the SPS contains a Swarm manager, which executes process level tasks on WRs and TRs.

This paper presents a framework for TM to orchestrate the topology planning, initiating the production process with the transfer of batch information from ERP and MES in the factory and ending with a local optimum topology for production.

2   State of the Art

Operations research has extensively studied the factory layout problem (FLP) associated with optimally localising manufacturing facilities to reduce cost. The nature of the problem of SPS topology planning has similarities with the FLP with the placement of WR on a shop floor.

2.1   FLP in Changeable Manufacturing

SPS is an applied case of conceptual RMS with a practical production philosophy involving autonomous WR and TR entities on the shop floor. The most practi-
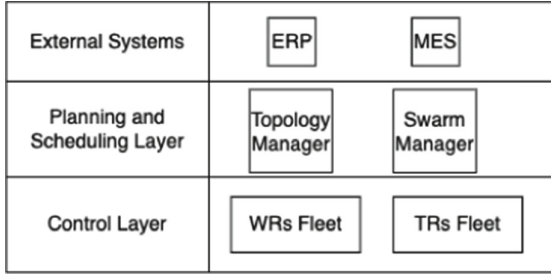
Fig. 1. SPS system level context [1]

cal research questions in the reconfigurable scenario is around addressing adaptability and scalability of transportation within production and deploying the dynamic layout over two dimensional plane [2]. Rosenblatt [3] first addressed the problem of dynamic plant layout (DLP). Heuristic-based solutions and dynamic programming techniques are proposed in [3], and [4] for DLP. Kusiak and Heragu [5] found that heuristics yield near-optimal, computational light solutions relevant to non-uniform spatial vacancies in Flexible Manufacturing Systems (FMS). A hybrid genetic algorithm has been implemented in [6] to optimise continuously change layout requirements in RMS. The effectiveness of this algorithm prevails over the standard genetic algorithms due to its broader search spaces capability. Metaheuristic techniques combined with deterministic ML algorithms are efficient in solving combinatiorial optimization problem in Changeable Manufacturing Systems [7]. The selection of optimal factory configuration is critical to quantifying machines, equipment, robots, and task assignments to all these entities. Line-less Mobile Assembly System (LMAS) [8], is a flexible production paradigm, incorporates a statistical assessment model created in [9] for the early planning stage; in comparison to a discrete event simulation (DES) model that is cumbersome to build in the absence of a suitable scheduler. A cost and time-driven dual approach [10] is proposed for task and location assignment in the planning of LMAS. Therefore, dividing shop floor area into a grid with uniform squares to reduce the time and computational complexities of estimating discrete x and y parameters instead of continuous ones.

2.2   Optimization Methods in FLP

Multiple FLP design problems in [11,12] are addressed with Mixed Integer Linear Programming (MILP) optimization methods. In [13] a branch and bound approach coupled with MILP is ineffective in solving a large-size problem, while metaheuristic algorithms fare better in comparison. MILP-based solvers shown to have exponential time complexity from medium to large grids impart low practicality in real-world FLP problems in [14]. Hybrid metaheuristic-based experiments performed in [15] for Capacity-based FLP optimally locate factories with the demand such that overall cost due to operation and product transportation

is minimal. Sets of metaheuristic solutions like Simulated Annealing (SA) and Genetic Algorithm (GA) are applied to the dynamic reconfiguration of factories in [16,17], and [18]. Quadratic assignment problem (QAP) in [19,20] addresses peculiar FLP where the cost is cumulative of distances between facilities and number flows between them. Various QAPs have been used in cross-disciplinary planning facilities for hospitals, supermarkets, and also in precision demanding electronic circuits design are presented in [19–21].

2.3   Optimization with Graph Theory

A two-stage cost optimisation model in [22] applies graph theory for evaluating the initial solution based on shortest path constraints followed by a selection of more optimal configurations from the first stage. The solution to FLP discussed in [23,24] could be modelled as an optimal location solution for vertices of a graph on 2D space with the edge weight representing cost. The spatial layout is optimised, transforming the supergraph into a subgraph that retains the parent's logical edge connections into a topological form, eventually marking it as a graph theory problem in [19,25].

2.4   Optimal Approach in Topology Manager(TM)

Most factory layout problems are centred around static location planning, and numerous reasonable heuristic solutions can be derived from them in SPS to perform the assembly operation of a batch mix. The cost and time-tradeoff are essential factors while planning SPS; therefore, a near-optimal approach in an ideal time frame is the best possible solution for a TM. Standalone MILP and metaheuristics are not enough to tackle the large search space needed in identifying topologies on a two-dimensional plane as it indefinitely increases the time and computational complexities as mentioned in [26–28].

The linear increase in topology size with WRs exponentially increases the search space in planning the optimal topology. Most heuristic solutions are deployed in a constrained-based scenario and a fairly static location planning objective. SPS needs a solution capable of identifying dynamic near-optimal topology in a viable time frame. Hence in the following, we will first propose several concepts in a TM. An example follows this at a practical implementation of TM, uncovering some of the abovementioned issues.

3   Topology Manager Framework

The outline of a TM framework is shown in the Fig. 2. The TM framework is based on utilising the existing enterprise software infrastructure of ERP and MES. The prerequisite to the TM planning process includes a PV list from products scheduled over the next day and an enumerated sequence of WRs for

every PV in the list. The pre-planning database hosts the prerequisite data that Batch Processing Module (BPM) later retrieves and processes into batch data. A batch data model is an aggregated data structure for the number of PVs in daily production and their WRs sequences. As seen in the Fig. 2 the output of the BPM is a logical topology comprising sets of WRs and their connections for every PVs in the batch without spatial information. The Topology estimation and optimization module (TEOM) in the Fig. 2 identifies a spatial topology based on the logical topology from BPM and optimize before the start of the production of the batch. The TEOM identifies different topologies relevant to the nature of the batch data. Different methods based on graph theory and heuristics are applied in TEOM to identify and optimize the search space for the near-optimal topologies. The set of topologies is the outcome of the TEOM and is forwarded to the Reconfiguration Decision module (RDM) for the decisive deployment of the most optimal topology on the production floor. The last module in Fig. 2 is an RDM, an inference engine for the selection of the most optimal spatial topologies and decision on the changeover from the currently deployed topology. The changeover process implies a temporal loss in production due to downtime. A reconfiguration process is triggered only if the sum of reconfiguration time loss and the production time with the new topology is less than the production time with the existing topology. In short, the reconfiguration process is skipped when the existing topology fares better, considering the reconfiguration cost. The production topology database has the final plan to be ready for SPS runtime production.

4   Exemplification

The TM represented a generic framework for macro-level planning inside an SPS. Practical implementation requires defined data structure, methods and algorithms in every stage of the TM. The problem TM tries to achieve is a most optimal topology that estimates locations for WRs and the material flow enabled by the topology. We do not consider the processes on WRs nor their flexibility and redundancy, assuming a single process per workstation. Thus the planning goal becomes to optimise the material flow and thereby the distance between WRs and the potential congestion between TRs.

4.1   Batch Processing Module

In Fig. 3 a multi phase process depicting data flow from ERP and MES to a structured, logical topology representing an SPS batch is shown. A logical topology is an undirected graph data structure with nodes representing a set of WRs for all PVs in a batch and the linkages between WRs as the edges. Phase 1 describes a database cluster hosting a separate schema for daily production PVs and WRs sequence data for each PV. Phase 2 is the interface between TM and the pre-planning database retrieving the PVs list and WRs schema. Lists of WRs are extracted from the phase 2 data depending on the PVs in the production list and
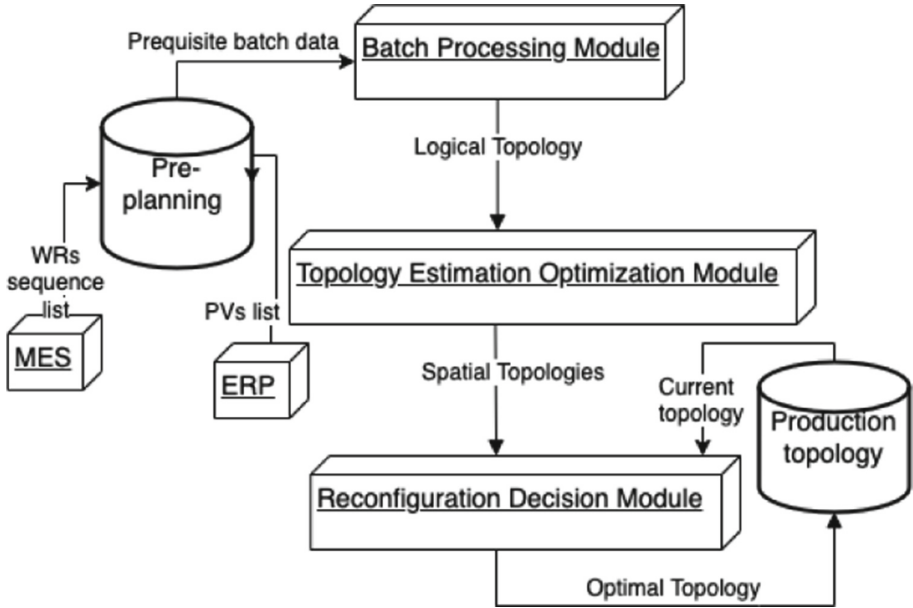
Fig. 2. Topology Manager Framework

aligned in the same order as PVs in the production list, and phase 3 represents this sorted 2D list data form, also known as batch data. The terminology PVs is replaced with Product Instance (PI) in phase 3, indicating that the PV is a product template before being enlisted in a batch. At the same time, PI is the physical entity associated with the PV in production. The batch data from phase 3 is converted to a logical 2D graph topology denoted by G = {V, E} where V represents a set of WRs nodes and edges E retaining the information from the WRs sequences.

4.2   Topology Estimation and Optimization Module

Different graph theory-based approaches are undertaken in TEOM to generate an optimal topology from the input logical topology. The logical topology represents a graph for a complete batch, while every PI in a batch is a subgraph of batch topology.

Objective Function. SPS differs significantly from a conventional production philosophy; hence, it is at the preliminary stage to understand the cost required to produce a PV. Since the planning stage demands topologies relevant to a batch of PVs, travel distances between the WRs influence the makespan and are hence used as a cost function. Throughout the estimation and optimization process, objective function is based on cumulative eucliedean distances between
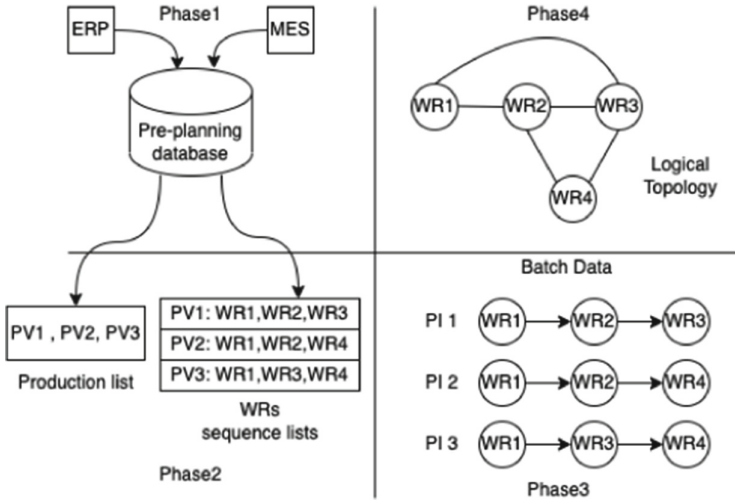
Fig. 3. Formation of Batch and Logical Topology

WRs in PI subgraphs. As mentioned in Eq. 2 the cost of a topology is the total travel distance required to visit every WR in a sequence for each PI in a batch. The objective function for estimation and optimization stage is minimal travel cost for the complete batch as stated in Eq. 1

$$C = min \sum_{v=1}^{V} T_v \qquad (1) \qquad T_v = \sum_{i=1}^{n-1} d(i, i+1) \qquad (2)$$

where, C = Cost of a batch topology
T_v = Travel cost in a PI subgraph
v = PI enumeration in batch
i = WRs enumeration in PI subgraph
n = Total WRs in a PI
d = euclidean distance between WRs

Factory Planning with Logical Topology. The placement of workstation nodes on the shop floor is the layout deployment to enable a batch of multiple PVs. Ideally, the workstations could be placed at the closest possible locations to minimize the conveyance time of the product after every process cycle on individual WRs. The distance between the WRs is constrained by factors like minimum spacing required for TR navigation and structural constraints (safety and environmental blockages). Therefore, WR nodes cannot be placed in an overlapped topological configuration even if it establishes a global minimum cost for production but unrealistic in a physical scenario. The Logical topology

provides the connected planar graph, and every WRs require spatial coordinates based on the minimal cost function in Eq. 1.

Estimation and Optimization with Spring Topology. The freedom in placement of the WR nodes from logical topology increases the complexity of the problem. The edges represents a preliminary path between the nodes which can be redefined in the later stage of path planning for TRs. Physical analogy embedded in a graph with every edge as a spring force that attracts the connecting nodes in the logical topology provides effective heuristic handling for undirected graphs [29,30].

A graph layout algorithm for drawing positions on a plane known as Force-directed Placement (FDP) [31] layout injects a repelling spring force among the nodes while expanding and contracting the edges in the whole process. FDP tries to draw positions based on the principles of uniform nodes distribution, minimal edge crossings, and uniform edge length but does not guarantee the implementation of these principles in the final layout [31,32]. An implementation is done using Networkx python API spring layout [33] that uses the FDP algorithm to draw the position on a logical topology without any spatial information. The implicit parameters to this API are the repulsive force (k) value and a number of iterations (ITR) determining the node separation on a planar surface and the maximum iterations required to draw the graph, respectively. Topologies based on Networkx spring are displayed in Fig. 4 for different values of K, illustrating swelling of the topology as the repulsive forces with increasing value of K.

(a) $k = 0.70$ (b) $k = 1.30$
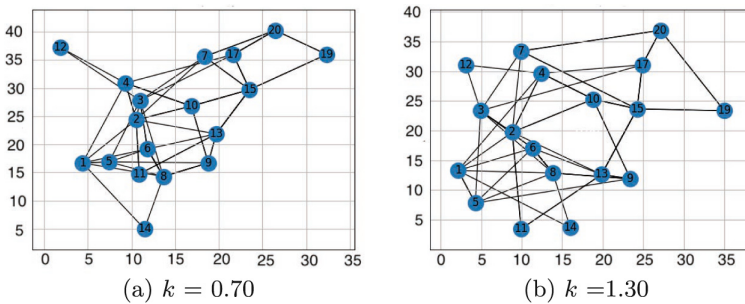
Fig. 4. Spring FDP topologies

4.3 Reconfiguration Decision Module

The RDM is the inference engine for selecting the topology with the least production cost in Eq. 1 from the set of spatial topologies generated by the TEOM. Since the topology is optimized over the cumulative travel cost of every PV in a batch, an additional layer of a performance assessment model is required to

evaluate the potential of topology in terms of production KPIs, e.g. throughput and cycle time. The edges in the topology in the SPS planning stage indicate job routing paths between the WRs. Therefore, the stochastic losses due to congestion causing time delay in TRs on overlapping edges are prominent in the Spring topology. A dynamic Discrete Event Simulation (DES) model provides a test-bed for testing scheduling algorithms, which eventually predicts SPS's PI and batch-specific KPIs. Such a DES model is not built yet for an SPS. Hence, in the absence of a suitable scheduler for SPS, a statistical model is described in the equations below with integer-valued uniform distribution stochastic variable X representing the number of occurrences of congestion. The dispatch time of the final product from v^{th} PI from the first WRs in a topology indicates the vacancy for the next PI loading. The product leaving the first WR is dependent on the cumulative process times of subsequent workstations and the stochastic time losses during the TR's conveyance and therefore, it can be written as

$$DT_v = P_1 + \sum_{i=2}^{n} (P_i + I \cdot X) \tag{3}$$

where P_i is a process time on workstation with range [1,n] and X represents uniform distribution stochastic in range [1,x] with a unit time loss of I on every crossing. The start time of a PI depends on the dispatch time of all quantities from the previous PI and its start time as seen in Eq. 4.

$$ST_v = \begin{cases} 0, & \text{if } v = 1 \\ ST_{(v-1)} + DT_{(v-1)}, & \text{otherwise} \end{cases} \tag{4}$$

The end timestamp ET_v of vth PI is evaluated in Eq. 5.

$$ET_v = ST_v + t_v \tag{5}$$

where, t_v represents total makespan for the vth PI of quantity Q_v and stated below. The end times stamp ET_V of the last product V for PI in the batch provides the total required batch production time as seen in Eq. 6.

$$BT = ET_V \tag{6}$$

Equation 7 is for calculation of makespan for each PI with quantity Q and estimated throughput λ.

$$t_v = \frac{Q_v}{\lambda_v} \tag{7}$$

where, Q_v is total quantity to be produced for v^{th} PI and λ is the throughput for v^{th} PI .Throughput calculation in Eq. 8 is based on cycle time for 1st product shown in Eq. 9 and cycle time for later products shown in Eq. 10.

$$\lambda_v = \frac{1}{CT_1 v + CT_2 v} \tag{8}$$

$$CT_1v = \sum_{i=1}^{n} P_i + \frac{T_v}{S_v} + I \cdot X \tag{9}$$

$$CT_2v = min(TT_v) + I \cdot X \tag{10}$$

where, T_v denotes total travel cost in v^{th} PI subgraph topology scaled by TRs at a speed S_v for that PI and TT_v is an expeted takt-time for vth PI. In the final stages a local optimum topology OT is selected with minimum batch production time $BTmin$ from Eq. 11.

$$OT = BTmin \tag{11}$$

The reconfiguration depends on the performances of the newly found near-optimal topology against the currently deployed topology. Reconfiguration in Eq. 12 is performed only when the sum of the changeover span and the batch production time does not exceed the makespan with the current deployed topology $BTcurr$.

$$RCF = \begin{cases} 1, & \text{if } RT + BT < BTcurr \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

where, RT denotes required reconfiguration time for a new topology for a batch with BT production time or makespan.

5 Experimental Results

This section describes the numerical implementation performed in Python for the TM. A test batch with seven PIs were used in the numerical exemplification with quantities and sequences illustrated in Fig. 5. Each PI has two different quantities for standard and larger batch experiments. Uniform process times are assigned to all the WRs in a PI, with every PI having unique process times, as shown in the Fig. 5. The BPM generates a logical topology of the test batch and feeds it to the TEOM. The population of Spring layouts are generated in the TEOM's topology estimation stage with K value from 1.2 to 2.0 with a step increase of 0.2, and ITR value from 0 to 45 with a step increase of 5.0. The process continues until the cost function converges on the objective function mentioned in the Eq. 1. The best candidate from the population of Spring topology is found at values K at 1.3 and ITR at 40. The best Spring topology with minimal cost is displayed in the Fig. 6a. At the same time, the crossings were found on subgraphs for PIs 2,3 and 4 to be 2, 1 and 3, respectively. The variance of the discrete stochastic variable X in Eq. 3 depends on the number of crossings. The random.randit API generates the integer stochastic variable with a lower limit of 0 and a higher limit as the total number of crossings for respective PIs. A grid-based FLP solution based on the optimal Spring topology in 6b illustrates the two-dimensional spatial positions for WR nodes in the test batch.

The optimal Spring topologies is subjected to performance evaluation through a statistical model from Sect. 4.3 in RDM. The results in Fig. 7 are generated for smaller batch sizes and a relatively large batch size with different process times and quantities for individual PIs displayed in the legend of the individual figures. From the Fig. 7a, Spring topology takes 2421 unit time to finish the batch production as compared to 6836 unit time for a larger batch seen in Fig. 7b.

PI	WRs Sequence	Qty	Large	Process times
1	①→⑤→⑨→⑩→②→⑪→⑬→⑮→⑦→⑳	10	100	5
2	①→②→⑦→③→⑤→⑥→⑧→⑨→⑬→⑮→⑲→⑳	30	100	8
3	①→⑤→⑧→⑥→③→②→④→⑩→⑯→⑰→⑳	50	100	10
4	①→⑧→⑨→⑩→②→⑪→⑬→⑮→⑦→⑳	20	100	12
5	①→④→⑰→③→⑧→⑨→⑬→⑮→⑲→⑳	60	100	7
6	①→⑥→⑧→⑯→③→⑫→⑦→⑩→⑮→⑰→⑳	20	100	9
7	①→⑭→⑧→⑥→⑬→②→④→⑩→⑮→⑰→⑳	40	100	13

Fig. 5. Batch for numerical exemplification

(a) $k = 1.3$ and $ITR = 40$ (b) Shop floor layout based on Spring

Fig. 6. Optimal Topology from TEOM

(a) Smaller batch   (b) Large batch
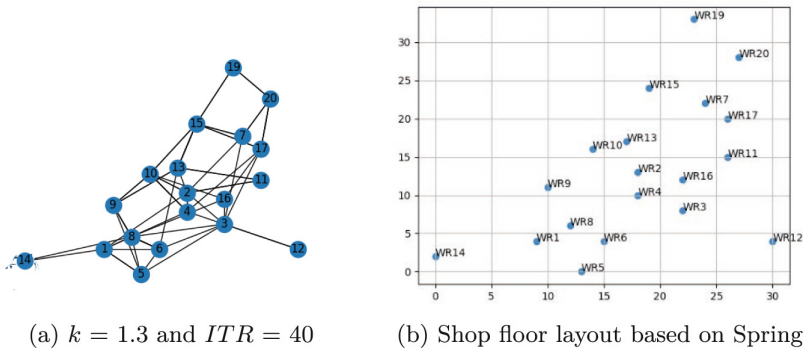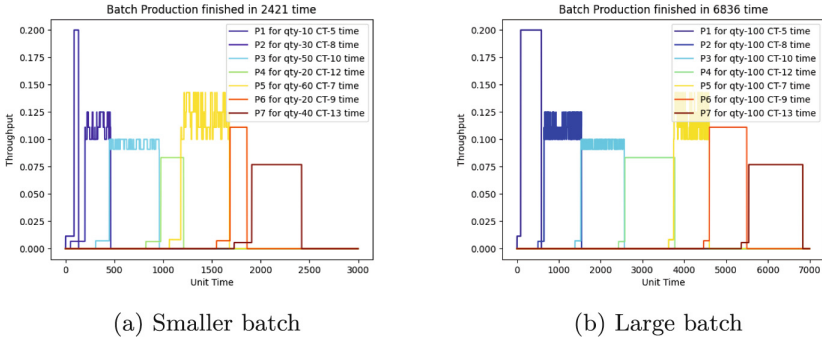
Fig. 7. Performances of the Spring Topology

6   Discussion and Conclusion

A flexible and reconfigurable paradigm like SPS enables and requires continuous
adaptation to a varying batch mix of product variants. A model-based TM frame-
work presented can achieve a strategic planning objective in SPSs by yielding
adaptable topologies to the changing batch mix. The development of the TM is
pivoted on integration with the existing manufacturing software ecosystem and
extending the capability of an enterprise to plan factory layout for a SPS. The
overall goal of the TM is to identify the best possible topology in a defined search
space and decide whether to change the new topology or keep the existing one.
The reconfiguration process will require an SM that executes task for WRs and
TRs on the shop floor.

A near-optimal heuristic approach with graph theory is more computation-
ally viable than a global optimization method as the major challenge is deploying
solutions in a short span. Later exemplified with graph theory FDP based opti-
mization with Spring Topology, and a statistical model in RDM. The grid-based
approach reduces the computational requirement by optimising discrete space
coordinates compared to the continuous ones. The performance KPIs of the
topologies are evaluated in TEOM with a mathematical stochastic model in the
absence of a suitable simulation tool. The results are outcome of a methodol-
ogy that assesses planned topologies for the potential performance before being
deployed on the shop floor. SPS planning objective can be associated with
multiple isomorphic topological graphs apart from the Spring topology. A non-
overlapping edge topology is capable of avoiding stochastic losses due to collision
and therefore, an extensive study is required in this direction to improve the SPS
planning.

Conventional FLP methods are focused on static layout and suited to a
defined set of PVs; on the contrary, graph-based TM provides a faster deliv-
ery of topological layouts that can be adapted to a batch of changing PVs mix.
TM provides a holistic framework that can support relevant graph-based opti-
mization methods apart from Spring topology with an approximated assessment

of the performances of the planning stage. TM also represents a digital twin for SPS planning capable of data modelling, optimization methods and decision making for deploying optimal configuration on the production floor.

Due to congestion, the stochastic nature of losses during material flow is subjected to efficient path planning for TRs in SPS. These uncertainties can be reduced with a topological form that enables a collision-free path for TRs in every possible sequence of PVs in a batch. A comprehensive graph theory-based method shall assist explore topologies to deploy shop floor layouts that can lead to a predictive performance assessment. In the future, a grid-based shop floor design, when incorporated with spatial constraints like safe passages, structural blockages, and no deployment zones, enables a pragmatic planning solution in a real-world factory scenario. Furthermore, solutions requiring more expansive search space can be yielded if applied with metaheuristic algorithms, especially in an upscaled production environment with multiple good solutions.

References

1. Schou, C., Avhad, A., Bogh, S., Madsen, O.: Towards the swarm production paradigm. In: Andersen, A.-L., et al. (eds.) Towards Sustainable Customization: Bridging Smart Products and Manufacturing Systems. CARV MCPC 2021 2021. Lecture Notes in Mechanical Engineering, pp. 105–112. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-90700-6_11
2. Literature review and research trend: Marco Bortolini, Francesco Gabriele Galizia, and Cristina Mora. Reconfigurable manufacturing systems. J. Manufact. Syst. **49**, 93–106 (2018)
3. Rosenblatt, M.J.: The dynamics of plant layout. Manag. Sci. **32**(1), 76–86 (1986). http://dx.doi.org/10.1287/mnsc.32.1.76
4. Conway, D.G., Venkataramanan, M.A.: Genetic search and the dynamic facility layout problem. Comput. Oper. Res. **21**(8), 955–960 (1994)
5. Kusiak, A., Heragu, S.S.: The facility layout problem. Eur. J. Oper. Res. **29**(3), 229–251 (1987)
6. Wei, X., Yuan, S., Ye, Y.: Optimizing facility layout planning for reconfigurable manufacturing system based on chaos genetic algorithm. Prod. Manufact. Res. **7**(1), 109–124 (2019). http://mc.manuscriptcentral.com/tpmr
7. Renzi, C., Leali, F., Cavazzuti, M., Andrisano, A.O.: A review on artificial intelligence applications to the optimal design of dedicated and reconfigurable manufacturing systems. Int. J. Adv. Manufact. Technol. **72**(1), 403–418 (2014)
8. Huettemann, G., Buckhorst, A.., Schmitt, R.: Modelling and assessing line-less mobile assembly systems. In: 52nd CIRP Conference on Manufacturing Systems (CMS), vol. 81, pp. 1–1452 (2019)
9. Schmitt, R.H., Huttemann, G., Munker, S.: A priori performance assessment of line-less mobile assembly systems. CIRP Annals (2021)
10. Buckhorst, A.F., Schmitt, R.H.: Multi-staged, multi-objective optimization for operation management in line-less mobile assembly systems (LMAS). Procedia CIRP **93**, 1121–1126 (2020)

11. Amaral, A.R.S.: On the exact solution of a facility layout problem. Eur. J. Oper. Res. **173**(2), 508–518 (2006)
12. Heragu, S.S., Kusiak, A.: Efficient models for the facility layout problem. Eur. J. Oper. Res. **53**(1), 1–13 (1991)
13. Solimanpur, M., Jafari, A.: Optimal solution for the two-dimensional facility layout problem using a branch-and-bound algorithm. Comput. Industr. Eng. **55**(3), 606–619 (2008)
14. Noor-E-Alam, Md., Mah, A., Doucette, J.: Integer linear programming models for grid-based light post location problem. Eur. J. Oper. Res. **222**(1), 17–30 (2012)
15. Souto, G., Morais, I., Mauri, G.R., Ribeiro, G.M., Gonzalez, P.H.: A hybrid matheuristic for the Two-Stage Capacitated Facility Location problem. Expert Syst. Appl. **185**, 115501 (2021)
16. McKendall, A.R., Shang, J., Kuppusamy, S.: Simulated annealing heuristics for the dynamic facility layout problem. Comput. Oper. Res. **33**(8), 2431–2444 (2006)
17. Dou, J., Dai, X., Meng, Z., Dou, J., Dai, X., Meng, Z.: A GA-based approach for optimizing single-part flow-line configurations of RMS. J. Intell. Manuf. **22**, 301–317 (2011)
18. Maniraj, M., Pakkirisamy, V., Jeyapaul, R.: An ant colony optimization-based approach for a single-product flow-line reconfigurable manufacturing systems. J. Eng. Manufact. **231**(7), 1229–1236 (2017)
19. Loiola, E.M., De Abreu, N.M.M., Boaventura-Netto, P.O., Hahn, P., Querido, T.: A survey for the quadratic assignment problem. Eur. J. Oper. Res. **176**(2), 657–690 (2007)
20. Koopmans, T.C., Beckmann, M.J.: Assignment problems and the location of economic activities. Econometrica **25**(1), 53–76 (1957)
21. Geoffrion, A.M., Graves, G.W.: Scheduling parallel production lines with changeover costs: practical application of a quadratic assignment/LP approach. Oper. Res. **24**(4), 595–610 (1976). https://doi.org/10.1287/opre.24.4.595
22. Dou, J., Dai, X., Meng, Z.: Graph theory-based approach to optimize single-product flow-line configurations of RMS. Int. J. Adv. Manufact. Technol. **41**(9–10), 916–931 (2009)
23. Boswell, S.G.: Discussion a reply to a note on 'similarity of a new greedy heuristic for facility layout by graph theory to an existing approach. **32**(1), 235–240 (2007). http://dx.doi.org/10.1080/00207549408956928
24. Singh, S.P., Sharma, R.R.: A review of different approaches to the facility layout problems. Int. J. Adv. Manuf. Technol. **30**, 425–433 (2006)
25. Kreher, D.L., Stinson, D.R.: Combinatorial Algorithms. CRC Press, Boca Raton (1999)
26. Maganha, I., Silva, C., Ferreira, L.M.D.F.: The layout design in reconfigurable manufacturing systems: a literature review. Int. J. Adv. Manufact. Technol., 683–700 (2019). https://doi.org/10.1007/s00170-019-04190-3
27. Usuga Cadavid, J.P., Lamouri, S., Grabot, B., Pellerin, R., Fortin, A.: Machine learning applied in production planning and control: a state-of-the-art in the era of industry 4.0. J. Intell. Manufact. **31**, 1531–1558 (2020)
28. Benderbal, H.H., Dahane, M., Benyoucef, L.: A new robustness index formachines selection in reconfigurable manufacturing system. In: Proceedings of 2015 International Conference on Industrial Engineering and Systems Management, pp. 1019–1026. IEEE IESM 2015 (2016)
29. Eades, P.: A heuristic for graph drawing. Congr. Numer. **42**, 149–160 (1984)

30. Gansner, E.R., North, S.C.: Improved force-directed layouts. In: Whitesides, S.H. (ed.) GD 1998. LNCS, vol. 1547, pp. 364–373. Springer, Heidelberg (1998). https://doi.org/10.1007/3-540-37623-2_28

31. Fruchterman, T.M., Reingold, E.M.: Graph drawing by force-directed placement. Softw.: Pract. Exper. **21**(11), 1129–1164 (1991)

32. Walshaw, C.: A multilevel algorithm for force-directed graph drawing. In: Marks, J. (ed.) GD 2000. LNCS, vol. 1984, pp. 171–182. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-44541-2_17

33. Hagberg, A.A., Schult, D.A, Swart, P.J.: Exploring network structure, dynamics, and function using NetworkX. In: 7th Python in Science Conference (SciPy 2008), (SciPy), pp. 11–15 (2008)

Using Meta-Learning to Reduce the Effort of Training New Workpiece Geometries for Entanglement Detection in Bin-Picking Applications

Marius Moosmann[1](✉), Julian Bleifuß[1], Johannes Rosport[1], Felix Spenrath[1], Werner Kraus[1], Richard Bormann[1], and Marco F. Huber[2]

[1] Department Robot and Assistive Systems, Fraunhofer Institute for Manufacturing Engineering and Automation IPA, Nobelstraße 12, 70569 Stuttgart, Germany
`marius.moosmann@ipa.fraunhofer.de`

[2] Department Cyber Cognitive Intelligence (CCI), Fraunhofer IPA and Institute of Industrial Manufacturing and Management IFF, University of Stuttgart, Stuttgart, Germany

Abstract. In this paper, we introduce a scaling method for the training of neural networks used for entanglement detection in Bin-Picking. In the Bin-Picking process of complex-shaped and chaotically stored objects, entangled workpieces are a common source of problems. It has been shown that deep neural networks, which are trained using supervised learning, can be used to detect entangled workpieces. However, this strategy requires time-consuming data generation and an additional training process when adapting to previously unseen geometries. To solve this problem, we analyze and compare several Meta-Learning techniques like Reptile, MAML and TAMS for their feasibility as a scaling method for the entanglement detection. These methods search for a strongly generalized model for entanglement detection by learning from the training process of various workpieces with different geometries. Using this generalized model for entanglement detection as initialization helps to increase the learning success with only few training epochs and reduces the required amount of data and therefore the setup effort significantly.

Keywords: Machine Learning · Meta Learning · Bin Picking · Entanglement Detection · Scaling

1 Introduction

In the Bin-Picking process entangled workpieces are a common source of problems for incorrect handling. To increase the robustness of successful grasps in Bin-Picking, the application can be extended with an entanglement detection and furthermore with separation methods [1–5]. It has been shown, that such entanglement detection can be realized with the use of neural networks in a model-based approach [6]. However, when applied to new workpiece geometries, this supervised learning approach requires an expensive deal of effort.

The current state of the supervised learning approach of the entanglement detection [6] uses a deep convolutional neural network. The architecture is inspired by DenseNet [7] and is trained with grayscale depth maps of potentially entangled situations using the supervised learning approach.

The depth maps are generated in a simulation and later transferred to reality using Sim-to-Real-methods. To conduct the Sim-to-Real-Transfer we use Cycle-GAN as a domain adaptation method and several domain randomization parameters, for example Gaussian noise on the input images. As the approach is model-based, the simulation needs the geometric information of the workpiece. Each workpiece therefore requires its own specific entanglement detection data generation and training. In order to receive a high performance entanglement detector, the training of the neural network requires up to 20,000 depth maps as training inputs. The training process and the data generation amount to 46 h on a standard hardware. In summary, the current state holds potential to reduce the effort on adapting the entanglement detection to new workpieces.

Meta-Learning shows great success in accelerating the adaption of neural networks and creating strong classification models with only few data samples. For this reason, different Meta-Learning methods were investigated for their suitability to reduce the effort on training new workpiece geometries for the entanglement detection.

In summary, the main contributions of this paper are:

- a base dataset for Meta-Learning based entanglement detection
- the comparison of the Meta-Learning methods applied to the entanglement detection
- the validation of the practical feasibility to reduce the effort on adapting the entanglement detection on new workpiece geometries with Meta-Learning

2   Meta-Learning

Meta-Learning enables machine learning models to use experience gained from related tasks [8]. It transfers previously learned knowledge of the training process and enables a neural network to perform this task faster and better. Meta-Learning is a learning process on two levels. The general procedure depends on the current Meta-Learning method and will be explained in the course of this work. Meta-Learning is used to realise powerful classification models with only a small amount of training data. With this procedure, several Meta-Learning methods achieve a high performance in few-shot image classification [9–14] or object detection [15–17].

As Meta-Learning grows in interest, a variety of Meta-Learning methods exist, which can be divided into gradient-based and metric-based algorithms, among others [18]. For the entanglement detection we perform experiments with the gradient-based algorithms Reptile [10] and the more complex MAML [9]. MAML achieves great success in generating a task-agnostic network which can

adapt to new tasks in few gradient steps. Therefore MAML uses the second-order derivatives as meta-gradient. The Reptile algorithm simplifies the method and is able to successfully meta-learn with fewer classes that are sufficiently populated [19]. As metric-based algorithm we test TAMS [20], which is based on prototypical networks [11] and dedicated for medium-shot applications. Since the entanglement detection tends to follow the character of medium-shot classification with less available classes but sufficient shots, we chose Reptile and TAMS for the experiments in addition to MAML.

3 Meta-Learning Applied to the Entanglement Detection

This section presents a brief overview over the base dataset used for the Meta-Learning based entanglement detection. Furthermore, it introduces the different investigated Meta-Learning methods applied to the entanglement detection.

3.1 Base Dataset

Successful Meta-Learning requires a base dataset of source tasks closely related to the later target task. In this case, the target task is the classification between entangled and non-entangled workpieces of an unknown geometry. Therefore, the base dataset consists of 54 different workpieces with various geometries in total. Each workpiece provides the entanglement detections as classification task with 200 synthetic depth maps, half of them showing entangled workpieces.

To validate the Meta-Learning implementation, the Omniglot [21] dataset was used. This dataset consists of images from letters and is similar to the depth maps in the manner that both have only one channel. Even though the two image datasets differ in their structure, one could observe a benefit on Meta-Learning by pretraining the models with Omniglot. The Omniglot dataset was consulted to pretrain and to verify the functionality of the Meta-Learning method.

3.2 Implementation to the Entanglement Detection

The Meta-Learning based entanglement detection was applied as K-shot N-way classification, where K describes the quantity of the training data and N the number of classes distinguished.

The implementation of the Meta-Learning based entanglement detection is realised in such a way that for each meta-batch N/2 workpiece geometries are sampled randomly from the base dataset. Both gradient-based algorithms are based on a 5-shot 6-way classification due to experimental experience. Therefore one meta-batch represents the simultaneous entanglement detection of three different workpiece geometries. This scheme is sketched in Fig. 1. In case of TAMS a 5-shot 20-way classification is selected as best hyperparameter for the entanglement detection application. After each Meta-Learning epoch the K-shot N-way classification is repeated and evaluated with unknown geometries for testing the Meta-Learning model without updating it.
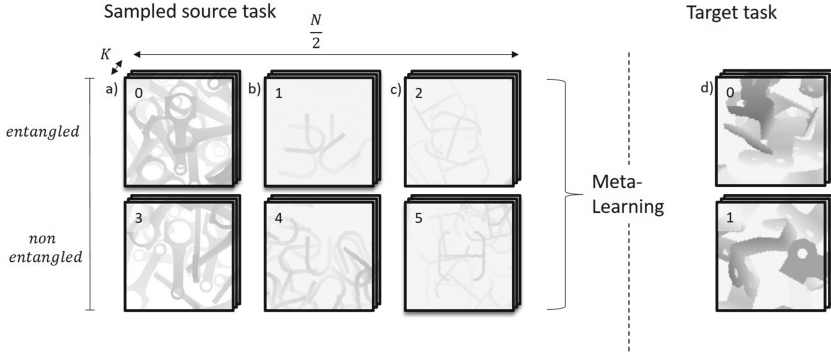
Fig. 1. Meta-Learning with depth maps of **a)** connecting rods, **b)** u-bolts, and **c)** double hooks, for faster adaptation on the entanglement detection of **d)** metal holder.

The Meta-Learning results in a strongly generalised network for the entanglement detection of multiple workpiece geometries. In order to finetune it to the unknown workpiece geometrie with transfer learning later, the classification layer of the generalised network is modified to a binary classifier.

3.3   Training of the Meta-Learning Methods

The training of the Meta-Learning applied to the entanglement detection is monitored using a subset of previously separated workpiece geometries from the base dataset. We use a split of 46 workpiece geometries for training and eight workpiece geometries for validating the adaptability. Figure 2 exemplarily shows the training plots from the Reptile and MAML Meta-Learning. In a training it happens that the validation accuracy is better than the training accuracy. We explain this behavior by the quality of the training and validation data. The training data was generated some time ago with an outdated physics simulation, while the validation data is from a revised version. In detail, data acquisition with a virtual depth image sensor and the physical interaction of the components in the bin have been improved through optimizations in simulation.

The Reptile Meta-Learning converges within 5,000 Meta-Learning epochs which takes about 4.5 h. The MAML Meta-Learning needs about 24 h for 8,000 epochs and then starts overfitting. This is due to the few source tasks in the base dataset. With a larger base dataset with hundreds of different workpiece geometries the performance of MAML is expected to improve. The TAMS-algorithm also suffers from the few classes and does not make any significant progress in the Meta-Learning. The structure of the base dataset with sufficient labeled data points per class but only few training tasks fits the Reptile algorithm best [19].
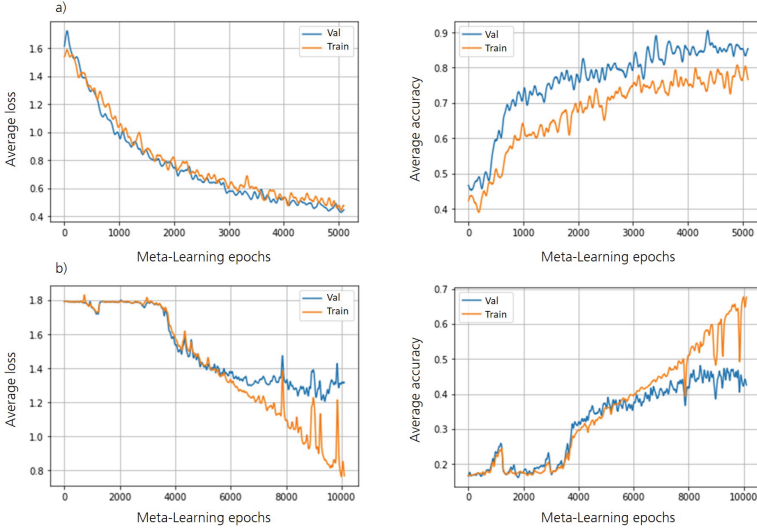
Fig. 2. Meta-Learning applied to the entanglement detection. **a)** Reptile, **b)** MAML

4   Results

### 4.1   Comparison of Applied Meta-Learning Methods

To compare the Meta-models generated by Reptile, MAML and TAMS, we use three new workpiece geometries not utilised in the Meta-Learning process yet. We also add a model with randomly initialized network parameters to the comparison which has to train the entanglement detection from scratch. We test the adaptation of the four models in dependency of the number of training data. While varying the amount of training data, the 2,500 depth maps for testing remain the same for each workpiece. We repeat each adaptation training eight times with different dropout-rates for regularisation and capture the best performance for each training data amount afterwards. Figure 3 shows the results for the three chosen workpiece geometries.

Comparing the workpiece geometries with each other, it is noticeable that the entanglement detection of the connecting rods is easier to learn than of the u-bolts, which results in higher test accuracies. The four models therefore do not differ much in the performance after the adaptation to the connecting rods. In case of the double hook and the u-bolt it can be seen that the Reptile model outperforms the other Meta-models and the random model in nearly every training data amount by far. The biggest benefit of the Meta-Learning can be observed in the adaptation to the double hooks with 2,500 depth maps.

4.2   Performance Validation of the Meta-trained Entanglement Detection on Unseen Workpiece Geometries

The model comparison leads to the choice of the Reptile algorithm as Meta-Learning to reduce the effort on training new workpiece geometries for the
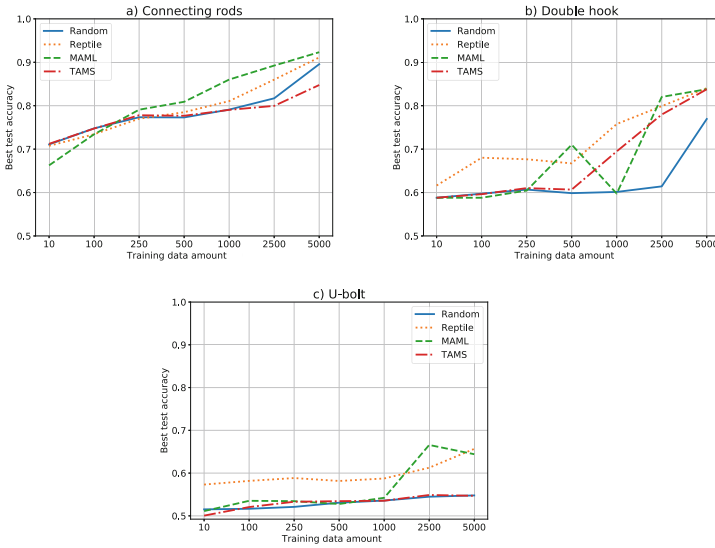
Fig. 3. Best models after 100 epochs of adaptation training to new workpiece geometries for **a)** connecting rods, **b)** double hooks, **c)** u-bolts

entanglement detection. This method is once more validated with the metal holder, shown in Fig. 1d, as workpiece which is interesting for industrial applications at a customer site. In doing so, the direct gain of Reptile as method to adapt to new workpieces with less effort, in later contexts abbreviated scaling method, is recorded.

Therefore the Reptile model and a model without prior knowledge from a Meta-Learning are compared in the adaptation with the same depth maps in a training with identical parameters. Figure 4 shows the progression of test accuracy and test loss during training with a training dataset consisting of 2,500 (green) data samples for the Reptile model and 2,500 (blue) and 5,000 (purple) data samples for the random model. The test dataset amounts to 5,000 equal depth maps.

One can observe that the Reptile model immediately starts adapting to the new workpiece and reaches a high classification performance in significantly less training epochs than the model without the Meta-Learning. If the training data is doubled to 5,000 instances, it is possible to reach a similar performance to the Meta-Learning, but with significantly larger number of epochs. In all cases, the loss of the adaptation training indicates the start of overfitting the model after it converged. The test-accuracy however remains stable and Reptile outperforms the current state of the entanglement detection by 100 epochs and 2,500 training samples.
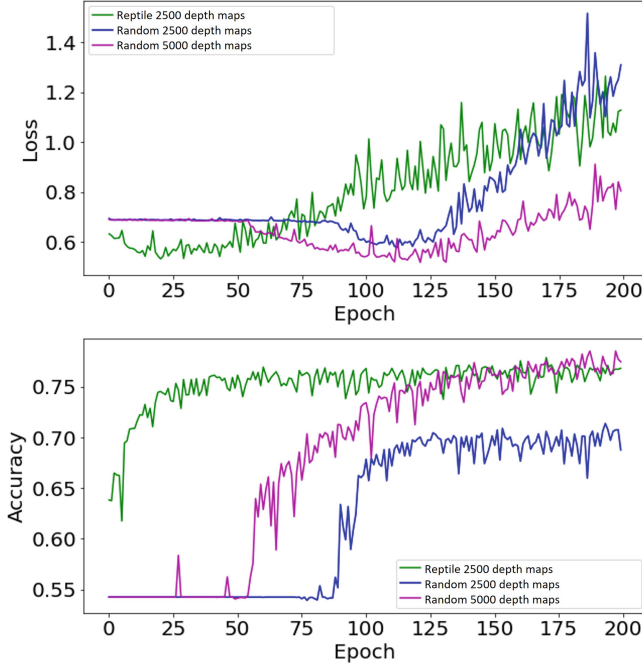
Fig. 4. Adaption to the new workpiece geometry with and without Meta-Learning

5 Summary and Outlook

In this paper we introduced a scaling method to reduce the effort of training new workpiece geometries for entanglement detection in Bin-Picking. We compared three different Meta-Learning methods from the current state of the art in their usability for the entanglement detection. The chosen scaling method based on the Reptile algorithm helps reducing the amount of training epochs and therefore the training time by about 80 percent points and halves the amount of training data, which also halves the simulation time.

The scaling method makes the entanglement detection feasible for faster responses to new workpiece geometries in the Bin-Picking application. However, to achieve an entanglement detection with a higher performance than with Meta-Learning, more training data can be used to train a specific machine learning model. In conclusion, the Meta-Learning method helps reacting quickly on customer requests and a more accurate entanglement detection model can be updated later.

To further improve the scaling method in future work, it is of interest how the Meta-Learning becomes stronger with the growing base dataset through new workpiece geometries.

References

1. Moosmann, M., et al.: Using deep neural networks to separate entangled workpieces in random bin picking. In: SCAP Stuttgart Conference on Automotive Production (2020)
2. Moosmann, M., et al.: Separating entangled workpieces in random bin picking using deep reinforcement learning. In: CIRP Conference on Manufacturing Systems (2021)
3. Matsumura, R., Domae, Y., Wan, W., Harada, K.: Learning based robotic bin-picking for potentially tangled objects. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (2019)
4. Leão, G., Costa, C.M., Sousa, A., Veiga, G.: Perception of entangled tubes for automated bin picking. In: Iberian Robotics Conference, pp. 619-631 (2019)
5. Leão, G., Costa, C.M., Sousa, A., Veiga, G.: Detecting and solving tube entanglement in bin picking operations. In: Applied Sciences (2020)
6. Moosmann, M., et al.: Increasing the robustness of random bin picking by avoiding grasps of entangled workpieces. In: CIRP Conference on Manufacturing Systems, Chicago (2020)
7. Huang, G., Liu, Z., Van Der Maaten., L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700-4708 (2017)
8. Hospedales, T., Antoniou, A., Micaelli, P., Storkey, A.: Meta-learning in neural networks: a survey. In: IEEE Transactions on Pattern Analysis and Machine Intelligence (2021)
9. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: International Conference on Machine Learning, pp. 1126-1135. PMLR (2017)
10. Nichol, A., Achiam, J., Schulman, J.: On first-order meta-learning algorithms. In: arXiv preprint arXiv:1803.02999 (2018)
11. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
12. Raghu, A., Raghu, M., Bengio, S., Vinyals, O.: Rapid learning or feature reuse? Towards understanding the effectiveness of MAML. In: International Conference on Learning Representations (2020)
13. Ravichandran, A., Bhotika, R., Soatto, S.: Few-shot learning with embedded class models and shot-free meta training. In: International Conference on Computer Vision (2019)
14. Sun, Q., Liu, Y., Chua, T., Schiele, B.: Meta-transfer learning for few-shot learning. In: Conference on Computer Vision and Pattern Recognition (2019)
15. Wu, X., Sahoo, D., Hoi, S.: Meta-RCNN: meta learning for few-shot object detection. In: ACM International Conference on Multimedia, pp. 1679-1687 (2020)
16. Kang, B., Liu, Z., Wang, X., Yu, F., Feng, J., Darrell, T.: Few-shot object detection via feature reweighting. In: IEEE International Conference on Computer Vision, pp. 8420-8429 (2019)

17. Wang, Y., Ramanan, D., Hebert, M.: Meta-learning to detect rare objects. In: IEEE International Conference on Computer Vision, pp. 9925-9934 (2019)
18. Chen, W. Y., et al.: A closer look at few-shot classification. In: International Conference on Learning Representations (2019)
19. Al-Shedivat, M., Li, L., Xing, E., Talwalkar, A.: On data efficiency of meta-learning. In: International Conference on Artificial Intelligence and Statistics, pp. 1369-1377. PMLR (2021)
20. Jiang, X., Ding, L., Havaei, M., Jesson, A., Matwin, S.: Task adaptive metric space for medium-shot medical image classification. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11764, pp. 147–155. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_17
21. Lake, B.M., Salakhutdinov, R., Tenenbaum, J.B.: Human-level concept learning through probabilistic program induction. Science **350**(6266), 1332–1338 (2015)
22. Zhu, J.-Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: IEEE International Conference on Computer Vision ICCV (2017)

Performance Comparison of Supervised and Reinforcement Learning Approaches for Separating Entanglements in a Bin-Picking Application

Marius Moosmann[1]([✉]), Manuel Kaiser[1], Johannes Rosport[1], Felix Spenrath[1], Werner Kraus[1], Richard Bormann[1], and Marco F. Huber[2]

[1] Department Robot and Assistive Systems, Fraunhofer Institute for Manufacturing Engineering and Automation IPA, Nobelstraße 12, 70569 Stuttgart, Germany
`marius.moosmann@ipa.fraunhofer.de`
[2] Department Cyber Cognitive Intelligence (CCI), Fraunhofer IPA and Institute of Industrial Manufacturing and Management IFF, University of Stuttgart, Stuttgart, Germany

Abstract. Machine Learning helps to separate entanglements in Bin-Picking Applications. The goal is to create a system that finds a path to separate an entanglement, starting from a single visual input. To realize such a system both supervised and reinforcement learning methods can be implemented. For both of these approaches we set up a motion model and the remaining properties of the real robot cell are implemented in a simulation scene. While the simulation scene can be used to create training data for the supervised learning approach, it is also the learning environment for the reinforcement learning model. Therefore, there are similar premises for comparing the two models. What needs to be investigated is which of the two methods separates the most entanglements and offers the least setup effort. The setup effort in general and the performance are examined for both approaches in simulation and also in real-world experiments. The reinforcement learning model outperforms both of the supervised learning models in the setup effort and the separation rate by over 15 percent points.

Keywords: Bin-Picking · Machine Learning · Reinforcement Learning · Supervised Learning · Entanglement Separation

1 Introduction

The ability to automatically pick chaotically stored workpieces from bins creates many new opportunities in productions. While some workpiece geometries can be picked in a robust manner, there are a lot of workpieces which are prone to entangle. Therefore an automated process that is supposed to consistently pick single workpieces needs to have the ability to detect and separate entanglements. Since the detection of entanglements has been handled in previous work [1,2], this paper mainly focuses on the separation of these entanglements. For this

purpose a new convolutional neural network (CNN) architecture for a supervised learning approach has been developed and tested alongside two already existing supervised learning [3] and reinforcement learning [4] approaches.

The main contributions of this paper are:

- the introduction of the updated supervised learning architecture
- the introduction of a threshold evaluation to deny gripping points for impossible separations
- real-world entanglement separation experiments
- comparison of supervised and reinforcement learning approaches

2   State of the Art

In motion-planning for bin picking applications multiple different approaches have emerged recently. Ellikide et al. [5] and Iversen et al. [6] prioritize finding a motion path, which avoids collisions with the environment. In the case of separating entangled harnesses Zhang et al. [7] uses a set of eight possible motion schemes with increasing complexity. Matsumura et al. present a model-free entanglement detection approach, but without separation strategies [8]. Leao et al. calculate the robot trajectory based on the size of the workpiece and move the robot on its x-y-plane [9]. Moosmann et al. [3] proposed a motion model in the shape of a hemisphere consisting of 25 points, which is based around the center of the entangled workpiece. The amount of hemisphere points have later been reduced to 17 points [4]. Each of the points has a specific translational and rotational offset that gets added to the original workpiece position. The workpiece is moved to the point, which has the highest probability to separate the entanglement. To calculate these probabilities a supervised learning [3] and a reinforcement [4] learning approach has been developed. Additionally two more hemispheres are created centering around the selected path point of the preceding hemisphere. Once all three path points have been passed the workpiece is lifted up. Figure 1 displays an example for a separation in the simulation environment.

| Starting position | Hemisphere 1 | Hemisphere 2 | Hemisphere 3 |
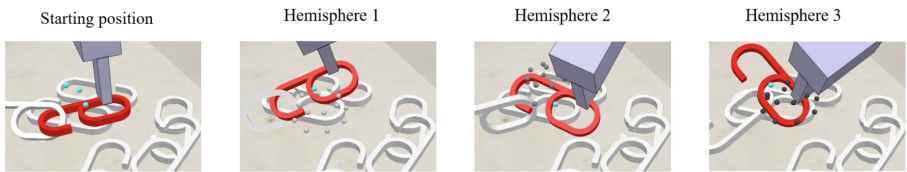
Fig. 1. Example for an entanglement from the simulation with the corresponding separation path generated by the Entanglement Separation Network

3 New Supervised Learning Entanglement Separation Method
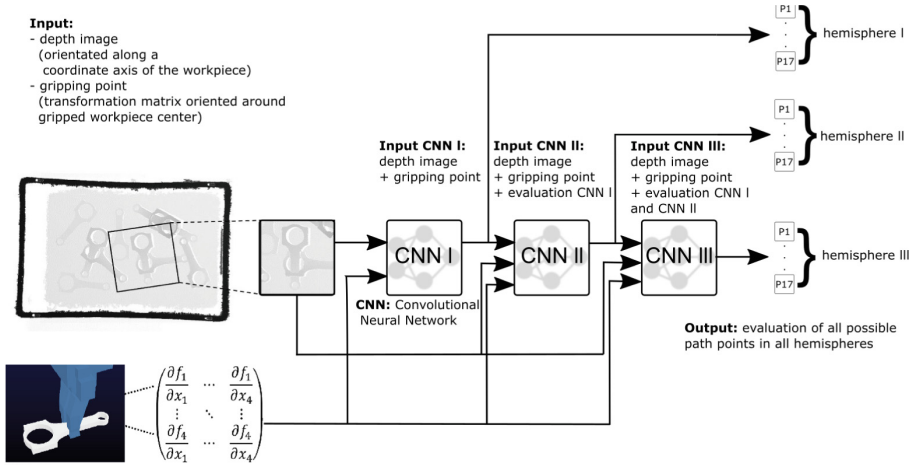
Fig. 2. Supervised Learning A: Serial connection of CNNs presented in [3].

In [3] a supervised learning method was presented, which uses a serial connection of three convolutional neural networks to predict the optimal trajectory for entanglement separation as shown in Fig. 2. This approach has been unified into a single network within this work in order to simplify the usage and reduce training time.

3.1 Data Generation

The training data is generated using the simulation environment CoppeliaSim[1]. In the simulation scene several objects such as the workpieces and bins in multiple sizes are integrated. All are based on a CAD-model with real-world proportions. A simulation cycle starts by filling the bin with a random amount of background workpieces, which vary between 0 and 20. After that a random entanglement is selected from of previously generated entangled workpiece poses and placed with a random x and y offset in the bin. To make sure the entanglement is still valid after being placed in the bin, the entangled workpiece is lifted up and it is checked how many workpieces are located above the bin. As soon as the conditions for a valid entanglement are met, a set of possible gripping points is checked in the simulation. For every gripping point that does not collide with the bin or the surrounding workpieces, a simulation cycle will start.

---

[1] https://coppeliarobotics.com/.

After a valid gripping point has been chosen, the path points of the first hemisphere are checked. A separation path is considered successful, if neither the gripper nor the workpiece collide with the bin and the entanglement has been separated. The second and third hemisphere is created for the best point of the preceding hemisphere. In the case of multiple or none successful separations, the path which caused the least movement for the surrounding workpieces is selected as the center for the next hemisphere. One simulation cycle is finished as soon as the 17 path points of each hemisphere have been checked. The separation motion model is presented in Fig. 3. On the left, the distribution of the 17 possible path points on the hemisphere, on the right, a possible trajectory with three hemispheres is shown.
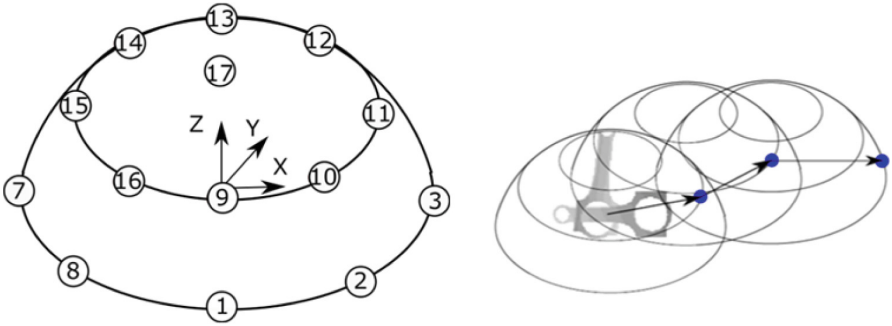
Fig. 3. Separation motion model - left: 17 possible path points of a hemisphere; right: possible trajectory with three hemispheres

The depth images, which are taken for every cycle have a size of 128×128 pixels. Before training the network with these images they are transformed using transfer learning methods as shown in [1]. This is necessary to minimize the differences between simulation and real-world data. For domain adaptation, we use CycleGAN [12] to generate real-looking depth maps from simulation. To get more realistic sensor data, we add different domain randomisation factors to the simulation depth map, for example Gaussian noise, translational and rotational offsets and brightness adaption [13,14]. We also use inpainting techniques [15].

3.2 Network Architecture

Since for the new architecture all three networks as presented in Fig. 2 have been merged into a single one, the information about the previously selected actions and the current hemisphere index needs to be transferred in a consistent manner. Therefore a matrix with 2×17 values is used as an additional input. The first column contains the value "one" at the index of the selected action in the first hemisphere and the value "zero" for the remaining indices. For the second column the value "one" is contained at the chosen action for the second

hemisphere. If the respective hemisphere has not been evaluated yet the vector contains 17 times the value "zero". The gripping point is represented by a 4×4 transformation matrix relative to the workpiece center. This means a simulation cycle as described in Sect. 3.1 creates three training labels with the same gripping point and depth image, but a varying previous action vector. The output of the network is the probability to solve the entanglement with each of the 17 hemisphere points. In Fig. 4 the complete architecture is summarized.
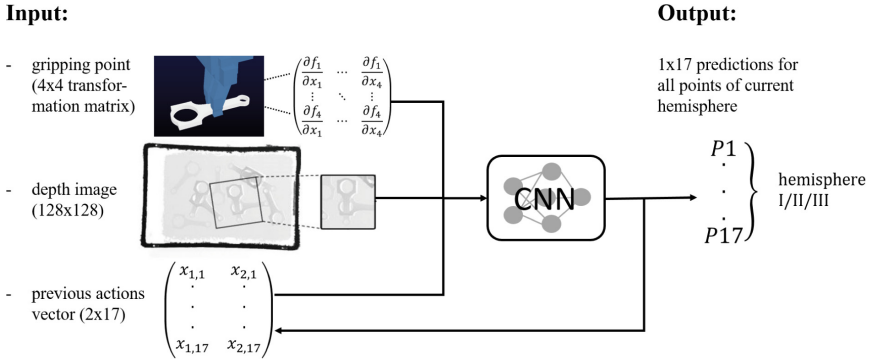
Fig. 4. Input and Output of the Entanglement Separation Network

For the separation task a DenseNet [16] architecture with 4 Denseblocks with depths of 6, 12, 24 and 12 layers is implemented.

3.3 Training

For all of the workpieces, which will be examined in this paper, around 20,000 data samples have been generated according to the procedure described in Sect. 3.1. The network has been trained using a Nvidia GeForce GTX 1080 Ti graphics card with a batch size of 128 for 80 epochs. To avoid overfitting, a dropout with a dropout rate of 35% has been used.

3.4 Threshold Evaluation

With the current motion model most entanglements have multiple different paths, which lead to separation. However some entanglements are impossible to separate. One reoccurring problem is the gripper blocking the path of the entangled workpiece as shown in Fig. 5 (a). Aside from that some entanglements are significantly easier to separate if another workpiece, which is part of the same entanglement, is gripped. To address these problems we implemented the ability to deny the gripping of a workpiece if the average value of the 17 predictions of the first hemisphere is below a threshold. To find the best threshold, the validation data of our training set has been examined for every workpiece on multiple thresholds between 0 and 0.1. The threshold with the highest accuracy has shown to be 0.022 as depicted in Fig. 5 (b).

<div align="center">(a)   (b)</div>
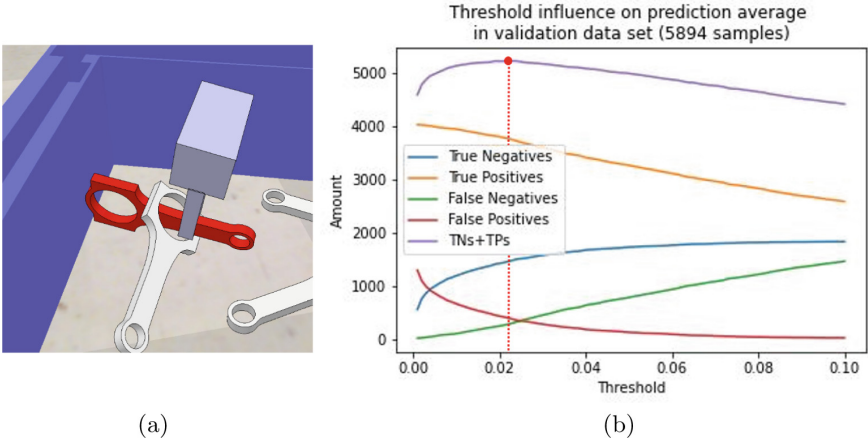
Fig. 5. (a) Example of an unsolvable gripping point (b) Threshold influence on prediction accuracy of the connecting rod with maximum at 0.022

4   Comparison

4.1   Setup Effort

For the supervised learning approach the amount of training data is a substantial factor for training success. Therefore it is necessary to generate a large amount of training data, which is the most time consuming part. To generate data from one simulation cycle as described in Sect. 3.1, a Lenovo Thinkpad with an Intel(R) Core(TM) i7-10750H CPU with 2.60 GHz processor and 16 GB RAM takes about 4.5 min on average. Accordingly generating 20,000 data samples would take 1,500 h. However since simulations can run on multiple processor cores simultaneously, the time to generate this data can be divided by the amount of cores on the respective system, reducing data generation time significantly. To keep the conditions for this comparison equal this aspect will be ignored. The training time of the updated supervised approach is 0.5 h lower than the previous approach.

In one episode of the reinforcement learning training a single separation path in the simulation environment is tried. Therefore the reinforcement learning approach takes up significantly less time per episode, compared to a cycle of the supervised learning data generation. Aside from that, the reinforcement learning training does not need any previously generated data and uses an Epsilon-Greedy strategy to explore the environment within the first 1,000 episodes. To achieve a sufficiently trained network about 30,000 episodes are necessary, which takes around 175 h. In Table 1 time consumption is summarized concluding that the reinforcement learning take up clearly less time.

Table 1. Time consumption comparison of Supervised and Reinforcement Learning approach, Supervised Learning A is referring to the new merged architecture and Supervised Learning B is referring to the old serial connection architecture

Approach	Required Sim Data for Training	Setup Time	Training Time	Total Time
Supervised Learning A	≈20,000	≈1,500 h	≈1 h	≈1,500 h
Supervised Learning B	≈20,000	≈1,500 h	≈1.5 h	≈1,500 h
Reinforcement Learning	0	0	≈175 h	≈175 h

4.2   Real-World Performance Comparison

To evaluate the performance of the networks, real-world experiments have been carried out. In order to see how the separation rate differs for a variety of workpiece geometries, u-bolts, connecting rods and hooks have been tried. The methods under consideration for the following comparison are the reinforcement learning approach introduced in [4], the supervised learning approach introduced in [3] and the new supervised approach introduced in this paper. For every combination of those workpieces and machine learning methods, the entanglement separation success rate for 200 workpieces has been determined as depicted in Table 2. All tests up to the last row do not involve the threshold, which has been introduced in Sect. 3.4. Comparing the results of the different supervised learning networks, the only minimal differences for all workpieces range between 0.5 and 1.5% points. However comparing the reinforcement learning network with the supervised learning approach the separation rates differ within a greater range. Here the u-bolt workpiece geometry shows the best separation rate with 98%, being 15.5% points better than the new supervised approach. For the hook and the connecting rod a slightly better separation rate can be achieved.

To evaluate the effectiveness of the threshold, another test series with the connecting rod has been carried out, in which predictions below the threshold were denied. This workpiece has been prioritised for the threshold evaluation, because it is more prone to impossible gripping points as demonstrated in Fig. 5 (a). The improvement of these test cases is shown in the last row of Table 2. Here an additional improvement between 4 and 5 percent points for both supervised learning approaches and the reinforcement learning approach is visible.

Table 2. Separation test comparison for all combinations of machine learning methods and workpieces (with 200 entanglements per combination)

Object	Approach	Entanglements separated	Separation Rate
U-bolt	Supervised Learning A	165	82.5%
	Supervised Learning B	162	81.0%
	Reinforcement Learning	196	98.0%
Hook	Supervised Learning A	92	46.0%
	Supervised Learning B	91	45.5%
	Reinforcement Learning	101	50.5%
Connecting Rod	Supervised Learning A	174	87.0%
	Supervised Learning B	176	88.0%
	Reinforcement Learning	176	88.0%
Connecting Rod	Supervised Learning A	183	91.5%
With	Supervised Learning B	184	92.0%
Threshold	Reinforcement Learning	186	93.0%

4.3   Integration into a Bin Picking System

The Bin-Picking Application, in which the separation strategies are integrated in, is able to detect and localize workpieces using a point cloud [10]. Furthermore to acquire a suitable gripping solution a heuristic search is used [11]. From the point cloud depth images are extracted and used as input for the entanglement detection. In case the workpiece is recognized as entangled a request for a separation path will be sent to the entanglement separation [1].

5   Conclusion and Future Work

In this paper it has been shown that the serial connection, consisting of three networks, which was previously implemented, can be reduced to a single network. This reduces training and load up time of the weights from the neural networks, without compromising in terms of performance. Additionally the real-world tests have shown that the reinforcement learning model achieves a 15% points higher separation rate, while having a lower setup effort. Furthermore, we introduced a threshold evaluation to deny gripping points, on the basis of which the separation of entanglements is impossible. This evaluation is also done in real-world experiments. In future work we will try to improve the reinforcement learning approach with any additional rotations and simplify the pipeline to teach in new workpiece geometries. Furthermore we will try to reduce training time with meta-learning methods.

References

1. Moosmann, M., et al.: Transfer learning for machine learning-based detection and separation of entanglements in bin-picking applications. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2022)
2. Moosmann, M., et al.: Increasing the robustness of random bin picking by avoiding grasps of entangled workpieces. In: CIRP Conference on Manufacturing Systems (CIRP CMS) (2020)
3. Moosmann, M., et al.: Using deep neural networks to separate entangled workpieces in random bin picking. In: Procedia CIRP (2020)
4. Moosmann, M., et al.: Separating entangled workpieces in random bin picking using deep reinforcement learning. In: CIRP Conference on Manufacturing Systems (CIRP CMS) (2021)
5. Ellekilde, L.P., Petersen, H.G.: Motion planning efficient trajectories for industrial bin-picking. In: IJRR, pp. 991-1004 (2013)
6. Iversen, T.F., Ellekilde, L.P.: Benchmarking motion planning algorithms for bin-picking applications. Ind. Robot. **44**(2), 189–197 (2017)
7. Zhang, X., Domae, Y., Wan, W., Harada, K.: Learning a sequential policy of efficient actions for tangled-prone parts in robotic bin picking. In: arXiv (2021)
8. Matsumura, R., Harada, K., Domae, Y., Wan, W.: Learning based robotic bin-picking for potentially tangled objects. In: IEEE/RSJ IROS (2019)
9. Leao, G., Costa, C.M., Sousa, A., Veiga, G.: Detecting and solving tube entanglement in bin picking operations. Appl. Sci. **10**, 3390 (2020)
10. Palzkill, M.: Heuristic method for object pose detection using point clouds for automated feeding systems (2014)
11. Spenrath, F., Spiller, A., Verl, A.: Gripping point determination and collision prevention in a bin-picking application. In: 7th German Conference on Robotics (ROBOTIK) (2012)
12. Zhu, J.-Y., Park, T., Isola, P., Efros, A. A.: Unpaired image-to- image translation using cycle-consistent adversarial networks. In: IEEE International Conference on Computer Vision (ICCV) (2017)
13. Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W., Abbeel, P.: Domain randomization for transferring deep neural networks from simulation to the real-world. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2017)
14. Borrego, J., Dehban, A., Figueiredo, R., Moreno, P., Bernardino, A., Santos-Victor, J.: Applying domain randomization to synthetic data for object category detection. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2018)

15. Telea, A.: An image inpainting technique based on the fast marching method. In: Journal of Graphics Tools (2004)
16. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K. Q.: Densely connected convolutional networks. In: IEEE CVPR (2017)

Machine Learning-Based Identification of Root Causes for Defective Units in Manufacturing Processes

Dominik Rotter[1(✉)], Florian Liebgott[1], Daniel Kessler[1], Annika Liebgott[2], and Bin Yang[2]

[1] Balluff GmbH, 73765 Neuhausen a.d.F., Germany
`dominik.rotter@balluff.de`
[2] Institute of Signal Processing and System Theory, University of Stuttgart, 70569 Stuttgart, Germany

Abstract. To achieve a high overall equipment effectiveness in a manufacturing process, reducing the number of defective units is crucial. It is therefore vital to identify the root causes of defects to be able to rectify them. However, the analysis of defective units can be a time-consuming and costly task.

By using machine learning, we can leverage data of the manufacturing process, like process states and different measurement values, to identify the root causes for the defects. We propose to use this data as features for the classification of a unit as defective or as belonging to a specific defect class. We can then identify the root causes for the defects by calculating the importance of the features.

In this paper, we compare our feature-based approach with deep learning methods based on the attention mechanism. The evaluation of our approach on data of a complex production process shows, that our approach clearly outperforms the deep learning methods. It also revealed the challenges in the collection of meaningful data

Keywords: root cause analysis · feature importance · attention mechanism

1 Introduction

In the age of Industry 4.0, the automation and digitization of manufacturing processes is playing an increasingly important role. With the help of sensors, a large number of process parameters are recorded in order to obtain as accurate a picture as possible of the current system status or to automate as many subprocesses as possible. In order to obtain profitable information from this amount of data to increase the efficiency of a manufacturing process, machine learning is a suitable tool. Here, models are trained with training data to automatically detect errors or plant conditions.

In today's manufacturing plants, machine learning is used for a wide variety of tasks. For example, machine learning is often found in quality control at the end of a manufacturing step [1,2]. With the help of images and previously trained models, defective parts can be automatically classified and sorted out. In this way, defective products can be identified at an early stage, sometimes even before the final inspection. Another area in which machine learning is already widely used in production is condition and process monitoring [3–5]. Condition monitoring or predictive maintenance can be used to monitor machines and systems and detect critical conditions before a machine breaks down. As a result, production downtimes can be avoided, maintenance costs reduced, and maintenance work as a whole made more plannable.

One topic where machine learning has hardly been used so far is the reduction of scrap by directly identifying and eliminating the causes of defects in a manufacturing process [6,7]. Although a lot of data is often collected during the production of a part and large datasets are available, these are rarely used for optimization. Often, the experience of the person responsible for the machine is relied on alone or no optimization is carried out at all [8].

Especially in the interpretability of models, which is crucial for the identification of root causes, great progress has been made in recent years and new approaches have been explored. For example, in the image domain, many new methods exist, such as GradCAM [9], guided backpropagation [10] and integrated gradients [11], to gain insight into how models make decisions. Similarly, for tabular data, which is the most common in production, there are some new deep learning approaches based on the attention mechanism [12,13] in addition to the traditional algorithms for computing feature importance [14,15].

To the best of our knowledge, this paper is the first to compare different machine learning based approaches for the detection of root causes in production processes. For this purpose, various approaches for the determination of feature importance are first presented and then compared with each other on the basis of a real dataset from automated sensor production.

2 Theoretical Background

There are different approaches to determine the root causes of errors using machine learning methods. In this section, classical methods and methods based on the attention mechanism are explained in more detail.

2.1 Classical Methods

The classical methods use a classifier to determine which features are responsible for a high classification result. The worse the classifier is without a feature, the more important this feature is for the classification. Since the goal of the classifier is to learn the usually complex relationship between features and result, it can be concluded that the features important for the classifier in error classification are also causes of errors in the real production plant. The classical methods

work completely independent of the type of classifier or model. The two most important representatives are the permutation feature importance and the drop column feature importance.

Permutation feature importance is a concept introduced by [14] to determine the feature importance of trained models. It measures the decreasing accuracy $A_{k,j}$ of a model when permuting a feature j compared to the reference accuracy A_{ref} of the original dataset D_{test}. Permuting a feature breaks its relationship with the result. Accordingly, if the accuracy of a model $A_{k,j}$ drops sharply after permutation, it is a feature that is important to the model. Insignificant features, on the other hand, are not relevant for the decision process of the model. If these features are permuted, the model ignores them because its decision is based on the important features. Consequently, the feature importance F_j can be calculated as

$$F_j = A_{\mathrm{ref}} - \frac{1}{K} \sum_{k=1}^{K} A_{k,j} \tag{1}$$

where K is a defined number of permutations to reduce the influence of random swapping [15,16].

An obvious criterion for a feature importance corresponding to reality is a well-predicting model. Only if the model has learned to pay attention to the right features, a good feature importance can be derived.

Highly correlated features are problematic for the permutation feature importance algorithm [17]. Despite the permutation of a feature, the model still has access to it via the correlated feature. This leads to reduced feature importance values for both features, when in fact their importance could be much higher. Therefore, care should be taken with permutation feature importance to remove highly correlated features.

The drop column feature importance method goes one step further than the permutation feature importance method. The importance of a feature is not determined by decreasing accuracy when that feature is swapped, but by reducing accuracy when the feature is ignored. A disadvantage of this is the high computational effort required. The model must be relearned for each feature. In addition, the algorithm exhibits the same shortcomings in regards to highly correlated features as the permutation feature importance method.

2.2   Method Based on Attention Mechanism

The attention mechanism was introduced in 2015 by [18] as an improvement to neural machine translation systems in natural language processing (NLP). Since then, this revolutionary concept has been transforming the application of deep learning in a wide variety of fields [19]. Not only in NLP, but also in computer vision applications [20], multiple instance learning [21], and language processing [22], numerous improvements have been achieved due to the attention mechanism.
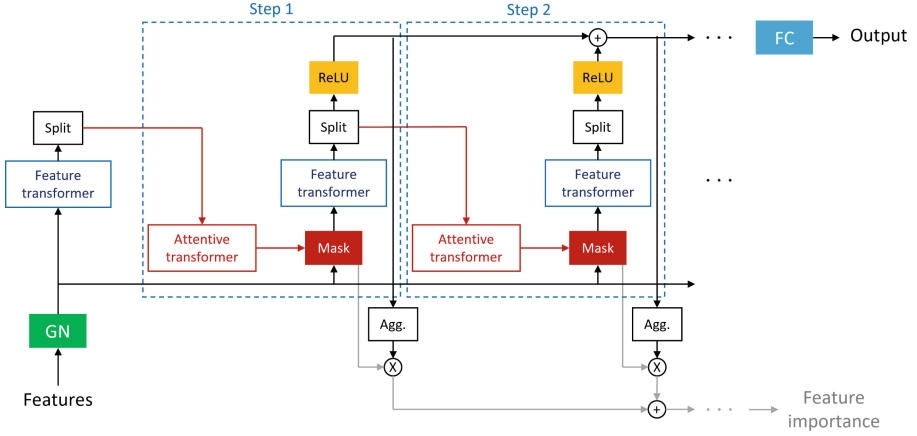
Fig. 1. TabNet architecture consisting of the encoder for classification. This is composed of several decision steps, each having a feature transformer, an attentive transformer, and a learnable selection mask. Based on [12]

Machine attention can be compared well with the cognitive attention of humans. The human brain is perfectly adapted to focus its attention only on the most important features despite an enormous flood of information. The attention mechanism implements the same principle for deep learning architectures, in that the neural network learns to pay attention only to the most important inputs [19]. This allows for more efficient learning and, by evaluating attention values, interpretability of DNN architectures.

TabNet is a neural network developed by Google LLC researchers for interpretable learning of tabular data based on the attention mechanism [12]. It uses a sequential attention mechanism to prioritize/select the correct features at each decision step.

Figure 1 shows the architecture of TabNet. It consists of N_{Steps} sequential decision steps. That is, the ith decision step operates on the results of the $(i-1)$th decision step to select which features to use. This is done using learnable masks $\mathbf{M[i]} \in \mathbb{R}^{B \times N_F}$. Each decision step is given the same N_F-dimensional features $\mathbf{F} \in \mathbb{R}^{B \times N_F}$, where B is the batch size and N_F is the number of features. The data are preprocessed only by group normalization (GN) [23].

The essential features are selected by multiplying the learnable mask $\mathbf{M[i]}$ by the feature matrix $\mathbf{F}$. Since the values of the mask are $\mathbf{M[i]} \in [0, 1]$, this corresponds to a soft feature selection. The masks are computed by the attentive transformer from the processed features of the previous step $\mathbf{A[i-1]}$ through

$$\mathbf{M[i]} = \text{sparsemax}\left(\mathbf{P[i-1]} \cdot h_i\left(\mathbf{A\,[i-1]}\right)\right). \tag{2}$$

The sparsemax activation function [24] is an extension of the softmax activation function to obtain sparse results, i.e. in this case weakly populated masks. As a result, the mask selects only the most essential features. h_i corresponds

to a trainable function consisting of a fully-connected (FC) layer and a batch normalization. $\mathbf{P}[\mathbf{i}]$ is a scaling term that tells how many times a given feature has been used so far.

The features selected by the mask are then processed by the feature transformer f_i, whose structure is described in detail in [12]. The feature transformer is followed by a split of the resulting matrix into the decision step output $\mathbf{D}[\mathbf{i}]$ and the information propagation $\mathbf{A}[\mathbf{i}]$ to

$$[\mathbf{D}[\mathbf{i}], \mathbf{A}[\mathbf{i}]] = f_i(\mathbf{M}[\mathbf{i}] \cdot \mathbf{F}). \tag{3}$$

Finally, the output $\hat{y}$ of the TabNet model is composed of a sum of the individual decision step outputs $\mathbf{D}_{\text{out}} = \sum_{i=1}^{N_{\text{steps}}} \text{ReLU}(\mathbf{D}[\mathbf{i}])$, a final FC layer, and a softmax function:

$$\hat{y} = \text{softmax}(\mathbf{W}_{\text{final}}\mathbf{D}_{\text{out}}). \tag{4}$$

The importance of a feature is included in the learned feature selection masks $\mathbf{M}[\mathbf{i}]$. The feature importance can therefore be calculated from the weighted sum of the individual masks:

$$\mathbf{M}_{\text{agg}-\mathbf{b,j}} = \frac{\sum_{i=1}^{N_{\text{steps}}} \eta_b[i]\mathbf{M_{b,j}}[\mathbf{i}]}{\sum_{j=1}^{N_F} \sum_{i=1}^{N_{\text{steps}}} \eta_b[i]\mathbf{M_{b,j}}[\mathbf{i}]}, \tag{5}$$

where the denominator is added for normalization. To factor in the importance or decision power of an instance b of a decision step, the coefficient $\eta_b[i]$ is used, which is calculated from the decision step outputs $\mathbf{D}[\mathbf{i}]$.

3   Data

A real dataset from a production plant of a sensor element is available for the investigation and comparison of the presented methods for the determination of failure causes. The production line consists of several production steps, including bonding, coil winding and welding processes, and contains several functional tests as well as quality controls. During the development of the production line, a special focus was placed on the generation, storage and display of process data. Therefore, during the production of a part, data such as process states, pressures, positions, distances, measurement results and other values are measured and stored. In addition, error codes are automatically generated in the event of errors. This enables separate investigation of the individual defect cases. In total, a dataset with 726,004 product instances is available.

During the analysis and pre-processing of the data, a frequent absence of parameters in the event of a defect was noticed. If a defect is detected in an early production step, the affected part is sorted out, further production steps are omitted and many parameters are not measured or stored. This is problematic for determining the root causes of defects, since the missing parameters cannot be used to determine the cause. Replacing them with specific values or distributions

Table 1. Size of the remaining datasets in the most frequent error cases.

Error code	Number of features	Number of instances
A	26	403,442
B	12	478,009
C	34	413,769
D	37	369,144
E	35	415,782
F	13	548,760

would cause the classifier to classify according to these values or distributions, breaking the actual relationship between the features and the output value. A calculation of the feature importance would be falsified by this. Table 1 shows how many features and instances the datasets still contain after sorting out for the most common error cases.

A calculation of the feature importance of all faulty instances against the correct instances as well as a multi-class classification is only possible either with the smallest denominator of the common remaining features or few instances. Otherwise, some error cases, due to the reasons described above, could be easily classified and thus falsify the result. We therefore chose to investigate the error cases individually. In the following, we present the results for error code A (position coil wire before welding out of order) as an example.

4   Results

For the determination of fault causes by means of machine learning, two criteria are particularly decisive. First, how well can the model associated with the method classify the fault case and second, what is the quality of the feature importance ranking determined with the respective method. In addition to the methods already presented, the techniques ReliefF and mutual information, which are known from the related topic of feature selection, as well as the Gini index automatically computed in the random forest classifier are evaluated. They serve as a basis for comparison of the different methods.

In a first step, the classification results of the different approaches are examined. For this purpose, the mean and standard deviation of the metrics balanced accuracy A_{bal}, sensitivity of error detection TPR, and specificity TNR are calculated for a random forest (RF) classifier, a support vector machine (SVM), and the TabNet model with a 5-fold cross-validation. The respective results are shown in Table 2. The comparison clearly shows that the random forest classifier gives the best classification results. For this reason, the feature importance rankings obtained using this classifier are expected to be the most reliable. The random forest classifier is followed by the TabNet model and the support vector machine in terms of their suitability.

One of the most important investigations is to determine the quality of the feature rankings. For this purpose, random forest classifiers are trained with a

Table 2. Comparison of classification results for error code A achieved with the support vector machine (SVM), random forest (RF) and TabNet models. The results are determined with a 5-fold cross validation. The mean value plus/minus the standard deviation for the test dataset is given in each case.

| Model	A_{bal} in %	TPR in %	TNR in %
SVM	75.34 ± 0.28	72.21 ± 0.55	78.47 ± 0.43
RF	79.16 ± 0.53	77.85 ± 1.19	80.46 ± 0.19
TabNet	76.33 ± 0.41	76.22 ± 1.72	76.45 ± 2.42

dataset f_j containing the j features with the highest importance and their classification result is computed. Random forest classifiers are suitable for this task because their training process is fast enough to obtain acceptable computation times and, as stated in the previous paragraph, they have a good classification result. The features are added one by one according to the rankings computed by the methods until finally the whole dataset F is used for training. The quality $Q(f_j)$ is defined as

$$Q(f_j) = \frac{A_{\mathrm{bal}}(f_j)}{A_{\mathrm{bal}}(F)}. \tag{6}$$

Figure 2 shows the result for the error code A. The first observation is that the drop column feature importance method produces the best results. Possible error causes of this error case can therefore best be determined with this method. It is closely followed by the calculated ranking of permutation feature importance, which is plausible given the strong similarities between the two methods. In
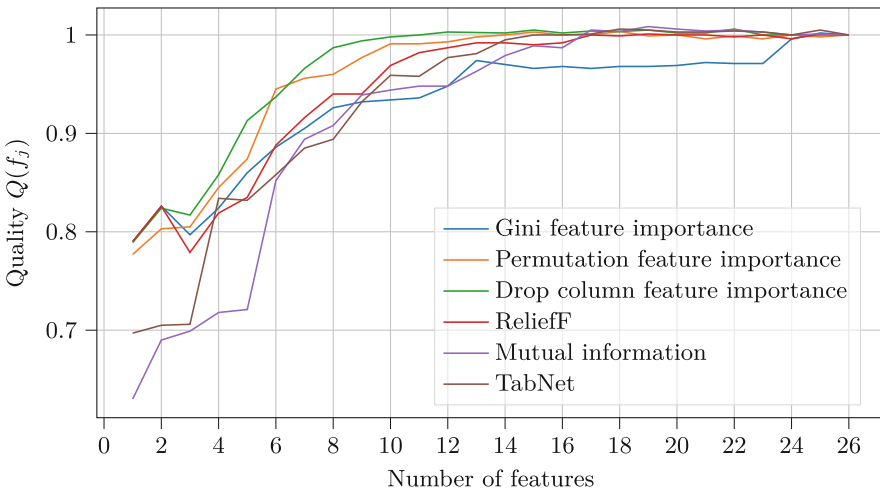
Fig. 2. Quality of feature importance rankings for the different methods for error code A.
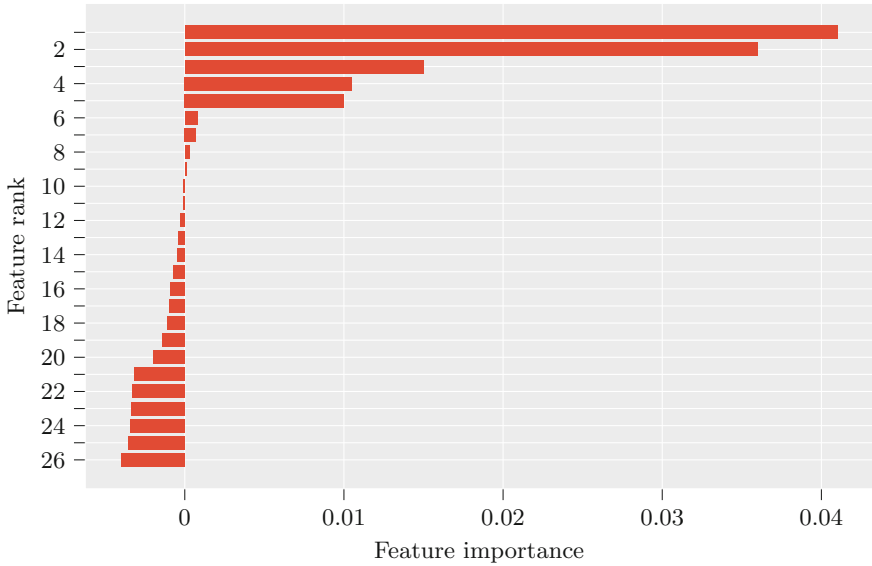
Fig. 3. Drop column feature importance for the error code A.

contrast, the Gini feature importance algorithm performs worse. Although the approach has placed the correct features in the top ranks, important features seem to be located only at the end of its ranking, as can be seen from the late jumps in feature ranking quality with many features already used. The opposite is the case for the rankings determined with the TabNet model and the mutual information. These have incorrect features calculated as most important, having lower quality $Q(f_j)$ first and having jumps in the curve after some features j have been added.

5 Discussion

The comparison of the classifiers and the quality of the feature importance rankings show, that the best results for the studied error case can be obtained with a random forest classifier in combination with the drop column feature importance. This is consistent with the results of the other error cases. For the error code A this corresponds to the result shown in Fig. 3.

The two most important features *position of the left/right wire* here correspond to the error description of the error code *wrong position of the coil wire before welding*. Since the error is determined directly from these features, this is a plausible result. The next features are of particular importance. They allow conclusions to be drawn about possible causes or consequences of the defect. In the following, these features will be considered individually:

1. **Wire spacing of the coil outgoing wires**: This features is measured in an early process step, the winding of the coil. It is therefore recorded before the

parts are assembled and before the defect occurs. A large wire gap near the tolerance limit can cause the wire to be out of position before welding after the parts are joined. This feature can thus be a cause of the defect.

2. **Lead number of transfer track**: The lead number of the transfer track is in close contact with the next placed feature, the workpiece number. In the production line, there are two different transfer tracks and several workpiece carriers. Due to an even number of workpiece carriers, these are always assigned to one transfer track. If a workpiece carrier produces worse parts, there will be an uneven distribution of defects among the transfer tracks.

3. **Workpiece carrier number**: As mentioned in the previous feature, there are several workpiece carriers. It is quite possible that one or more workpiece carriers negatively influence the parts.

4. **Maximum contact force of the coil into the ferrite**: This feature is detected when the coil is pressed into the ferrite. An effect on the position of the coil wire is therefore possible.

5. **Contact path of the coil into the ferrite**: The contact path of the coil is closely related to the maximum contact force of the coil into the ferrite. The same conclusions apply.

6. **Distance of the left edge to the core opening**: The distances of the edges to the core opening are measured in the same step as the position of the coil wire and serve as a reference for it. A direct connection is thus given.

All the above features are in direct contact with the position of the coil wire. It is therefore quite possible that they are causes of the error case.

6   Conclusion

In this paper, we compared feature-based approaches with a deep learning method based on the attention mechanism to determine root causes in an automated sensor manufacturing process. The results achieved on a dataset containing real production data show that the best results can be obtained with a random forest classifier in combination with the drop column feature importance method. This made it possible to determine potential root causes, as illustrated with one exemplary error case. All identified root causes are directly related to the error case and can now be optimized in further investigations.

This paper also revealed the challenges in the collection of meaningful data. Especially in fault cases, many parameters are not written to the database. As a result, many parameters in some error cases contain no or only a few measured values and are therefore not usable for the evaluation of this error case.

References

1. Bonomi, N., et al.: Smart quality control powered by machine learning algorithms. In: 2021 IEEE CASE, pp. 764–770 (2021)
2. Mohammadi, P., Wang, Z.J.: Machine learning for quality prediction in abrasion-resistant material manufacturing process. In: 2016 IEEE CCECE, pp. 1–4 (2016)

3. Gawde, S.S., Borkar, S.: Condition monitoring using image processing. In: 2017 ICCMC, pp. 1083–1086 (2017)
4. K, M.K., Kannadaguli, P.: IoT based CNC machine condition monitoring system using machine learning techniques. In: 2020 IEEE CSNT, pp. 61–65 (2020)
5. Zhao, X., Li, M., Xu, J., Song, G.: Multi-class semi-supervised learning in machine condition monitoring. In: 2009 ICIECS, pp. 1–4 (2009)
6. Abdelrahman, O., Keikhosrokiani, P.: Assembly line anomaly detection and root cause analysis using machine learning. IEEE Access **8**, 189,661–189,672 (2020)
7. Carletti, M., Masiero, C., Beghi, A., Susto, G.A.: Explainable machine learning in industry 4.0: evaluating feature importance in anomaly detection to enable root cause analysis. In: 2019 IEEE SMC, pp. 21–26 (2019)
8. Root cause analysis of failures and quality deviations in manufacturing using machine learning. Procedia CIRP **72**, 1057–1062 (2018)
9. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-CAM: visual explanations from deep networks via gradient-based localization. In: 2017 IEEE ICCV, pp. 618–626 (2017)
10. Springenberg, J.T., Dosovitskiy, A., Brox, T., Riedmiller, M.: Striving for simplicity: the all convolutional net. arXiv preprint arXiv:1412.6806 (2014)
11. Sundararajan, M., Taly, A., Yan, Q.: Axiomatic attribution for deep networks. In: Proceedings of the 34th ICML, vol. 70, pp. 3319–3328. PMLR (2017)
12. Arik, S.O., Pfister, T.: Tabnet: attentive interpretable tabular learning. arXiv preprint arXiv:1908.07442 (2019)
13. Skrlj, B., Dzeroski, S., Lavrac, N., Petkovic, M.: Feature importance estimation with self-attention networks. In: ECAI 2020, vol. 325, pp. 1491–1498 (2020)
14. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
15. Molnar, C.: Interpretable Machine Learning (2022). https://christophm.github.io/interpretable-ml-book/. Accessed 14 May 2022
16. scikit-learn developers: 4.2. Permutation feature importance (2022). https://scikit-learn.org/stable/modules/permutation_importance.html. Accessed 16 May 2022
17. Hooker, G., Mentch, L.: Please stop permuting features: an explanation and alternatives. arXiv preprint arXiv:1905.03151 (2019)
18. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 (2014)
19. Hore, P., Chatterjee, S.: A Comprehensive Guide to Attention Mechanism in Deep Learning for Everyone (2019). https://www.analyticsvidhya.com/blog/2019/11/comprehensive-guide-attention-mechanism-deep-learning/. Accessed 14 May 2022
20. Xu, K., et al.: Show, attend and tell: neural image caption generation with visual attention. In: Proceedings of the 32nd ICML, vol. 37, pp. 2048–2057 (2015)
21. Ilse, M., Tomczak, J.M., Welling, M.: Attention-based deep multiple instance learning. arXiv preprint arXiv:1802.04712 (2018)
22. Chan, W., Jaitly, N., Le, Q.V., Vinyals, O.: Listen, attend and spell: a neural network for large vocabulary conversational speech recognition. In: ICASSP (2016)
23. Wu, Y., He, K.: Group normalization. arXiv preprint arXiv:1803.08494 (2018)
24. Martins, A.F.T., Astudillo, R.F.: From softmax to sparsemax: a sparse model of attention and multi-label classification. In: Proceedings of the 33rd ICML, pp. 1614-1623 (2016)

Enabling Single-Sensor Simultaneous Condition Monitoring of Several Vibration-Emitting Machine Parts Using Neural Networks

Andreas Seitz[1(✉)], Florian Liebgott[1], Dominik Rotter[1], Daniel Kessler[1], and Hans-Peter Beise[2]

[1] Balluff GmbH, 73765 Neuhausen auf den Fildern, Germany
andreas.seitz@balluff.de
[2] Department of Computer Science, Trier University of Applied Sciences, 54293 Trier, Germany

Abstract. In production environments, monitoring the vibration of a machine or parts thereof can yield important information about the condition of the machine. The most common recommendation for vibration-based condition monitoring is to place a vibration sensor on each part of interest. These vibration sensors usually output preprocessed data, for example the root mean square value of a predefined time window.

We propose to use machine learning to simultaneously monitor several vibration-emitting machine parts using only one single vibration sensor. Due to the superposition of multiple vibrations, this is not feasible using preprocessed sensor data. Our approach therefore consists of a one-dimensional convolutional neural network, which uses the raw vibration signal as input to classify the status of the monitored machine parts.

As a first proof of concept, we monitored the status of three different motors. Using our approach, we were able to detect, which motors were running at a given time with high accuracy. We were able to significantly improve the classification results by transforming the raw data into the frequency domain. Our results are promising and show, that monitoring several vibration-emitting machine parts at the same time using only one vibration sensor is feasible.

Keywords: condition monitoring · vibration sensing · convolutional neural network

1 Introduction

Condition monitoring enables an efficient and failure-free operation of production plants, machines and manufacturing processes. One way to get information about the condition of a machine or a machine part is to monitor its vibration. For a successful monitoring, it is commonly advised to place a vibration sensor on each part that is to be monitored. By using sensors, which output preprocessed

data, for example the root mean square value or the peak-to-peak value in a given time interval, simple processes can be monitored by comparing the sensor output to predefined thresholds [1,2].

By placing the sensor directly on the part of interest, it is often assumed that the sensor values are mainly reflecting the vibrations emitted by this part and that vibrations emitted by other parts hardly influence the sensor values.

For complex processes and machines, using thresholds and assuming that the sensor values depend only on the machine part the sensor is attached to, may not be feasible. In these cases, using machine learning to extract the desired information from the sensor values or the values of several sensors, can yield very good results [3,4].

Instead of using machine learning to improve the condition monitoring of a single machine part, we use it to simultaneously monitor several machine parts using only a single vibration sensor. As preprocessed sensor data may not contain enough information to extract the condition of each machine part, we use the raw vibration data. Nguyen et. al [5] proposed to treat monitoring several machine parts using only a single vibration sensor as a blind source separation problem. However, their approach detects the overall condition of a machine and is not able to classify the individual condition of each machine part. For the classification of the condition of individual parts, the use of a support vector machine has been proposed in [6] for agro-industrial machinery and in [7] for centrifugal pumps. In both cases, the authors calculated statistical properties of the time domain signal and the corresponding frequency domain signal and used them as features. These features included mean, standard deviation, skewness, kurtosis, crest factor and root mean square value amongst others.

We propose to use a convolutional neural network (CNN) to classify the condition of several machine parts using only a single vibration sensor. This approach eliminates the need for explicit feature extraction and selection. Furthermore, the CNN may find features more suitable for the task than statistical properties of the time or frequency domain signal and thus yield better results. To the best of our knowledge, simultaneous condition monitoring using CNNs has not been studied before.

2   Theoretical Background and Methods

2.1   Convolutional Neural Network

Deep learning is a part of machine learning and offers the possibility to learn complex relationships in data using artificial neural networks (ANNs) [8]. ANNs are inspired by the brain of humans and animals and are based on the mathematical concept of artificial neurons [9]. Artificial neurons are arranged in interconnected representation layers to filter information from the data layer by layer [10].

Convolutional neural networks (CNNs) were introduced in [11] as a form of ANNs which are in particular used for image or speech recognition tasks as well as time-series tasks. CNNs use convolution layers to extract local features. A convolutional layer can have multiple convolution matrices, also called kernels

or filters, which each create a feature map. The feature maps get stacked up and passed on to the next layer. To reduce shifts and distortions, a convolution layer is usually followed by a pooling layer, which performs a local averaging or subsampling. This results in a resolution reduction. Because convolutional operations are linear, a non-linear activation function must be applied to the output [8].

## 2.2   Discrete Fourier Transform

For periodic time signals, it is usually beneficial to investigate the spectrum of the signal. The transformation of a discrete-time signal into the discrete frequency domain using the discrete Fourier transform (DFT) is therefore a commonly used preprocessing step [12]. A DFT maps N finite discrete samples of a signal x_n onto N complex finite spectral values X_k with the index $0 \leq k \leq N - 1$ and is defined as

$$X_k = \sum_{n=0}^{N-1} x_n \, \mathrm{e}^{-\,\mathrm{i}\,2\pi nk/N} \quad . \tag{1}$$

2.3   Methods

In all of our experiments, we applied automated hyperparameter optimization using a Bayesian optimization algorithm to find the optimal parameters for the models. Figure 1 displays the structure and the hyperparameter space for the hyperparameter optimization. Our architecture contains one or more convolutional (conv) blocks, which consist of a convolutional layer, an average pooling or a max pooling layer, followed by a batch normalization and a ReLU activation layer. The optimization algorithm can choose from one up to five conv blocks with a default number of three. It can also choose the number of filters and size of the kernels in the convolution layer. For the number of filters, the selectable options are from four filters up to 64 with a step size of four and a default of eight. The kernel size options are from three up to a maximum of 59 with a step size of eight and a default size of 27. Because we chose a maximum kernel size of 59 for the hyperparameter optimization, we had to limit the number of conv blocks to five. After the conv block, the algorithm can choose between a global average or a global max pooling layer. Next, it has the option to add a dropout layer with a minimal dropout rate of 0.2 up to a maximum of 0.5 with a step size of 0.05 and a default of 0.25. At the end, there is a flatten layer followed by a dense layer with a ReLU activation and finally a dense layer with a softmax activation function.

   It would have been possible to add more layers and increase the hyperparameter space, but to reduce the number of trials, we chose a few basic parameters based on the current state of research, preliminary assessment and experience. As optimizer we use Adam, which was proposed in [13]. Furthermore, we chose categorical cross-entropy because of the one-hot encoding. The optimization metric
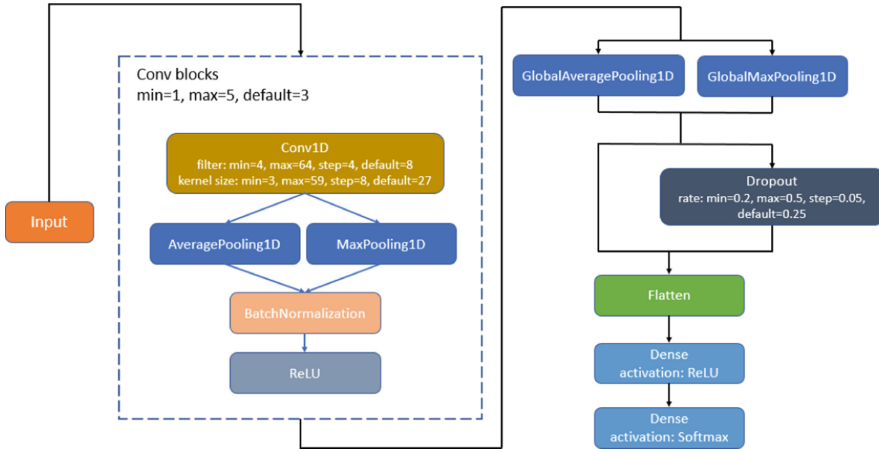
Fig. 1. The architecture we used for the hyperparameter search. The convolution block is highlighted using a dashed line.

is the classification accuracy over the test data set. As batch size we chose 32 as recommended in [14]. The output layer is, as mentioned above, a fully connected layer with eight neurons and a softmax activation function.

We use raw and transformed data to train the CNN model and measure the performance of each by comparing the correct classification rates. The transformed data is generated by taking the absolute value of the discrete Fourier transform of the raw data. Every run consists of three sections. First, we use hyperparameter optimization to find an architecture, which fits the data best. Next, we create ten models based on this architecture and train them separately. In the last step, we calculate a classification average. For every run, we make sure, that the data gets imported anew and that it gets distributed randomly into the training, test and validation set to avoid a favorable distribution. Finally we calculate the mean test accuracy and the standard deviation of the test accuracy over the ten loops for each run.

3   Experimental Setup

To create training data, we used a hardware setup with motors and a vibration sensor, that are installed on a perforated plate. We used one DC motor and two servo motors. This setup emulates a machine that was retrofitted with a single vibration sensor. The positions of the motors were chosen randomly. For the vibration sensor, we chose a specific position. In our scenario, this is possible, because the sensor was installed after the initial setup of the machine. Every motor has its own characteristic curve and therefore a specific vibration pattern. In this case, the DC motor is running continuously with the same frequency. The servo motors switch between movement and standstill with a different moving speed and idle time. The vibration sensor measures acceleration on three axes

Table 1. Mapping of motor combinations to classes.

Motors	Class
No motor is running (idle)	0
Servo 1	1
Servo 2	2
DC 1	3
Servo 1 - Servo 2	4
Servo 1 - DC 1	5
Servo 2 - DC 1	6
Servo 1 - Servo 2 - DC1	7

with a sampling frequency of 6.4 kHz. It provides raw vibration data without preprocessing.

We used software developed in-house to assign characteristics to the motors, control the start and stop time, as well as the start of the recording of the vibration with a predefined duration. The usage of three motors results in eight possible combinations. These combinations, as well as the class labels we assigned to the combinations, are listed in Table 1. For each class, 20 recordings were made.

The data is normalized with the L2 norm. Preliminary investigations showed, that an independent distribution normalization achieved better results than for example a min-max scaler or standardization.

We used 70% of the recordings for the training. The remaining 30% were split equally between validation and test data. Ahead of the split, the data was shuffled randomly without a seed to be sure that the distribution of the data differed from the runs before. This ensured, that an initial advantageous distribution of the data is avoided.

From the recordings, we created windows with 1600 data points, which equals a quarter of a second. We chose this time interval to reduce the inference time to a minimum with still enough information to reach a satisfying classification rate. To consider the transition between the windows, we chose a shift of 200 data points. Overall, we had 160 recordings, which resulted in 16000 windows for the training, validation and test sets. The split into training, validation and test data is made before creating the windows to ensure that all windows of a recording belonged to the same data set.
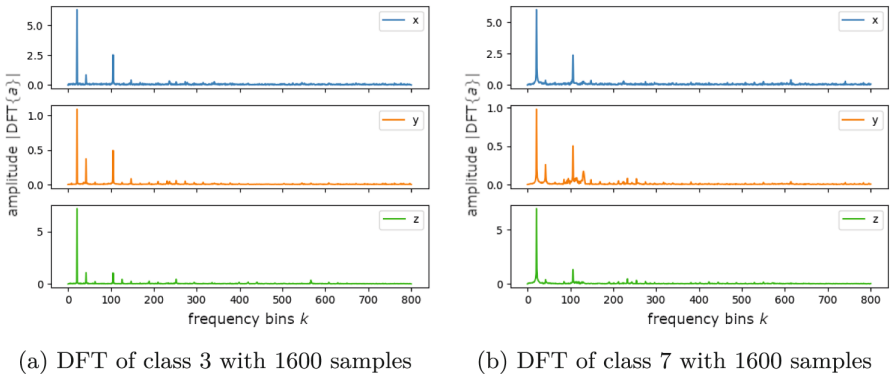
4   Results

4.1   Data Transformation

Figure 2 shows an example of raw data of the classes 3 and 7. In Fig. 2a, only the DC motor is running. In Fig. 2b, the DC motor and both servo motors are

(a) Raw data of class 3
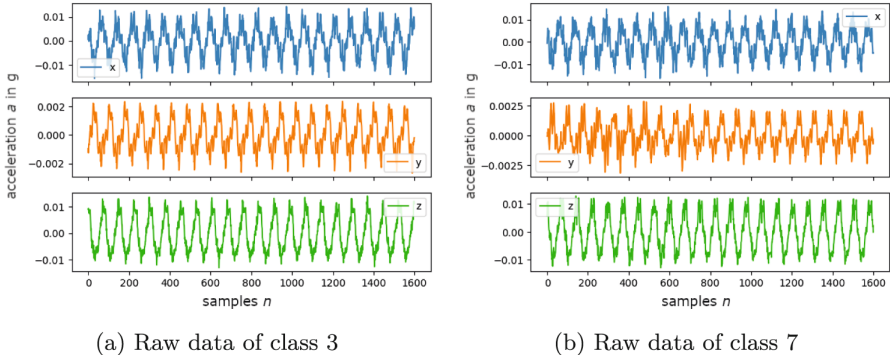
(b) Raw data of class 7

Fig. 2. Raw data of the classes 3 and 7, given as acceleration in g over 1600 data points

(a) DFT of class 3 with 1600 samples

(b) DFT of class 7 with 1600 samples

Fig. 3. Amplitude of the DFT with a length of 1600 for the classes 3 and 7.

in use. Both plots look very similar due to the strong vibrations emitted by the DC motor. In Fig. 2b, a small distortion is visible on the y axis between data point 100 and 800. This can be caused by the servo motors but could also have an external source of origin as well.

The absolute value of the Fourier transform of the time signals in Fig. 2 is shown in Fig. 3. The transformation into the frequency domain was applied over the whole time axis (1600 samples). As it was the case with the raw data, the dominance of the DC motor is visible. The high amplitudes are located in the low frequency area of the spectrum. Additionally, Fig. 3b displays spectral components between the frequency bins 80 and 120 as well as between 200 and 300. The amplitudes of these frequencies are small compared to the ones caused by the DC motor.

Considering our goal to use this classification algorithm on a sensor with limited resources, we decided to reduce the window size and thus the number of input values to 1200, 800 and 400 samples, respectively, by cropping before applying the DFT.
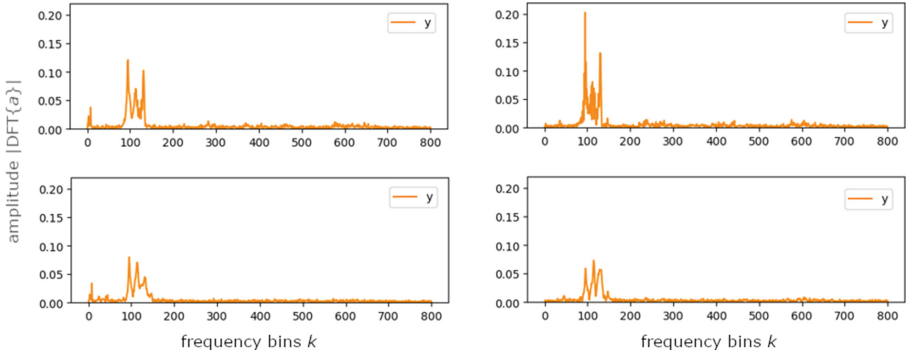
**Fig. 4.** Examples for the spectra occuring in class 2

In Fig. 4, four examples of the spectrum of class 2 are shown. This reveals significant differences in the magnitudes of the amplitudes. Although the plots have a similar pattern, the maximum magnitude of the plot in the top right is more than double compared to the bottom two plots. Also, there is a spike around frequency bin 5 visible in the plots on the left, which is missing in the plots on the right.

4.2 Training and Evaluation

The hyperparameter search process indicates that a decrease of input values exacerbates the discovery of architectures. Especially for the number of convolutional layers, we observed that the smaller the input size, the more the results varied between runs. We decided to choose one architecture for all input values except the raw data set because preliminary investigations showed, that the architecture found for 1600 samples reached high accuracies for the other input sizes as well. The architecture can be seen in Fig. 5. With five conv blocks it uses the maximum amount. In conv block one and three, 64 filters, that is the maximum possible number, are used where conv block four and five use the minimal number of four filters. Conv block two runs with the default eight filters. With a kernel size of three, only conv block one uses a smaller kernel than the default size of 27. The rest uses a kernel size of 35, 59, 59 and 35, respectively, where the maximal capacity of 59 is reached in the third and fourth conv block. Additionally, except conv block two, which uses average pooling, all conv blocks use max pooling. After the conv blocks, a global average pooling layer is used.

The classification results achieved with this architecture are listed in Table 2. The experiment with raw data to classify the motor status in row one reached the lowest mean classification accuracy value with 86.20%. Additionally, it has the highest standard deviation of all trials with 6.11%. The data shows, that the highest mean classification accuracy of 94.07% is achieved, when a Fourier transform with 1600 samples is used before training the model. When reducing the number of samples to 1200, 800 and 400, it shows a steady decline in correct

186 A. Seitz et al.
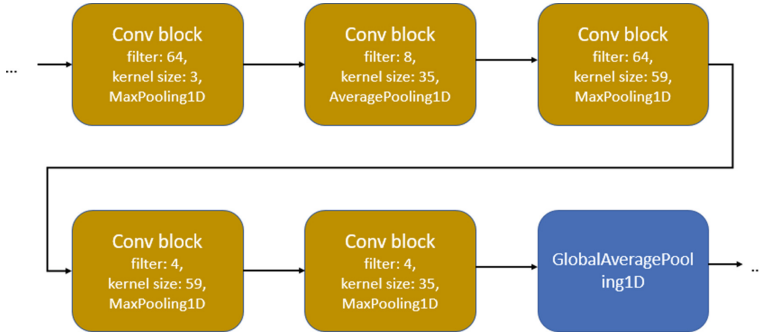
Fig. 5. Architecture found by the optimization algorithm

Table 2. Mean accuracy and standard deviation of the test accuracy for ten runs each with raw data and Fourier transformed data with a DFT length of 1600, 1200, 800 und 400 samples.

No.	Data	N	mean test accuracy	std
1	raw	–	86.20%	6.11%
2	DFT	1600	94.07%	2.41%
3	DFT	1200	92.98%	2.16%
4	DFT	800	90.13%	3.61%
5	DFT	400	87.33%	3.01%

classification accuracy with 92.98%, 90.13% and 87.33%, respectively. Although the lowest standard deviation is reached with 1200 samples, it comes with a reduced correct classification rate, which is 1.09% points lower compared to the mean accuracy with 1600 samples. The results also show, that even with a DFT length of 400 samples, a 1.13% points higher correct classification accuracy and a 3.1% points lower standard deviation is reached than with no transformation beforehand.

4.3 Discussion

The results show that it is feasible to simultaneously monitor several vibration-emitting machine parts using only one single vibration sensor. However, the plots in Fig. 4 suggest, that thresholds cannot be used to classify the motors, which supports our decision to use neural networks instead.

Although the classification accuracy of the CNN model, when using raw data, reached a satisfying level with an average of 86.20%, it also had a standard deviation of 6.11%. As shown in Fig. 2, this is likely due to the servo motors being masked by the DC motor, which dominates the shape of the time signal. As motors exhibit different frequencies, we transformed the signal into the frequency domain to achieve better separability. By using a DFT length of 1600 we

were able to achieve the highest classification accuracy with 94.07%. The steady decline of the accuracy when reducing the DFT length to 1200, 800 and 400 indicates that due to the shorter window sizes, the window may lie completely in the pause between two movements of a servo motor. The example thus contains no vibrations from the servo motor despite being labeled as such.

The results also show, that the model's greatest issues lie with distinguishing, if only the servo motor from class 2 or both servos are running. This false classification error from class 2 and 4 as well as class 6 and 7 represents 58.9% of all classification errors of the trial runs with 1600 samples.

Furthermore, we determined a correlation tendency with the number of hyperparameters and conv blocks with the number of input data when using hyperparameter optimization. Whereas the decrease of the correct classification rate with less data seems plausible, the average number of hyperparameters rose. As aforementioned, the optimization algorithm was not able to find a clear tendency for the number of conv blocks with a decreasing number of input data, which resulted in a lower average number of conv blocks. With a greater number of input data like 1200 or 1600 samples, the number of conv blocks tended towards the maximum number of five.

Our use case was limited to monitoring three vibration emitting machine parts. As this results in eight possible combinations, we were able to use multiclass classification. However, this approach is not feasible when more machine parts are in use. Therefore, we plan to introduce multi-task classification in the next step to achieve a scalable classification approach.

5   Conclusion

Our study shows, that it is feasible to use a single sensor to monitor several vibration-emitting machine parts simultaneously. We were able to reach an average correct classification rate of 94.07% by using a CNN model and transforming the data into the frequency domain beforehand. We also showed, that using raw vibration data with a CNN model can produce uncertain results because of its data dependency. Although it reaches a high accuracy of 86.20%, it still can't match the results reached by the CNN models in combination with a Fourier transform. The reduction of the input size proved not to be sensible due to considerably lower accuracy rates. However, with 1600 samples as input, the usage of the Fourier transform proved to be an effective preprocessing step to reach higher classification accuracies with CNN models. The results provide evidence, that a classification of superposed vibration-emitting machine parts with a CNN model in combination with a Fourier transformation is possible. Nonetheless, reducing the number of classification errors caused by similar vibration patterns, like servos, is an issue for future research to explore.

References

1. Ansari, S., Baig, R.: A PC-based vibration analyzer for condition monitoring of process machinery. IEEE Trans. Instrum. Meas. **47**(2), 378–383 (1998). https://doi.org/10.1109/19.744177

2. Jablonski, A., Barszcz, T., Bielecka, M., Breuhaus, P.: Modeling of probability distribution functions for automatic threshold calculation in condition monitoring systems. Measurement **46**(1), 727–738 (2013). https://doi.org/10.1016/j.measurement.2012.09.011

3. Paya, B., Esat, I., Badi, M.: Artificial neural network based fault diagnostics of rotating machinery using wavelet transforms as a preprocessor. Mech. Syst. Signal Process. **11**(5), 751–765 (1997). https://doi.org/10.1006/mssp.1997.0090

4. Su, H., Chong, K.T.: Induction machine condition monitoring using neural network modeling. IEEE Trans. Ind. Electron. **54**(1), 241–249 (2007). https://doi.org/10.1109/tie.2006.888786

5. Nguyen, V., Rutten, C., Golinval, J.C.: Fault diagnosis in industrial systems based on blind source separation techniques using one single vibration sensor. Shock Vibr. **19**, 795–801 (2012). https://doi.org/10.3233/SAV-2012-0688

6. Ruiz-Gonzalez, R., Gomez-Gil, J., Gomez-Gil, F.J., Martínez-Martínez, V.: An SVM-based classifier for estimating the state of various rotating components in agro-industrial machinery with a vibration signal acquired from a single point on the machine chassis. Sensors **14**(11), 20713–20735 (2014). https://doi.org/10.3390/s141120713

7. Kumar, A., Kumar, R.: Time-frequency analysis and support vector machine in automatic detection of defect from vibration signal of centrifugal pump. Measurement **108**, 119–133 (2017). https://doi.org/10.1016/j.measurement.2017.04.041

8. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MITP Verlags GmbH, Das umfassende Handbuch (2018)

9. Kruse, R., Borgelt, C., Braune, C., Klawonn, F., Moewes, C., Steinbrecher, M.: Computational Intelligence. Springer, Wiesbaden (2015). https://doi.org/10.1007/978-3-658-10904-2

10. Chollet, F.: Deep Learning mit Python und Keras. MITP Verlags GmbH, Frechen (2018)

11. LeCun, Y., Bengio, Y.: Convolutional networks for images, speech, and time-series. In: Handbook of Brain Theory and Neural Networks (1995)

12. Neubauer, A.: DFT - Diskrete Fourier-Transformation. Vieweg+Teubner Verlag (2012). https://doi.org/10.1007/978-3-8348-1997-0

13. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization (2014)

14. Bengio, Y.: Practical recommendations for gradient-based training of deep architectures. arXiv preprint arXiv:1206.5533 (2012)

Unlocking the Potential of Digital Twins

Development of the Virtual Commissioning Model to a Hybrid Predictive Maintenance System

Fabian J. Pohlkötter[1,2]([✉]) [iD], Dominik Straubinger[1], Alexander M. Kuhn[1], Christian Imgrund[1], and William Tekouo[1]

[1] BMW AG, Petuelring 130, 80809 Munich, Germany
`Fabian.Pohlkoetter@bmw.de`
[2] TU Dresden, Institute of Mechatronic Engineering, Helmholtzstr. 7, 01069 Dresden, Germany

Abstract. Increasing competitive pressure is confronting the automotive industry with major challenges. As a result, conventional reactive maintenance is being transformed into predictive maintenance. In this context, wearing and aging effects no longer lead to plant failure since they are predicted at an earlier stage based on comprehensive data analysis.

Furthermore, the evolution towards Smart Factory has given rise to virtual commissioning in the planning phase of production plants. In this process, a Hardware-in-the-Loop (HiL) system combines the real controls (e.g., PLC) and a virtual model of the plant. These HiL systems are used to simulate commissioning activities in advance, thus saving time and money during actual commissioning. The resulting complex virtual models are not further used in the series production.

This paper builds upon virtual commissioning models to develop a Digital Twin, which provides inputs for predictive maintenance. The resulting approach is a methodology for building a hybrid predictive maintenance system. A hybrid prediction model combines the advantages of data-driven and physical models. Data-driven models analyse and predict wearing patterns based on real machine data. Physical models are used to reproduce the behaviour of a system. From the simulation of the hybrid model, additional insights for the predictions can be derived.

The conceptual methodology for a hybrid predictive maintenance system is validated by the successful implementation in a bottleneck process of the electric engine production for an automotive manufacturer. Ultimately, an outlook on further possible applications of the hybrid model is presented.

Keywords: Predictive Maintenance · Digital Twin · Data-Driven Production · Virtual Commissioning · Data Science

1 Introduction

The increasing competitive pressure combined with global crises is a tremendous challenge for manufacturing companies. The transition from the internal combustion engine to electric mobility, combined with governments' environmental protection targets, poses

major hurdles for the automotive industry in particular [1]. It is therefore important to optimize existing processes in order to produce as efficiently and cost-effectively as possible while maintaining quality, supply capability and punctuality. Maintenance plays a major role in this context, as it accounts for an average of 9% of manufacturing costs [1, p. 1]. Approximately 50% of these costs can be attributed to the repair of equipment during shutdowns.

Predictive maintenance makes it possible to detect upcoming failures in time and thus perform maintenance proactively, rather than just reactively [2, 3]. Advances in connectivity between machines are making this digital transformation possible. Nevertheless, predictive strategies, are only used by about 15% of the manufacturing companies [4, p. 5]. In addition, suitable methods for data evaluation and forecasting are not yet sufficiently established [4, p. 2f.].

The transformation towards the digital factory developed the virtual commissioning (VC). A complex interconnected model of the planned system is generated in order to shorten the real commissioning time and thereby save costs. However, the resulting model is currently not further used after the real commissioning of the plant [5, p. 9]. Continued use of the model would relativize the investment in its creation.

This paper investigates if and how the model from the virtual commissioning can be used in later stages. Furthermore, it is described how the model has to be improved in order to build a hybrid predictive maintenance system. The combination of these two elements of Industry 4.0 and Digital Factory represents a significant scientific contribution to automation in production.

2 State of the Art

In the following, the fundamentals of the life cycle of production plants will be covered, as well as the digitalisation concepts relevant for this paper.

2.1 Life Cycle of Production Plants

The life cycle of production plants can be divided into three phases. The planning and realization phase, the operating phase, and the redistribution. In the following, the first two of these phases will be discussed in more detail since the development of the virtual commissioning model is situated between them.

Planning and Realization Phase
In the beginning, sequences and processes are designed and, if necessary, simulated according to the customer's requirements. Subsequently, mechanical, electrical, and pneumatic systems are developed and designed. The plant is then manufactured based on the construction plans. After the individual components have been completed, the final step is assembly and commissioning at the customer's site [6, p. 11f.]

The elimination of software errors takes up to 90% of the entire commissioning. The software is created after the design phase, i.e., in parallel to the manufacturing and assembly phase. Therefore, there is usually not enough time for extensive function tests of the

control software before commissioning. To meet this challenge, virtual commissioning has been established in the digitalisation of planning [7, p. 19].

Operating Phase
After the takeover of the production system by the customer, the ramp-up phase follows. During this stabilization process, technical defects and failures occur that were not detected during commissioning. One of the main reasons for this is the control software, due to the short test cycles. After the ramp-up has been completed, series production begins. In this phase, the system must be maintained in case of failures [6, p. 4].

2.2 Digitalisation of the Planning, Realization and Operation Phases

Dynamic and growing customer requirements demand more product variants with shorter time-to-market and reduced product life cycles. OEMs have to demonstrate flexibility and transformability due to the transformation towards electromobility. In the following, digitalisation and Industry 4.0 instruments are presented to meet these requirements [8, p. 103].

Virtual Commissioning (VC)
In virtual commissioning, a large part of the commissioning activities is simulated in advance. Based on the construction plans, a virtual model of the plant is created, with which processes and material flows can be modelled and thereby the control software can be tested. On average, the software deployment time is reduced by 75%. The quality of the control software increases from 37% to 84% [9]. Furthermore, critical and potentially dangerous situations can be simulated safely and without damage. By simulating complex interlinked processes, not only individual processes, the entire system behaviour can be defined and optimized in an early stage [10, p. 273f.].

Maintenance Strategies
In general, there are five different maintenance strategies. Reactive maintenance corrects faults only after they have occurred. Preventive maintenance avoids recurring faults through periodic servicing. The next strategy is monitoring the condition of a wearing part and maintenance is carried out when certain limits are reached. Predictive maintenance analyses patterns and trends of past failures and predicts the time of failure, including a confidence interval, based on current process data. The prescriptive strategy goes beyond and defines concrete solutions for what needs to happen at the predicted time of failure in order to maintain the plant [2].

A recent literature review shows that there are two basic approaches for building predictive maintenance strategies [2]. First, data-driven approaches that extract patterns and trends from process data of a plant and generate a prediction of the failure using machine learning methods [11]. On the other hand, physically driven approaches offer the possibility to generate prediction models based on the underlying physics [12]. A hybrid concept uses a combination of the two approaches above [13, 14].

Herein lies the novelty of the following concept. Building on existing research, this paper investigates how the virtual commissioning model can be developed into a digital twin to create a hybrid predictive maintenance system.

3   Concept, Industrial Use Case and Challenges

In order to understand how the potential of digital twins can be maximized, this chapter presents a method to identify beneficial use cases and further develop the VC model into a digital twin. Finally, the concept is implemented in an industrial use case.

3.1   Use Cases for Digital Twins

The developed concept (Fig. 1) suggests a way to further expand the virtual commissioning model into a digital twin. This is used to build a hybrid predictive maintenance system. The advantage lies in the fact that the virtual model from the planning phase has a practical use in the operating phase of production plants.

According to [15] a virtual model can be developed into a digital twin by including an automatic data exchange between real and virtual machine. In case of a semi-automatic information exchange this is called a digital shadow.

VC = Virtual Commissioning | PLC = Programmable Logic Controller | HiL = Hardware-in-the-Loop
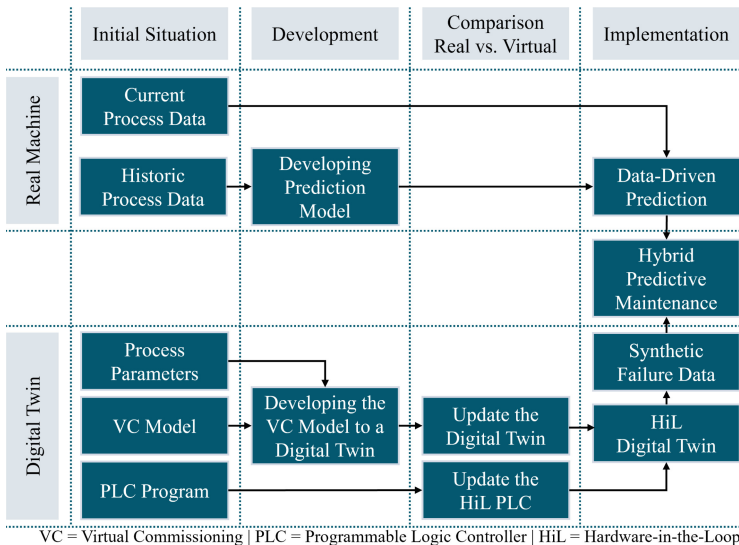
Fig. 1. Architecture for developing a hybrid predictive maintenance system

However, a predictive maintenance system should not be set up during the ramp-up phase, instead it should be built when the production plant is in stable series production. Further necessary information for setting up a hybrid predictive maintenance system is provided by experience of similar wear patterns, as well as literature. Predictive maintenance is particularly well suited for predicting wear patterns of mechanical components.

This paper presents the questions that can identify an appropriate use case for a given machine as shown in Fig. 2. By diligently answering those questions, a use case for predictive maintenance based on digital systems can be identified and specified.
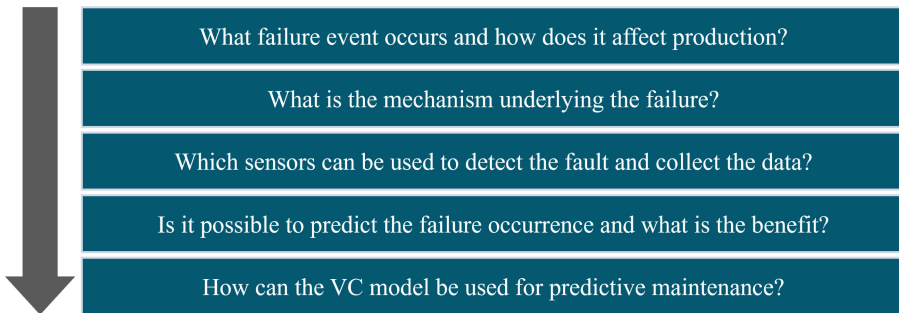
After the basic requirements for creating a hybrid predictive maintenance system have been fulfilled, it is examined whether the benefit exceeds the cost. This turns a simple use case into a business case. Exemplary benefits of predictive maintenance are listed below [16, p. 5]:

- Reduction of plant downtime and therefore increase of the OEE.
- Optimal utilization of the operating lifetime of wearing parts.
- Reduction of maintenance costs.
- Assurance of the quality of products.

As soon as the economic viability has been proven, the mechanism underlying the wear is inspected. First, it is assessed if the existing sensor technology is capable of detecting the fault or if additional external sensors are required. Then it must be checked whether the data integration is sufficient to process the sensor information and to record process data. After a positive assessment, data collection can be continued.

For data acquisition, it is important that the data can be recorded not only from the real machine, but also from the extended virtual commissioning model via the PLC of the HiL setup - the Digital Twin. By combining both real process data and virtually generated failure data, a hybrid predictive maintenance system is realized. The advantage of using the digital twin is that more failure data can be generated. In practice, there often is not enough data available that contains the failure that is supposed to be predicted. This is where the synthetically generated failure data acts as an enabler by providing essential training data with as many failures as necessary.

What failure event occurs and how does it affect production?

What is the mechanism underlying the failure?

Which sensors can be used to detect the fault and collect the data?

Is it possible to predict the failure occurrence and what is the benefit?

How can the VC model be used for predictive maintenance?

Fig. 2. Identification of a use case for a hybrid predictive maintenance system

3.2   Development of the Virtual Commissioning Model Towards a Digital Twin

After the use case has been identified, the task is to extend the virtual commissioning model to generate synthetic fault data. There are several approaches for this.

The first approach reproduces the wear in the form of mathematical equations. In addition to a deep understanding of the process, this also requires a high computing power, since the formulas are mostly multidimensional and are iteratively optimized.

The second approach is material removal simulation. Using the example of a cutting process, the cutting edge gradually reduces in sharpness with each cut until the cutting

edges are so worn that the material can no longer be cut through. This is challenging, because such a detailed material simulation would require a tool that supports such a feature, thus limiting the choice. Secondly, and more importantly, the computational performance for simulating the degradation of the blades is substantial and as such would hinder a model or digital twin beyond feasibility and practicability.

The third and selected approach is synthetic failure data generation (Fig. 3) based on real patterns and trends. Real fault data are recorded and imported into the simulation program. These data are then randomized in the simulation within a statistically significant range. Hence, new synthetic fault data are generated, based on real patterns and trends. This could not be done if subprocesses were simulated separately. The connections between subsystems in the digital twin are necessary to understand the machine's behaviour and to correctly predict failures.

In this paper, an architecture model is developed that provides the essential steps for building a hybrid predictive maintenance system (Fig. 1). On one side, real process data are collected. On the other side, virtually generated failure data are recorded to create a broader failure database and to avoid overfitting.

A crucial factor is the further development of the virtual commissioning model into a digital twin. First, all process parameters from the real machine are entered into the virtual model, like e.g. axis parameters or closing and opening times of cylinders. Furthermore, the current PLC program of the real plant is imported into the HiL configuration and then the virtual model is optimized until the automatic mode is in operation.

Ultimately, the developed failure data generation module is implemented in the corresponding process in order to simulate failure data based on real patterns and trends. In case of a constructional change of the real system, the model has to be modified accordingly.

After the digital twin behaves like the real machine in automatic mode, synthetic failure data can be generated. These are then incorporated into the prediction model from the real plant. This solves the problem that there often is not enough data from the time of failure to robustly train a prediction model.
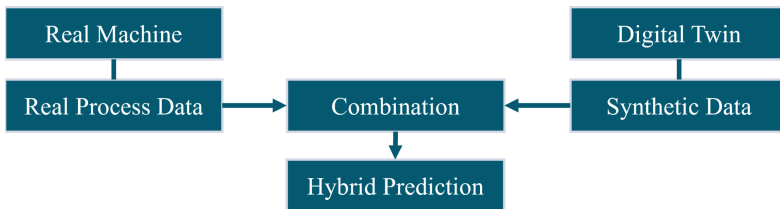
Fig. 3. Structure of a hybrid prediction

3.3 Example of an Industrial Use Case

A new approach in the field of electric engines for automobiles is to use stiff, u-shaped hairpins in the stator instead of a copper winding made of round wire. In the production
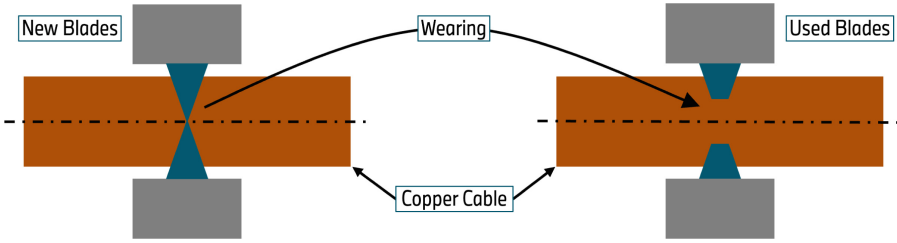
Fig. 4. Illustration of a cutting process in the Hairpin production

of the hairpins, there is a cutting process which cuts the raw material of the copper coil to the required length, which is then bent into the required shape.

The cutting tools of the process (Fig. 4) are worn down and lose their sharpness with every cut. Once the cutting edges are worn out, the machine stops. The blades have to be replaced. Changing them requires maintenance, which in turn requires production to be halted for some time. Since the wear depends especially on external influences, there is no exact regularity in the failures of the line. On the other hand, the costs for a preventive approach, e.g., replacing every week, are too high. Therefore, this example forms the ideal use case for a predictive maintenance application.

Prediction methods require a certain number of data sets to be able to predict with sufficient accuracy. The more irregular the patterns and trends that occur, the more data is needed to identify stable correlations. Conventional recording of process data therefore requires a significant time. Therefore, we developed a component in ISG-virtuos (Fig. 5) that incorporates the patterns and trends from real process data (*LookUp Table*) in the digital twin and then randomizes them (*Randomization*). The port in the top left corner (*TurnOn_SyntheticData*) allows the user to toggle this option.

By doing so, additional training data can be generated, in parallel to the real machine's data plus any perturbations and failures the user chooses to include. Thus, the training data can be gathered at twice the speed or even faster, by using more setups simultaneously. Additionally, the generated data can yield more variation than the real data. Yet another advantage of the synthetic data generation with the digital twin, based on the further developed virtual commissioning model, is that the effect of aging on subsequent processes can be directly observed and investigated.
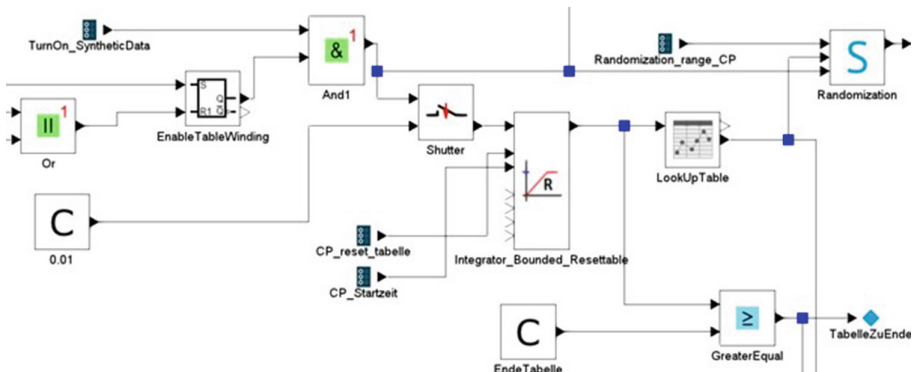
Fig. 5. Developed module for generating synthetic failure data with ISG-virtuos

3.4 Challenges

To build a hybrid prediction model, a significantly more profound understanding of the process is required so that the virtual model can be further developed into a digital twin. On one side, it must be understood how the digital twin can be used to provide additional information about the occurring failures. Use and business cases must be identified and evaluated. On the other side, the virtual commissioning model has to be extended correspondingly to be able to generate synthetic failure data.

Data pre-processing is particularly challenging, since in practice a lot of data are recorded, but usually only a small part of it is relevant for predicting a failure. When using machine learning algorithms to predict failures, it is important to define an accurate benchmark to select the best model. It is also possible to average multiple predictions depending on how well the forecasts perform.

4 Conclusion and Outlook

With virtual commissioning models becoming increasingly available, this paper presents a method for evolving them towards a hybrid predictive maintenance system. Finally, the presented results are concluded and an outlook on future work is proposed.

4.1 Conclusion

The concept of a hybrid predictive maintenance system makes it possible to generate predictions more efficiently and accurately. The virtual commissioning model is further developed into a digital twin of the machine. With this, synthetic failure data are generated based on the patterns and trends of real process data. This helps training prediction algorithms faster and more accurately.

The core requirement for the creation of such a hybrid system is the existence of the virtual commissioning model. In addition, the plant has to be producing in a stable condition and has to offer the possibility to collect data. First experiences of experts regarding the failures of the plant as well as the underlying mechanisms are of significant advantage. If these conditions are fulfilled, a use case can be identified where the benefit can outweigh the effort.

From an economic point of view, this concept is advantageous in several respects. First, the cost-intensive virtual commissioning models can be further utilized. Second, the synthetically generated fault data can be used to implement a predictive maintenance system at an earlier stage compared to conventional procedures, improving plant availability and reducing maintenance costs.

It has to be noted that in order to follow the presented method and to actually set up such a maintenance system, expert knowledge is required. Without it, the use cases cannot be identified, and the hybrid system cannot be implemented to the necessary extent. It is of utmost importance to also identify business cases and ensure economic benefits. A digital twin always needs to serve a concrete purpose. Additionally, the approaches have to be adapted to existing structures and processes. If the maintenance is not equipped to train models and use their output, the benefits will diminish. In conclusion, in the right hands, this work presents a further step towards the digitalisation of production and maintenance.

4.2   Outlook

Digital twins in general and the hybrid system in particular hold a lot of potential, as this work shows. However, current virtual commissioning models usually do not have the necessary modelling depth to generate synthetic failure data. Therefore, in the further development to the digital twin, the sub-process to be predicted must generally be modelled in even more detail. The automation of this could be subject of future research. If use cases are recognized early enough, the models can be created with enough detail from the start, thus avoiding additional costs later on.

The digital twin is also well suited for cycle time simulations. In this case, aging in certain sub-processes can be simulated and the effect of this on following operations can be analysed. The next step could be the observation of the decrease of parts produced, due to wearing effects. Ultimately, further knowledge for maintenance and production can be generated.

References

1. Pawellek, G.: Integrierte Instandhaltung und Ersatzteillogistik: Vorgehensweisen, Metho-den, Tools, 2nd edn. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-662-486 67-2. Accessed 28 Jan 2022
2. Werner, A., Angadi, V., Lentes, J., Mousavi, A.: "Vorausschauende Instandhaltung – Wenn der Digitale Schatten an seine Grenzen stößt", (in de). Zeitschrift für wirtschaftlichen Fabrikbetrieb **115**(5), 335–339 (2020)
3. Ansari, F., Glawar, R., Nemeth, T.: PriMa: a prescriptive maintenance model for cyber-physical production systems. Int. J. Comput. Integr. Manuf. **32**(4–5), 482–503 (2019). https://doi.org/10.1080/0951192X.2019.1571236
4. Nemeth, T.: Vorgehensmodell zur Einführung und Reifegradbewertung einer präskriptiven Instandhaltungsstrategie unter Zuhilfenahme von Qualitätsmetriken. Dissertation, Fakultät für Maschinenwesen und Betriebswissenschaften, Technische Universität Wien, Wien (2020). Accessed 5 Dec 2021
5. Bracht, U., Geckler, D., Wenzel, S.: Digitale Fabrik: Methoden und Praxisbeispiele, 2nd edn. Springer, Heidelberg (2018). https://doi.org/10.1007/978-3-662-55783-9. Accessed 22 Jan 2022
6. Wünsch, G.: Methoden für die virtuelle Inbetriebnahme automatisierter Produktionssysteme. Zugl.: München, Techn. Univ., Dissertation, 2007. Utz, München (2008). Accessed 8 Jan 2022
7. Kövari, L.: Konzeption und Realisierung eines neuen Systems zur produktbegleitenden virtuellen Inbetriebnahme komplexer Förderanlagen. Dissertation. KIT Scientific Publishing, Karlsruhe (2011). Accessed 20 Jan 2022
8. Bauernhansl, T., Hompel, M., Vogel-Heuser, B.: Industrie 4.0 in Produktion, Automatisierung und Logistik: Anwendung, Technologien, Migration. Springer Vieweg, Wiesbaden (2014). https://doi.org/10.1007/978-3-658-04682-8. Accessed 30 Dec 2021
9. Zäh, M., Wünsch, G., Hensel, T., Lindworsky, A.: Nutzen der virtuellen Inbetriebnahme: Ein Experiment. Zeitschrift für wirtschaftlichen Fabrikbetrieb **101**(10), 595–599 (2006). https://doi.org/10.3139/104.101070
10. Drescher, B., Stich, P., Kiefer, J., Strahilow, A., Bär, T., Reinhart, G.: Physikbasierte Simula-tion im Anlagenentstehungsprozess – Einsatzpotenziale bei der Entwicklung automatisierter Montageanlagen im Automobilbau. In: ASIM-Mitteilung, vol. 147, Simulation in Produktion und Logistik, pp. 271–281, Univ. Paderborn, Paderborn (2013)

11. Shen, Y., et al.: Predicting tool wear size across multi-cutting conditions using advanced machine learning techniques. J. Intell. Manuf. **32**(6), 1753–1766 (2020). https://doi.org/10.1007/s10845-020-01625-7

12. Aivaliotis, P., Georgoulias, K., Chryssolouris, G.: A RUL calculation approach based on physical-based simulation models for predictive maintenance. In: International Conference on Engineering, Technology and Innovation (ICE/ITMC), pp. 1243–1246 (2017). https://doi.org/10.1109/ICE.2017.8280022

13. Gonzalez, M., Salgado, O., Croes, J., Pluymers, B., Desmet, W.: A digital twin for operational evaluation of vertical transportation systems. IEEE Access **8**, 114389–114400 (2020). https://doi.org/10.1109/ACCESS.2020.3001686

14. Luo, W., Hu, T., Zhang, C., Wei, Y.: Digital twin for CNC machine tool: modeling and using strategy. J. Ambient. Intell. Humaniz. Comput. **10**(3), 1129–1140 (2018). https://doi.org/10.1007/s12652-018-0946-5

15. Kritzinger, W., Karner, M., Traar, G., Henjes, J., Sihn, W.: Digital Twin in manufacturing: a categorical literature review and classification. IFAC-PapersOnLine **51**(11), 1016–1022 (2018). https://doi.org/10.1016/j.ifacol.2018.08.474

16. Gouriveau, R.: From Prognostics and Health Systems Management to Predictive Maintenance 1: Monitoring and Prognostics. Wiley, Newark (2016). http://gbv.eblib.com/patron/FullRecord.aspx?p=4718318

Hybrid Digital Twins Using FMUs to Increase the Validity and Domain of Virtual Commissioning Simulations

Denis Pfeifer[1,2(✉)], Andreas Baumann[2], Marco Giani[3], Christian Scheifele[1], and Jörg Fehr[2]

[1] ISG Industrielle Steuerungstechnik GmbH, Gropiusplatz 10, 70563 Stuttgart, Germany
denis.pfeifer@isg-stuttgart.de
[2] Institute of Engineering and Computational Mechanics, University of Stuttgart, Pfaffenwaldring 9, 70569 Stuttgart, Germany
[3] Robert Bosch GmbH, Robert-Bosch-Campus 1, 71272 Renningen, Germany

Abstract. The main objective of virtual commissioning is to help design and validate the control systems of entire production plants. Therefore, simulations on a logical and kinematic level are performed, typically in a Software- or Hardware-in-the-Loop configuration using the original control software and controller [1].

However, the lack of level of detail means that this type of simulation is insufficient for an integrated system dynamics and control algorithms design. These engineering tasks are currently performed in separate tools, e.g. by finite element analysis, multibody simulations or by a combination, i.e. elastic multibody systems (EMBS) [2]. However, the designed components are only considered individually and not in the context of the control technology used. Therefore, primarily synthetic inputs are used and not the original control behavior. With a higher level of simulation detail, further questions about the system, such as the effect of control algorithms on the dynamic processes, can be virtually validated.

Therefore, this paper explores hybrid component-based digital twins to combine the advantages of both VC and EMBS. Hybrid components allow the simulation of the interactions between process, machine and control system with a high level of detail where this is beneficial. Such integration is achieved using the Functional Mock-up Interface (FMI) to couple different simulation models in a co-simulation environment [3]. This is demonstrated in a simulation use case of an inverted pendulum. The level of detail of individual components in the virtual commissioning tool ISG-virtuos [4] is increased by the modular integration of elastic multibody simulations via FMI so that the swing-up controller can be designed in the simulation.

Keywords: Digital Twins · Virtual Commissioning · Functional Mock-up Interface · Software-defined Manufacturing · Multibody Simulation

1   Introduction

In the past virtual commissioning was successfully used to help design and validate the control software of a production plant in simulations rather than on the real machinery. The real-time simulation models are called digital twins in the context of virtual commissioning, and they typically describe production facilities on a logical and kinematic level. These simulations are performed in real-time with the control software running on the real controller or a virtual controller, so-called hardware- or software-in-the-loop simulations (HiLS/SiLS). Virtual commissioning (VC) helped to significantly reduce commissioning times and improve software quality, thus reducing cost and increasing productivity of production plants [1].

So far, the simulation on a logical and kinematic level was sufficient for many tasks, such as I/O tests and the verification of the motion direction of machine axes, etc. However, certain tasks, such as the design of feedback controllers for industrial control systems, are still performed on the real machine rather than in the virtual world.

Following the idea of software-defined manufacturing (SDM) [5], where a machine's functionality is determined by the software running it, these tasks need to be shifted to the virtual world, where optimizations can be performed much easier and at a lower cost. This leads to better production plants with increased efficiency, minimized production cost and shorter development cycles. Therefore, the domain of virtual commissioning simulation needs to be extended to the dynamics of a production plant.

For successful integration of dynamics into virtual commissioning, the following requirements must be met:

- (Re-)Use of the know-how about a system's dynamics from previous engineering stages for virtual commissioning (not modeling the dynamics again).
- Virtual commissioning can parametrize dynamics models without the assistance of dynamics experts.
- Dynamics models can run autonomously in virtual commissioning software and do not need additional engineering software to run so that models can easily be transferred between simulation systems.
- Production plants of a similar size to which are simulated today in virtual commissioning can be simulated, and dynamics models can be real-time capable.

In Sect. 2, the state of the art is presented to lay the foundation why hybrid digital twins are needed. In Sect. 3, the methods to increase the level of detail of virtual commissioning simulations to the dynamic level while meeting these requirements are explained. This is achieved through the concept of hybrid digital twins on a component basis using functional mock-up units (FMUs), a standardized format for simulation models which follows the functional mock-up interface. This concept is validated in the concrete use case of designing a swing-up controller for an inverted pendulum in a Hardware-in-the-Loop simulation in Sect. 4. In the final Sect. 5, the conclusion is drawn.

2 State of the Art in Virtual Commissioning Simulations with Increased Level of Detail

Many researchers found ways to introduce a higher level of detail into virtual commissioning simulations, of which some representative examples are listed here, together with their benefits and shortcomings:

Hoher [6] developed a deterministic physics engine to calculate the behavior globally in the simulation scene. This engine is best suited for calculating material flow in deterministic real-time for HiLS. In order to achieve determinism and to keep the computational effort low so that the behavior can be simulated globally, Hoher's engine significantly abstracts collision detection and rigid body dynamics. This results in just a slight increase in level of detail, which suffices for basic material flow simulations but not, for example, feedback controller design.

Sekler [7], on the other hand, examines the vibration behavior of machine tools in detail using finite element (FE) analysis. FE models are exported from an FE software via a proprietary format and are then coupled within the virtual commissioning software to obtain the system. This approach shows great modularity. However, it is limited to a linear description of systems, and demands knowledge about dynamics from the VC engineer.

Heiland et al. [8] and Klingel et al. [9] couple the FE-simulation of a deep drawing process to a virtual commissioning tool using the standardized format for simulation models functional mock-up unit (FMU). However, their method relies on executing the FE-simulation program during VC. Further, it is not real-time capable and thus is limited to SiLS with a virtual timescale.

The research on increasing the level of detail in virtual commissioning simulations so far focuses on either global calculation of dynamics. It is thus limited in the level of detail reached by the available computational power. Alternatively, the emphasis is on reaching a high level of detail, but then only small systems can be regarded, and dynamics experts are required to set up the simulation.

3 Methods and Concept of Hybrid Digital Twins

In order to increase the level of detail of virtual commissioning simulations to the dynamic level while meeting the requirements proposed in Sect. 1, the concept of hybrid digital twins on a component-basis using the method of integrating functional mock-up units is proposed.

Digital twins are modelled on a component basis to achieve modularity and thus reusability for different applications [10]. Nowadays, the components manufacturers supply digital twins of their components next to the real ones [11]. The manufacturer's dynamics experts already possess the know-how about the dynamics of the components, but the boundaries between disciplines prevent the implementation of digital twins for virtual commissioning. The method to model dynamic behavior and how to integrate it into component models for VC is proposed as follows:

The method used to model a production plant's dynamics is as elastic multibody systems (EMBS) [2]. EMBS involve large rigid body movements and small elastic body movements, as typically found in production plants and machine tools. Starting from the CAD model, elastic bodies are modelled and then meshed and equipped with material properties in an FE software. A popular option is to then reduce the model's complexity through model order reduction (MOR). Next, the flexible bodies are coupled to the rigid bodies in an EMBS software, where the nonlinear equations of motion of the whole mechanical system are calculated. The equations of motion of EMBS are written as a differential equations system of the following form:

$$
\begin{bmatrix}
m\boldsymbol{I} & & sym. \\
m\,\widetilde{\boldsymbol{c}}\,(q) & \boldsymbol{J}(q) & \\
\boldsymbol{C}_\mathrm{t}(q) & \boldsymbol{C}_\mathrm{r}(q) & \boldsymbol{M}_\mathrm{e}
\end{bmatrix}
\cdot
\begin{bmatrix}
\ddot{\boldsymbol{q}}_\mathrm{t} \\
\ddot{\boldsymbol{q}}_\mathrm{r} \\
\ddot{\boldsymbol{q}}_\mathrm{e}
\end{bmatrix}
+
\begin{bmatrix}
\boldsymbol{k}_\mathrm{t} \\
\boldsymbol{k}_\mathrm{r} \\
\boldsymbol{k}_\mathrm{e}
\end{bmatrix}
=
\begin{bmatrix}
\boldsymbol{h}_\mathrm{t} \\
\boldsymbol{h}_\mathrm{r} \\
\boldsymbol{h}_\mathrm{e}
\end{bmatrix}
$$

In those equations, the indices t, r, e stand for the equations' translational, rotational and linear elastic pars. q are the generalized coordinates, k the Coriolis and centrifugal, as well as dampening and spring forces, and h the generalized applied forces. The mass matrix consists of the masses of the individual bodies m, the inertia matrix of rotation $\boldsymbol{J}(q)$, and the elastic mass matrix $\boldsymbol{M}_\mathrm{e}$. $\widetilde{c}\,(q)$ describes the variable center of gravity of the flexible bodies, and $\boldsymbol{C}(q)$ are the coupling terms [12].

FMUs are well-suited as a standardized format for such differential equations and thus for exchanging dynamics models between the two disciplines. This allows the dynamics to be modelled in domain-specific tools and to be exported together with a numerical solver as an FMU. The FMU can then be used in virtual commissioning, rather than modeling the dynamics in the VC tool, following the idea of domain-driven design [13]. Thus, the know-how of dynamics experts can be used further along the development process, ensuring that the models used are valid. Another advantage is that FMUs can be designed as black-box models with limited inputs, outputs, and parameters so that VC engineers without extensive knowledge about dynamics can parametrize simulations easily and time-efficiently.

Co-simulation techniques enable the seamless integration of FMUs in combination with different simulation setups, such as SiLS and HiLS. The distribution of computational effort onto different CPU cores means that large production plants can be simulated [14].

These methods used to model an EMBS and to bring it into virtual commissioning are displayed in Fig. 1 with the indication of which steps are typically performed by dynamics experts and which by VC Engineers.
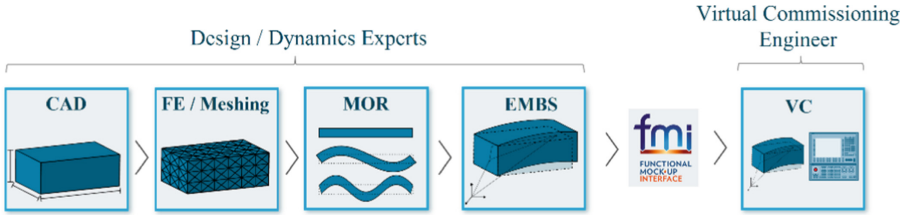
Fig. 1. Engineering workflow for using EMBS in virtual commissioning

However, even with this approach, the inclusion of the dynamic level results in an increase in computational cost and modeling efforts. Since this level of detail is not necessary everywhere in the simulation scene at the same time, but only locally, component models need the ability to turn on and off their dynamic behavior: When turned off, they perform like classic digital twins for VC on a logical and kinematic level, and when turned on they add the additional level of detail from the dynamics model, see Fig. 2. These digital twin components are called hybrid digital twins.
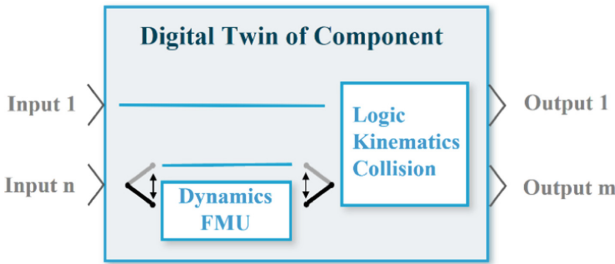
Fig. 2. Digital Twin of a component with hybrid level of detail

The inputs and output signals of hybrid digital twins are the same as the signals of regular component models [10, 11], for example position values or logical signals. Thus, they can be used just the same to build up the digital twin of a whole production plant or machine tool in a modular manner.

4   Validation Use Case Inverted Pendulum

The proposed method for extending the domain of virtual commissioning simulations by increasing the level of detail through the use of hybrid models is validated on the use case of an inverted pendulum (see Fig. 3). A Bosch Rexroth ctrlX industrial controller controls the pendulum of length $l = 400$mm and mass $m = 0.4$ kg. An IndraDrive drive with position feedback controls the force F on the linear sled of mass $m = 28.8$ kg, which moves along its axis of length $d = 880$ mm with position x. The pendulum body is attached to the sled, and its angle α is measured by a rotary encoder (between the sled and the pendulum's attachment point to the sled).
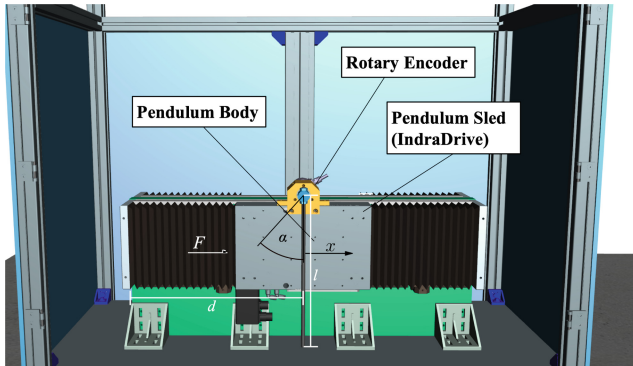
Fig. 3. Validation use case inverted pendulum

4.1 Modeling and Simulation Setup

The inverted pendulum was modelled following the workflow in Fig. 1. Starting from the CAD model, the pendulum body was meshed and equipped with material properties in the FE software Ansys [15]. From there, it was imported into the model order reduction software MatMorembs [16], where a modal reduction was performed, resulting in a flexible body with three degrees of freedom. Using the simulation software Neweul-M^2 [17], the flexible body was then coupled to the sled, which is modelled as a rigid body. This resulted in an elastic multibody system. Next, the nonlinear equations of motion where calculated and a numerical solver was chosen.

The elastic multibody model of the inverted pendulum could then be exported as a co-simulation FMU, consisting of the nonlinear equations of motion and the numerical solver.

The FMU is imported in the VC tool ISG-virtuos where it is integrated into the pendulum model for virtual commissioning, forming a hybrid component.

With the traditional VC model of the pendulum, containing the logical and kinematic information, the motion direction of the sled could be verified, as well as the pendulum's zero-point and the encoder's resolution. The dynamics of the pendulum can then be switched on in order to design a feedback controller for the swing-up motion of the pendulum, resulting in a simulation setup as displayed in Fig. 4. The controller of the pendulum is running on the Bosch Rexroth ctrl X industrial controller (cycle time = 2 ms) and communicates via the EtherCAT fieldbus with the simulation system running on the TwinCAT 3 real-time operating system (cycle time = 1 ms). In the TwinCAT partition of the simulation, the drive model (Sercos state machine), the rotary encoder and the logical and kinematic part of the hybrid pendulum component are simulated. The FMU containing the pendulum's dynamics is simulated in the Microsoft Windows partition.

The implemented controller used for the swing-up consists of an energy controller which swings up the pendulum to a certain angle and then switches to an LQR controller, as demonstrated by [18]. Using the hybrid component, the controller parameters could be tuned to achieve the desired swing-up behavior.
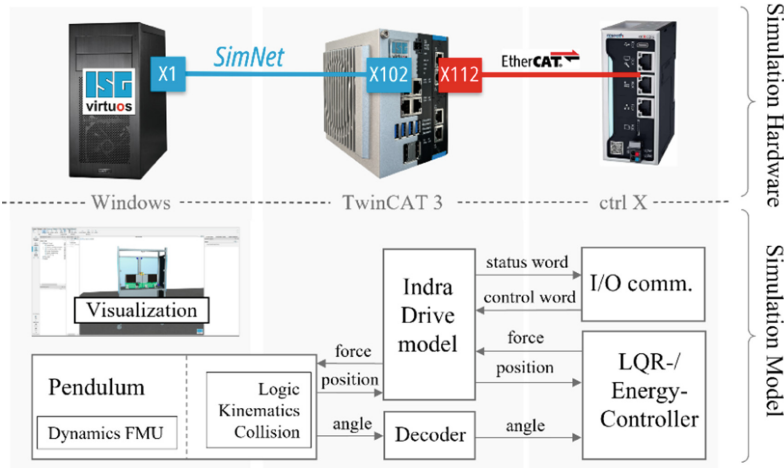

Fig. 4. Hardware-in-the-loop simulation setup with model partitioning

4.2 Simulation Results and Interpretation

In Fig. 5, the angle during the swing-up is displayed in blue, the sled's position in red. Simulations are plotted for a controller with a high gain (full lines) and a low gain (dashed lines). Both controllers manage to successfully swing up the pendulum and stabilize it once it reaches the upper equilibrium. However, only for the controller with the high gain, the sled stays within the boundaries of the linear axis it runs on. The downside to the high gain controller is the strong reactions to the disturbances. The disturbances originate from small time deviations of the Windows simulation, since Windows is not a real-time operating system. The effect that can be observed is the relatively strong reaction of the controller to these disturbances in the form of oscillations when the pendulum is in its upper equilibrium position.

The simulation with the low gain controller shows that the pendulum reacts less strongly to the disturbances, but the sled needs a longer axis for the swing up. In a development process, the pendulum's construction could be iterated by lengthening the axis, for example, in order to achieve a smoother behavior.
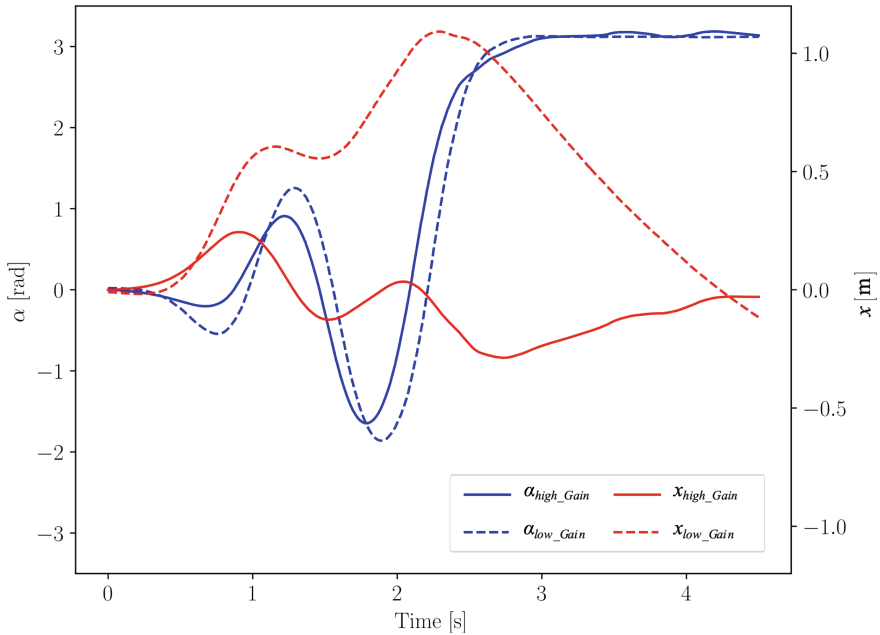
Fig. 5. Sled position and pendulum angle during swing-up

5 Conclusion and Outlook

Through hybrid components, the domain of virtual commissioning models can be extended to the dynamic level. Dynamic systems can be modelled in domain-specific tools and are then imported into VC component models using the standardized format FMU, enabling VC engineers to parametrize the simulation easily and time-efficiently. With the ability to selectively enable the dynamic behavior, a local increase in level of detail can be reached where needed, keeping overall computational cost low. Hybrid components thus extend the validity of VC simulations from only logic and kinematic-related engineering tasks to tasks such as the integrated system dynamics and control algorithms design.

The concept of hybrid components was validated on the model of an inverted pendulum. The pendulum was designed as a hybrid component with an optional dynamics part modelled as an elastic multibody system. This allowed the pendulum's controller to be successfully designed in a hardware-in-the-loop simulation. The controller design would have previously been possible only once the real system is built. Further, the controller parameters could be tuned, and the simulation results can even serve as a basis for reiterating the pendulum's construction. Reiteration can be done without much additional effort since the project is still in the development phase, and none of the pendulum's hardware has been built yet.

Future research can be directed at improving the real-time behavior of the dynamics simulation by transferring it to a real-time operating system. Further, model-based approaches, such as the implementation of a precontrol, can be researched.

References

1. Pritschow, G., Röck, S.: Hardware in the loop simulation of machine tools. In: CIRP Annals (2004)
2. Siedl, D.: Simulation des dynamischen Verhaltens von Werkzeugmaschinen während Verfahrbewegungen. In: Verlag, H.U. (ed.) Herbert Utz Verlag (2008)
3. Blockwitz, T., et al.: Functional mockup interface 2.0: the standard for tool independent exchange of simulation models. In: 9th International Modelica Conference (2012)
4. ISG-virtuos homepage. https://www.isg-stuttgart.de/de/digitaler-zwilling. Accessed 05 Jun 2022
5. Lechler, A., Kircher, C., Verl, A.: SDM - Software Defined Manufacturing, wt Werkstattstechnik online (2018)
6. Hoher, S.: Ein gekoppeltes Materialflussmodell zur durchgängigen Entwicklungsunterstützung von Materialflusssteuerungen, Ph.D. dissertation, Institut für Steuerungstechnik der Werkzeugmaschinen und Fertigungseinrichtungen Universität Stuttgart (2017)
7. Sekler, P.: Modellbasierte Berechnung der Systemeigenschaften von Maschinenstrukturen auf der Steuerung, Ph.D. dissertation, Institut für Steuerungstechnik der Werkzeugmaschinen und Fertigungseinrichtungen Universität Stuttgart (2012)
8. Heiland, S., et al.: Virtual tool commissioning using LS-DYNA functional mock-up interface. In: 13th European LS-DYNA Conference 2021 (2021)
9. Klingel, L., Penter, L., Mayer, P., Steffen, I., Verl, A.: Digital Twins in deep drawing for virtual tool commissioning and inline parameter optimization. In: International Deep-Drawing Research Group Conference (IDDRG 2022) (2022)
10. Scheifele, S.M.: Generierung des Digitalen Zwillings für den Sondermaschinenbau mit Losgröße 1, Ph.D. dissertation, Institut für Steuerungstechnik der Werkzeugmaschinen und Fertigungseinrichtungen Universität Stuttgart (2020)
11. Scheifele, C., Verl, A., Tekouo, W., Belgharda, S., Mauderer, T.: Eine Online-Plattform für Digitale Zwillinge: Austausch, Pflege und Bereitstellung von Simulationsmodellen zur virtuellen Inbetriebnahme, atp magazin (2020)
12. Shabana, A.A.: Computational Dynamics. Wiley, Hoboken (2009)
13. Evans, E.: Domain-driven design: tackling complexity in the heart of software. Evans, E. (ed) Addison-Wesley Professional, Boston (2004)
14. Scheifele, C., Verl, A., Riedel, O.: Real-time co-simulation for the virtual commissioning of production systems. Procedia CIRP **79**, 397–402 (2019)
15. Ansys Homepage. https://www.ansys.com/de-de. Accessed 05 Jun 2022. Available: https://www.ansys.com/de-de
16. Fehr, J., Grunert, D., Holzwarth, P., Fröhlich, B., Walker, N., Eberhard, P.: Morembs—a model order reduction package for elastic multibody systems and beyond. In: Keiper, W., Milde, A., Volkwein, S. (eds.) Reduced-Order Modeling (ROM) for Simulation and Optimization, pp. 141–166. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75319-5_7

17. Kurz, T., Eberhard, P., Henninger, C., Schiehlen, W.: From Neweul to Neweul-M2: symbolical equations of motion for multibody system analysis and synthesis. In: Multibody System Dynamics (2010)
18. Aström, K.J., Furuta, K.: Swinging up a pendulum by energy control. Automatica **36**(2), 287–295 (2000)

Towards the Smart Factory: Process Optimization in Virtual Commissioning

Alexander M. Kuhn[1(✉)], Michael Christ[1], Christopher B. Kuhn[1], Peng Liu[1], William Tekouo[1], and Hans A. Kestler[2]

[1] BMW AG, 80788 Munich, Germany
alexander.ak.kuhn@bmw.de

[2] Universität Ulm, Albert -Einstein -Allee 11, 89081 Ulm, Germany

Abstract. Since the early 1990s, virtual models have been used in production planning to digitally support production lines before beginning physical construction. These digital models, commonly referred to as "digital twins", are currently primarily used for virtual commissioning. Despite being first used decades ago, digital twins are still not established in production planning on a global scale. The benefits of developing and testing a planned system in a virtual model are often not fully capitalized. While this can be partially explained by the technological challenges of designing accurate virtual models, we argue that the current processes for production planning are another important factor that hinder the widespread use of digital twins. In this paper, we summarize and analyze each step and the involved participants in a typical production planning workflow. Based on this analysis, we discuss how current practices conflict with the goal of incorporating virtual models into the established work environment. Then, we derive concrete recommendations of how a production process can be adjusted to allow easier digital twinning, showing how comparatively few changes could make virtual models significantly easier to apply.

Keywords: Digital Twin · Virtual Commissioning · Information Management · Dynamic Production · Automotive Production

1 Introduction

Ever changing political requirements and social trends accelerate the ongoing transformation process in the automotive industry [1]. EU climate goals, the COVID-19 pandemic, and the ongoing chip shortage keep adding pressure to the companies. A growing product variety over the last years further increased process complexity [2]. One potential approach for handling this complexity is the concept of digitalization. Virtual commissioning (VC), in particular, allows for a reliable virtual validation of production systems without the physical machine having to be built. Despite this idea being almost 30 years old, a widespread establishment of this technology still remains to be seen. Even recent papers

such as [3] from 2021 still investigate how to use and integrate virtual commissioning, demonstrating the lack of practical experience in this field. Considering the significant potential benefits of digitalization for production systems, the question arises why tools such as virtual commissioning are not yet ubiquitous in the industry. While digitalization is not a new topic, there is still a lack of comprehensive work to investigate why digitalization sometimes fails to be established in the first place. This paper addresses this pressing issue and proposes a practical workflow for resolving these challenges. We investigate what the reasons for the hesitant adaptation of digitalization might be. We focus on the general process workflow in production planning and analyze it with regards to its potential for digitalization. We make the following main contributions:

1. Identifying the obstacles for digital technologies such as virtual commissioning in the processes typical to the industry
2. Conceptualizing mitigating measures that allow for removing these obstacles
3. Showing how these solutions can be applied in practice

The rest of this paper is structured as follows. In Sect. 2, we summarize the state of the art. In Sect. 3, we analyze a typical production process. In Sect. 4, we discuss how to optimize such a process. Section 5 concludes this work and gives an outlook on the next steps.

2 State of the Art

To understand the obstacles that prevent wide-spread use of digitalization in the industry, the current state of production plant planning is summarized first.

2.1 Conventional Planning

Traditionally, the planning process of production systems can be categorized into four distinct phases, as described by Weber [4] in 1992: draft, development&planning, construction, commissioning, and production.

After commissioning, the machine changes owners, along with all its responsibilities. While the subsequent production phase also faces challenges, this work focuses on the earlier stages, where digital tools are already being applied and a mitigation of errors can be observed [5]. Modern, automated production lines are highly complex mechatronic systems. Achieving an interdisciplinary and parallel collaboration of the disciplines of mechanics, electronics, and information technology requires novel approaches. To ensure smooth operation, data consistency appears to be a fundamental requirement to allow for the necessary exchange between the fields. In general, the development is attributed the most significant role, since its results will determine the future productivity and efficiency of the machine. An example of a modern way of development collaboration is presented in VDI 2206 [6].

Starting with the mechanical design, the electronics can only be specified once the 3D concepts are available and signed off on by the client. The electronic planning needs to be completed to allow for the software development to begin. Works such as [7] emphasize that software plays a major part in modern plants.

2.2   Digital Factory and Twin

Terms such as digital factory and digital twin have been around for a long time. In theory, they should enable production lines to be planned more efficiently and in a more flexible way. The core principle is to use tools such as simulation to evaluate concepts without the risk of costly changes late in the project or even hardware damages on the construction site [8].

The idea of a digital factory encompasses a digitalization of an entire factory. In this work, we restrict our focus on the concept of a digital twin. While there are different definitions, this paper will use this common understanding: A digital twin describes a digital, i.e., simulated, entity running in parallel to the real machine [9]. Only recently, more research has been dedicated to the question of how to methodically introduce digital technologies to existing structures [2]. This can be seen as yet another indication that, despite their potential, digital twins are not yet established everywhere, and careful decision making is required to turn the use cases into beneficial business cases. To support such decision making, the processes used in the industry need to be properly assessed and potentially changed to facilitate the introduction of digitalization. Barbieri et al. propose a high-level approach on how to get to a digital twin from virtual commissioning, but do not go into detail on how to properly introduce the foundation for digitalization [10]. Similarly, Leng et al. focus on digital twins and on how to use them for smart manufacturing systems [11]. The work in [12] provides a general overview of virtual commissioning with its benefits and pitfalls. So, while some work has been done on virtual commissioning and its practical implications such as the guideline in [12], work directly derived from production systems is still lacking. To address this, our paper focuses on the practical applicability of virtual commissioning. We derived all concepts proposed in this work and all results obtained from our analysis directly from the automotive production sector. The work by Liebrecht goes into more detail than the VDMA guideline in [2]. Similarly, we build on the VDMA guideline to investigate the claims from [12] hands-on in the automotive industry. We therefore perform an in-depth analysis to serve as a starting point for optimizing the processes used in this field.

3   Process Analysis

In this chapter, we analyze a typical process in production planning as a first step towards assessing why digitalization is not more commonly used. While addressing and improving technical aspects of the implementation is one factor, we focus

on how the underlying processes can be changed instead. Problems such as computing performance, hardware cost, or data consistency are present in most technological fields, but are much harder to change than the processes established in production planning. Changes to these processes can be as straightforward as changing existing specification sheets. Next, we first discuss the main obstacles faced in a planning process. Then, we derive potential solutions that are straightforward to implement. This problem analysis along with its considerations all stem from automotive production planning. We gathered insights by interviewing experts in the field of planning and virtual commissioning.

3.1 Metrics

Before analyzing the processes to identify possible problems, it is crucial to establish metrics by which to measure these problems by. They are derived from expert interviews and hands-on experience by the authors. We summarize the proposed metrics in Table 1.

While cost efficiency might be the most important goal in any company, we argue the most important metric is time. It can be dangerous to try and optimize for money, if it diminishes the output quality. In that case late improvements are required, which are significantly more expensive than investing a comparatively small amount during the early stage of a project. So, rather than trying to save money, we argue that the most important aspect should be to try and be as efficient as possible. If done correctly, this will automatically reduce costs, while simultaneously improving the acceptance.

Table 1. Metrics for process analysis and solution identification

Metric	Description	Prio.
Time efficiency	processes, automation, standardization	1
Employee acceptance	specifications, trainings, workshops	2
Information flow	automation tools, exchange formats	3

This leads to the second important metric, the acceptance amongst employees. Without it, no technology can last. The human factor plays a significant role and must not be neglected, because ultimately any technology ends up being used by human employees and support tools have to be focused on them.

Finally, this paper proposes to consider the information flow along the processes. By putting emphasis on its quality, problems such as lacking synchronization and having to do work multiple times by not realizing available synergies can be identified and avoided.

3.2   Problem Analysis

In this section, we identify four key problems that current production processes face regarding the implementation of digital twins and virtual commissioning. The first problem is a lack of synchronization across the planning pipeline. A typical planning process can be divided into several main phases: 3D construction, electrical planning, software programming, and the actual physical construction site. These phases are typically not sufficiently synchronized or connected. Following a sequential pattern, information can get lost and usually only flows in one direction: forward. Changes in a later phase are thus often not properly fed back to the initial phases. With such a restricted flow of information, digital twins are inherently restricted as well, which in turn prevents them from becoming established in the industry.

A second issue is the fact that the processes used in the industry have been established over decades. Changing such a fixed process is inherently challenging since these processes were not designed to be flexible or dynamic. By using digital twins and virtual commissioning, the CAD data is required much earlier. This requirement disrupts the conventional workflow. The programming of the software required for a process has to start significantly earlier.

Thirdly, the human factor is an important issue. Digital tools can only fulfill their potential if the people involved in the production have sufficient knowledge in the field. Considering the current lack of digitalization in a typical production planning, the expertise in corresponding tools subsequently is low.

The last problem, which is easily overlooked when focusing only on the technical side of digitalization, is the increased demands on the management of production plants. The managerial support that is required for any change to take lasting effect is significant.

3.3   Problem Solutions

For remedying the problems outlined in the previous section, a range of solutions can be considered. Building upon the work of [13], we categorize potential solutions into three main categories: process-related (PR), human-related (HR), and technical-related (HR), as described in Table 2. The right column illustrates the priority, i.e., the urgency, with which a solution should be implemented. To illustrate the problems described in the previous section and the identified solution categories, we take a look at a conventional planning process in Figure 1. It is shown, what exemplary problems can hinder the effective use of digital tools such as virtual commissioning. We depicted the identified problems along with applicable solution categories in the same figure.
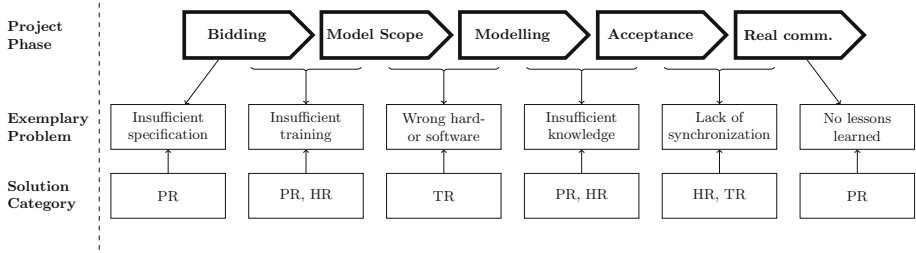
Fig. 1. VC process with identified problems and possible solutions from Table 2

Table 2. Solution categories, prioritized by urgency for efficient digitalization.

Category	Solution	Prio.
Process-related [PR]	Identification of starting points for digitalization	1
	Specification documents adapted to the growing processes	1
	Expert project managers for flexibility beyond these specs	2
	Training processes to ensure being ready for VC	2
	Lessons learned to adapt quickly and sustainably	3
Human-related [HR]	Communicating changes to allow them to be embraced	1
	Systematic training of required experts with specific roles	1
	Regular trainings and workshops to keep up to date	2
	Transparency of processes and responsibilities throughout	3
	Human in center of tailored digitized processes	3
Technical-related [TR]	Well-suited hardware and software	1
	Lab environment for uninterrupted testing	1
	Data consistency throughout for efficiency and automation	2
	Automation of generation and testing	2
	Tangible visualization to make digital tools approachable	3

While not an exhaustive list, these three categories cover the most important fields to incorporate digital technologies. More importantly, the given priority shows a clear roadmap to highlight where the focus should be put at what stage of the digitalization process. We argue that optimizing only one category will not yield sustainable solutions in the long run. Instead, a coordinated procedure that considers solutions prioritized in the right order from all three categories is needed.

4   Process Optimization

In this section, we present how such a coordinated procedure could look like in a concrete project scenario. We propose that the typical process can be optimized in the following ways.

4.1   Process Integration

First, digitalization is more than just a tool or a plug-and-play solution. It is a process that needs to be integrated thoroughly into existing structures. In order to accomplish this, qualified team leaders have to supervise this digital transformation. Managerial support has to be ensured to allow for a smooth transition. Following the priorities from Table 2, it is crucial to approach this step by step. While all-encompassing solutions might appear tempting, they rarely work. Incremental improvements, on the other hand, can take effect immediately and serve as building blocks for the next phases.

4.2   Educating and Training Employees

After integrating digitalization into the process, all involved departments have to be informed and trained to fully implement the techniques and obtain their benefits. This requires creating suitable training methodologies, training teaching personnel, and ensuring that the employees are willing and capable to adopt the new methods.

The focus to help the project and its team members should not be lost. Correct training is essential to assign the correct people to a given project. The documentation for this can be as simple as an excel sheet, containing requirements such as *PLC planning and programming experience* or *knowledge in the employed simulation tools.* On a less technical level, topics such as an understanding of the company's processes and the economical aspect are required. This will allow the assigned experts to competently decide when digitalization can be useful and when it can be too costly.

Not all trainings can be carried out in the same way. A suitable method for ensuring effectiveness has to be devised, but structuring and tracking trainings and required competencies is an important first step.

4.3   Technical Changes

Besides the comparatively soft process-related and human-related factors, it is also necessary to find the right technical tools. There are many different software tools available, making a careful assessment of their strengths and weaknesses necessary. A benchmark has to be devised with a fitting reference machine and suitable criteria. The results have to be representative, so they can be translated and extrapolated to upcoming projects. Ensuring that the basic technical framework runs smoothly can then lead to increased and successful automation.

4.4   Use Cases

To make the proposed solutions tangible and how they can be implemented, this section presents two concrete use cases from production planning. One of the most straightforward approaches for implementing process related solutions is to develop suitable tracking documents. This can range from simple tables to complexly linked and automated documents that provide an overview of the

entire project at any given point. As a first step, it is sufficient to create and maintain a tracking document in MS Word or Excel to track subprocesses and the respective progress. To ensure acceptance, it is important to start such processes small and grow the gradually over time.

Another approach is to implement training processes to ensure the necessary knowhow is present in every project. For this to succeed, expert knowledge is essential in order to break the field down into teachable parts. Hiring experts to create and give occasional training seminars does not add significant cost, bust allows to gradually accustom employees to the new workflows.

Hardware and licenses are easy to buy, but setting up a business plan for this endeavor should be given sufficient attention as well.

5   Conclusion and Outlook

This paper investigated, why tools such as virtual commissioning are still lagging behind other digital technologies. We identified the main problems as lack of synchronization, lack of flexibility in current processes, the human factor, and challenges for management. We categorized potential solutions into three groups and outline how a process can be optimized to facilitate digitalization, namely by a complete process integration, training for employees, and by a thorough assessment of potential technical tools.

While we believe these measures are a first step towards establishing digitalization, several limitations remain. Creating expert knowledge and changing team structures requires significant time and resources. Even with such a commitment, the right people have to be found, which is a challenge with a general global shortage of technical personnel. A technical benchmark requires extensive documentation and continuous updates to remain relevant. Finally, the determined use cases have to match actual business cases. While digital twins can reduce costs, creating an internal team for this purpose might not always be the more efficient that relying on external resources.

With this work laying the foundation, there still remains work to be done. The next step should be to continue gathering expert information by launching a wide range of interviews to get a broad view of how industry experts see the processes, what problems they identify and how they judge the proposed solutions. Additionally, the proposed solutions have to be implemented in practice to obtain real-life feedback, so that the proposed concept can be further improved iteratively. In summary, this work identified the main problems and proposed first steps towards solving them, serving as a starting point towards fully establishing the smart digital factory.

References

1. Kleimann, P., Kalmbach, R., Bernhart, W., Hoffmann, M.: Studie automotive landscape 2025: opportunities and challenges ahead. https://circabc.europa. eu/sd/a/197115bc-e691-4abd-a6a0-d7e5c9c20f45/Roland_Berger_Automotive_ Landscape_2025_E_20110228_lang.pdf/. Accessed 08 Jun 2022

2. Liebrecht, C.: Entscheidungsunterstützung für den Industrie 4.0-Methodeneinsatz–Strukturierung, Bewertung und Ableitung von Implementierungsreihenfolgen. PhD thesis, Karlsruher Institut für Technologie (KIT) (2020)

3. de Oliveira Hansen, J.P., da Silva, E.R., Bilberg, A., Bro, C.: Design and development of automation equipment based on digital twins and virtual commissioning. In: Procedia CIRP, vol. 104, pp. 1167–1172 (2021)

4. Weber, K.H.: Inbetriebnahme verfahrenstechnischer anlagen. Springer Vieweg **2**, 650–655 (1992)

5. Zäh, M.F., Wünsch, G., Hensel, T., Lindworsky, A.: Nutzen der virtuellen inbetriebnahme: Ein experiment. Zeitschrift für wirtschaftlichen Fabrikbetrieb **101**(10), 595–599 (2006)

6. VDI. VDI-Richtlinie 2206: Entwicklungsmethodik für mechatronische Systeme. Verein Deutscher Ingenieure e.V. (2004)

7. Wünsch, G.: methoden für die virtuelle inbetriebnahme automatisierter produktionssysteme. iwb TUM (2007)

8. Lechler, T., Fischer, E., Metzner, M., Mayr, A., Franke, J.: Virtual commissioning - scientific review and exploratory use cases in advanced production systems. In: Procedia CIRP, vol. 81, pp. 1125–1130 (2019). ISSN 2212–8271. 52nd CIRP Conference on Manufacturing Systems (CMS), Ljubljana, Slovenia, June 12–14 (2019)

9. Spitzweg, M.: Methode und Konzept für den Einsatz eines physikalischen Modells in der Entwicklung von Produktionsanlagen. Utz Verlag (2009)

10. Barbieri, G., et al.: A virtual commissioning based methodology to integrate digital twins into manufacturing systems (2021). https://doi.org/10.1007/s11740-021-01037-3

11. Leng, J., Wang, D., Shen, W., Li, X., Liu, Q., Chen, X.: Digital twins-based smart manufacturing system design in industry 4.0: a review. J. Manuf. Syst. **60**, 119–137 (2021). https://doi.org/10.1016/j.jmsy.2021.05.011

12. Vdma 2021 leitfaden virtuelle inbetriebnahme: Handlungsempfehlungen zum wirtschaftlichen einstieg. https://www.vdma.org/c/document_library/get_file?uuid=66118d8d-68fc-55a7-a25b-a90ffe89f1f2&groupId=34570

13. Paulsen, H., Inkermann, D., Zorn, V., Reining, N., Vietor, T., Kauffeld, S.: Produktentwicklung in der digitalisierten welt-virtuelle inbetriebnahme aus prozessbezogener, methodisch-technischer sowie personeller sicht. Arbeit interdisziplinär: analysieren-bewerten-gestalten. Beitrag C, vol. 6 (2019)

A New IT Architecture for the Digital Factory

Christopher Lorch[1(✉)] and Bernd Lüdemann-Ravit[2(✉)]

[1] Mercedes-Benz Operations Information Technology, 71059 Sindelfingen, Germany
christopher.lorch@mercedes-benz.com
[2] Institute for Production and Informatics, University of Applied Science Kempten, 87527 Sonthofen, Germany
bernd.luedemann-ravit@hs-kempten.de

Abstract. In general, current systems for the Digital Factory implement a product-process-resource (PPR) data model in a monolithic rich-client/server architecture with a single database persistence layer. Common data objects are the product bills of material, descriptions of the production processes, or the resource structure, e.g. bill of equipment. The main drawback of the current monolithic architecture is the slow rate of development, which prevents fast adoption of the software to the new production planning process (i.e., due to new technologies for the transformation of the automotive industry with the goal of electrification) is not possible. Furthermore, time-consuming and error-prone export-import operations characterize the collaboration of the engineering supply chain. Mercedes-Benz has created a new IT system architecture for their Digital Factory. The core idea of this architecture is a module-based approach. Each planning step has its own module, e.g. product analysis, layout planning or cost calculation. One single module consists of a server-based business logic, a web-based user interface and its own database. Each module is the source of master data objects that originate from the corresponding planning step and refers to data objects from predecessor planning steps. The single modules communicate mostly via KAFKA. The usage of a model based application engine allows the fast creation of different modules. Best-of-breed third-party systems for specific planning steps can be integrated into the system architecture. Web technologies allow suppliers to access the Mercedes-Benz systems directly for a fully integrated supplier collaboration. Roll-out has started and has already led to significant efficiencies.

**Keywords:** Digital Factory · Production Planning · IT Architecture

1 Introduction

1.1 Digital Factory

The term Digital Factory characterizes systems, data models, methods and business processes during the production planning [1]. The term is used for more than twenty years and received new attention within the paradigm of industry 4.0 as it is the origin of the digital twin of the production.

1.2 Data Model PPR – Product, Process, Resource

Data models in a Digital Factory consist of three main abstract data structures (business objects) [1, 2]: product, process and resource (PPR) objects. First, the product structure contains the outcome of the factory, e.g. described by the bill of material (BoM) including fasteners. The author of the product structure is the research and development department. The production planning uses the BoM in read-only mode. Second, the process structure represents the operations that are necessary to build the product including estimated and analyzed time attributes. For example, it contains processes to assemble, weld or glue product parts. Third, the resource structure contains all equipment inside the factory: e.g. robots, tools, carriers, building and humans. Both process and resource structure are created by the production planning (PP), in parallel to the product development process (simultaneous engineering). The PP also interconnects the business objects (product, process and resource) with predefined relations according to their usage to define which product is processed inside the factory with the help of the corresponding resources (see Fig. 1).
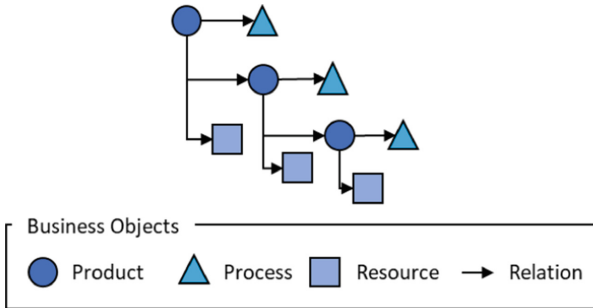
Fig. 1. Sample Data Model "Product Process Resource (PPR)" including relations to represent the Digital Factory

1.3 Scope

The scope of the current development activities of the new IT architecture focuses on body-in-white planning in an early stage starting from the preliminary planning and ending at the placing of the orders to the prime contractors for the automation equipment. The idea of this new IT architecture can also be used in other planning departments such as press shop, final assembly and intralogistics. Validation topics are out of scope in the current stage of the system design because they require extensive 3D simulation functionality. 3D visualization and modification functionality offered by this system is therefore not enough.

1.4 Main Requirements

The chance to build a new Digital Factory is also a challenge, because the corresponding IT architecture has to be suited for the requirements and upcoming ideas of industry 4.0 [3].
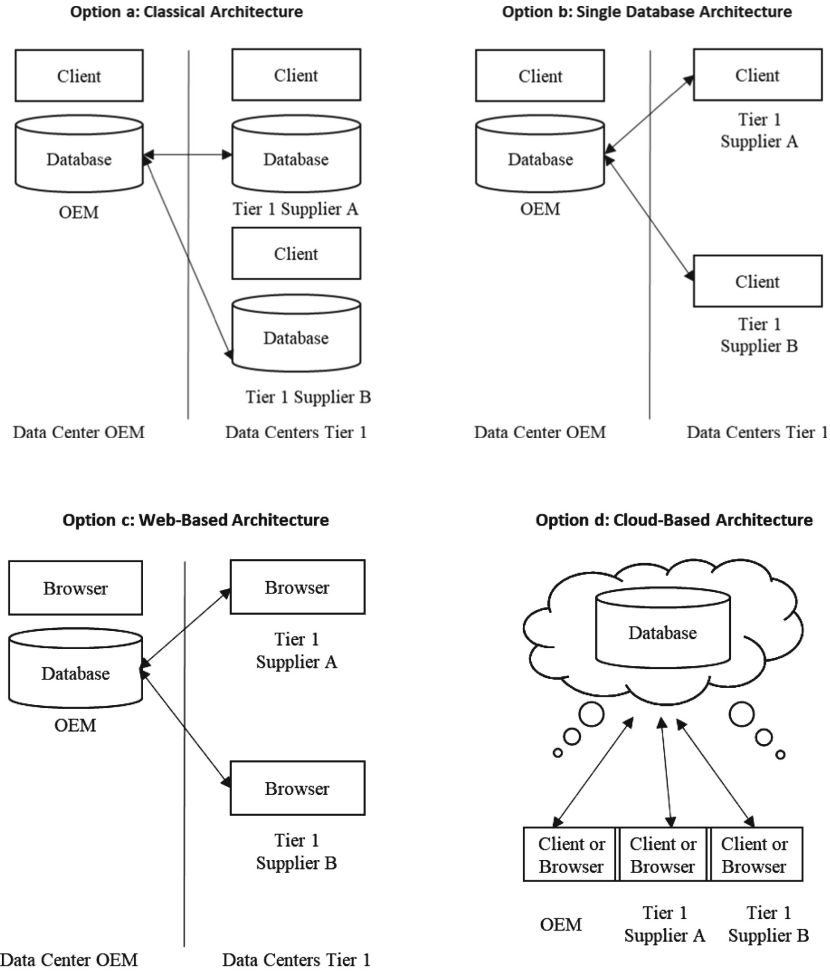
Requirements Due to Supplier Collaboration. Nowadays, original equipment manufacturers (OEMs) and their suppliers (Tier 1, 2, 3…) collaborate along the complete production planning process [2] until the automation equipment is ready to produce. This leads to the fact that supplier integration is one of the key aspects to establish efficient planning processes. The IT architects focused therefore on four approaches (Fig. 2).

Potential Architectures. The first option is the classical monolithic architecture offered commonly by standard software: an installation with one central database in the data center of the OEM and corresponding installations in the data centers of the suppliers (Fig. 2a). Most OEMs still use this approach for their Digital Factory. Thereby manual effort and complex interfaces are necessary to update and exchange product and planning data between the OEMs and their suppliers, since the product is still in development. Additionally, the IT departments of all participants have to synchronize the same database scheme, which is a challenging organizational task. Some software vendors propose a second approach (Fig. 2b): a supplier client accesses the OEM database.

A web-based (Fig. 2c) and a cloud-based architecture (Fig. 2d) promise a more efficient approach regarding applicability and operation.

Requirements Due to the Speed of Change. An additional and essential requirement to the IT architecture is a fast adaption and an easy extension of the Digital Factory systems according to changes in the planning process. The transformation of the car industry due to digitalization and electro-mobility will cause changes in the current planning process that no one can predict today. Adaptions and extensions of the standard Digital Factory software that take several years on the roadmap of a software vendor will significantly slow down the speed of adaption of the Digital Factory IT landscape and the support of agile production planning processes. Therefore, service-orientationed and modular organization of the Digital Factory system are also key requirements.

The implication of the new IT architecture to fulfill extensibility and modularization is the concept of a distributed data model (Fig. 3). What is the key idea of this requirement? The complete business process consists of single processes with a corresponding result out of each process, e.g. the production planning consists of a process "plan resources". Its result is the recourse bill of material. The IT architecture itself must have one module for each process with its own GUI and its own master data. One module is referring to the data of another module via a reference object. The architecture then needs mechanisms to exchange data between the modules. For example, the resource module to plan the resources contains a data model with reference objects to the product data model of the module to analyze the product (Fig. 3). The resource module stores the business objects that the planner has created within the process "plan resources" (master data resource objects).

Fig. 2. a + b. Classical and Single Database IT architecture for the supplier collaboration. **c + d.** Web- and Cloud-Bases IT architecture for the supplier collaboration

Requirements Stemming from Legacy Environment. An additional challenge is the integration of the software into the existing legacy IT landscape (e.g. product documentation management (PDM) systems) or IT guidelines (e.g. identity access management infrastructure). Production planners need up-to-date product data. Therefore, an interface must provide hundreds of car configurations from the PDM system to the Digital Factory. The Digital Factory must be capable to provide or access this data and enrich downstream processes efficiently with production planning information.
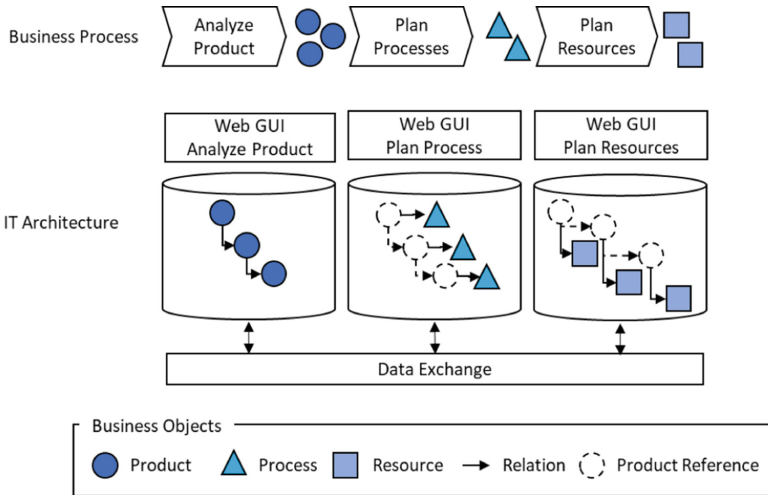
Fig. 3. Extensibility and modular organization of the system

1.5 State of the Art

An overview of commercial and scientific software tools for Digital Factories can be found in [4]. The market leaders [5, 6] started implementing either web-based architectures [7] or cloud-based architectures [8–10]. Nevertheless, all of them have one common characteristic in their architecture: the central database (Fig. 2a). In scenarios with suppliers one can find own central databases for the suppliers (Fig. 2b).

The Aras platform [11] and the ASCon Application Engine [12] have first implemented the idea of modularization of their Digital Factory solution. Both of them need additional 3D visualization, e.g. [13–15], and authoring functionality, e.g. [13].

Additionally, current IT landscapes for the domain of PLM (product lifecycle management) are dominated by few monolithic systems [21, 22]. New approaches for an IT architecture is necessary to implement the concept of a digital twin of the production [23–25].

2 System Design

2.1 General Approach

Advantages. We have chosen a modularized system design approach in order to fulfill the requirements listed above, using the ASCon Application Engine [12]. By this architecture we benefit from the potential of developing several modules in parallel and that we can focus on the use-cases most urgently needed without the necessity to implement all aspects which are needed in the implementation (or customization) of a monolithic system architecture. However, the open architecture approach is a sound basis for extending the ecosystem in order to match future use-cases like the realization of a digital twin

of the plant. Furthermore, from scratch design of the application allows a performant implementation of the product structure data model.

Disadvantages. Nevertheless, there are also drawbacks then creating such a modularized system architecture from scratch. On the one hand, the design of a shared data model and the modeling of the interactions between different applications are crucially important. On the other hand, this design makes it more difficult for the product owners of the individual modules to implement the correct features.

Web-Based Approach. The majority of the applications are web-based, which has several advantages. First, no sophisticated rich client installation on the end-user devices is necessary. The hardware requirements of the latter are lower for web-clients than for rich clients offering 3D modification features. In addition, less data needs to be transferred from the application server to the client if almost no processing is done on the client tier.

All of the above is very helpful to improve the collaboration with suppliers. OEM ask suppliers to work on the OEM's infrastructure (Fig. 2c, b). This shortens the overall planning time since no exchange of large data sets, complex export and import mechanisms and distribution of application software is needed. Furthermore, the suppliers do not need to purchase specific software or licenses. These advantages outnumber in the long run, the enlarged investment of the OEM in infrastructure due to a higher number of users.

Table 1. Overview of implemented modules.

| Application Name	Description of the application
*DiFa*Project	Central project structure; user management; rights and roles;
*DiFa*Library	Life cycle and information of single equipments; definitions of equipment modules
*DiFa*Cost	Costs of the different equipments; overhead cost
*DiFa*Product	Analysis of the product and fasteners; comparison of different product configurations
*DiFa*Planning	Planning of the mechanical part of the resource structure incl. Layout
*DiFa*Controls	Planning of the electrical part of the resource structure

2.2 Interface Design

In general, the architecture uses two different technologies for the exchange of information between modules.

Streaming. Firstly, KAFKA [16] is the technology to implement use cases with an asynchronous communication based on an event-driven approach. This technology has several benefits for our modularized architecture. Applications can release events and

any number of applications can consume the information from the events. This makes the scenario easily extendable for the implementation of further use-cases, since the application providing the information does not need to be changed. Furthermore, different applications in the downstream processes can directly use the very same information provided by one source.

An example for a use case implemented via KAFKA is the release of a new equipment in the *DiFa*Library (Fig. 4b). An event containing the equipment information triggers a new calculation task in *DiFa*Cost to plan costs. Additionally, the central planning application, *DiFa*Planning consumes the message and provides the new equipment for resource planning. In the latter use case, *DiFa*Planning also automatically checks if this new equipment replaces an older one, and, in this case, provides the information to the users so that they can update their resource planning with the new version of the equipment. Future use cases which could be implemented directly in the event of a new released equipment could be either the trigger for the purchasing department to start negations with the supplier or the contract department to start with the detailed design phase of the new equipment.
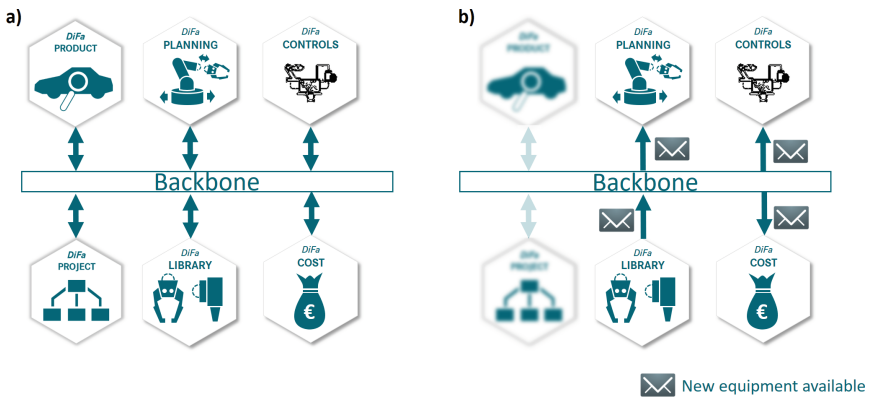
Fig. 4. a) Schematic overview of the six main applications generated in the first iteration of the implementation. Table 1 summarizes the idea behind the single applications. **b)** Example flow of information shared by *DiFa*Library, which is processed by several other applications

REST APIs. On the other hand, additional use cases were implemented using Representational State Transfer (REST) APIs [17, 18] for synchronous communication needs. Some data is just too big or changes happen too frequently for an efficient use of the KAFKA bus. A good example for this is the simultaneous mechanical and electrical resource planning. Traditionally, different user groups work in these two domains, however they are mutually dependent. For this use case both planning teams work on the same resource structure, using also the same equipment library, but with different optimized tools for the requirements of the individual user group.

2.3   Example Application

One typical and rather complex application is the resource planning application *DiFa***Planning**. We will highlight the most interesting parts. We created *DiFa***Planning** on a web software framework provided by the company ASCon Systems [12] and the 3D rendering service Uber provided by the company Netallied [13].

This application is used to create and modify the resource structure of the production plants and to provide a 3D layout of the production lines. Figure 5 shows a screenshot of a typical scenario. On the left, a tree structure of the resources is visible to navigate to the desired planning area and to list the planned resource bill of material. The middle section contains a view of reference objects to the equipment library originating in *DiFa***Library**. The window on the right contains a 3D representation of the current planning scenario. The 3D scene is rendered on the server side and the information is shared with the user via WebSockets. Information is exchanged between the web framework and the 3D scene using REST APIs.
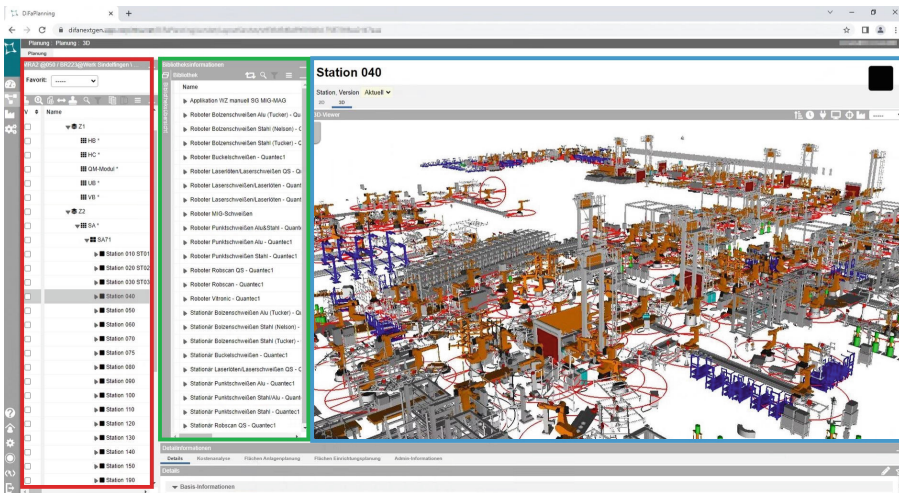
Fig. 5. Screenshot of *DiFa***Planning**. On the left hand side (marked with red), one can see the tree representation of the resource structure. The middle column (marked with green) shows only equipment that can be used for planning out of *DiFa***Library.** On the right (marked with blue), 3D planning of the resources is shown.

3   Outlook

The implementation of the Digital Factory of Mercedes-Benz is by no means done. With the architecture described above, the fundamentals for the implementation of future use cases is set.

Additional Applications.   One could think in various directions for the enhancement and expansion of the application landscape. So far, the three branches of the PPR are not

completely linked. Even though this results in some liberties for the users, a closer link between the branches will help to analyze necessary changes in the process or resource structure that lead to changes of the product structure. In addition, this is necessary when implementing an approach to create automatically the process and resource structures from a given product structure [19].

Integrated Change Management. Another interesting aspect to follow is a fully integrated change management over the whole planning and production cycle. This would be the basis for creating a full digital twin of the production facilities. In order to realize this, a complex branching concept of the different structures is necessary which allows, on the one hand, to see the evaluation in time, and, on the other hand, different planning alternatives representing different possibilities to realize the planning.

Online Validation. During the collaboration process, Mercedes-Benz requests usually different deliverables from the engineering partners. The integration of online validation tools in the architecture will further improve the data quality and reduce the realization time for new production facilities.

4   Conclusion

Implementation Speed. After the successful go live of the first applications, we have seen that the chosen approach is fruitful and sustainable. The implementation was approximately two times faster compared to the approach of customizing a commercial software solution to fit the OEM-specific requirements. We achieved this faster development speed mainly because we could develop several modules in parallel and we have implemented or customized only parts necessary for our processes. By setting up the already exciting application engine ASCon Map [12] we have benefited from the already available functionality, such as user management, monitoring, logging and default user control features.

Performance. A great benefit of the architecture is the enormous speed of loading and visualization of 3D data. For a fully configured product structure, both the loading of the product structure and the visualization takes approximately 30 s, thanks to the visualization engine Uber [13]. In comparison, loading and visualization took usually around 30 min within the former Digital Factory system of Mercedes-Benz.

Additionally, we achieved the requirement of loading hundreds of car configurations from the PDM system into the Digital Factory with this architecture approach.

Integration with Existing Applications. We were able to add further, earlier excised, applications to the ecosystem and we are very positive that additional applications will be part of the Digital Factory in the future.

References

1. Digital Factory – Fundamentals, VDI 4499 Part 1, Verein Deutscher Ingenieure e.V. 8, Beuth, Berlin (2008)
2. Bracht, U., Geckler, D., Wenzel, S.: Digitale Fabrik, 2nd edn. Springer Vieweg, Berlin (2018)
3. Duda, J., Oleszek, S., Santarek, K.: Product lifecycle management (PLM) in the context of industry 4.0. In: Trojanowska, J., Kujawińska, A., Machado, J., Pavlenko, I. (eds.) MANU-FACTURING 2022. LNME, pp. 171–185. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-99310-8_14
4. Chen, D., Kjellberg, T., von Euler, A.: Software tools for the digital factory – an evaluation and discussion. In: Huang, G.Q., Mak, K.L., Maropoulos, P.G. (eds.) Proceedings of the 6th CIRP-Sponsored International Conference on Digital Enterprise Technology. Advances in Intelligent and Soft Computing, vol. 66. Springer, Berlin (2010). https://doi.org/10.1007/978-3-642-10430-5_62
5. Top 10 PLM and Engineering Software Vendors, Market Size and Market Forecast 2020–2025, https://www.appsruntheworld.com/top-10-product-lifecycle-management-engineering-software-vendors-and-market-forecast. Accessed 27 May 2022
6. Product Lifecycle Management (PLM) Software Market - Groth, Trends, COVID-19 Impact, and Forecasts (2022–2027). https://www.mordorintelligence.com/industry-reports/product-lifecycle-management-software-market. Accessed 27 May 2022
7. Active Workspace | Siemens Software. https://www.plm.automation.siemens.com/global/en/products/collaboration/active-workspace.html. Accessed 27 May 2022
8. Liu, L.: Application of Dassault system 3D experience platform in enterprise digitalization. In: Hung, J.C., Yen, N.Y., Chang, J.W. (eds.) Frontier Computing. FC (2021)
9. Sharma, S., Segonds, F., Maranzana, N., Chasset, D., Frerebeau, V.: Towards cloud based collaborative design – analysis in digital PLM environment. In: Chiabert, P., Bouras, A., Noël, F., Ríos, J. (eds.) PLM 2018. IAICT, vol. 540, pp. 261–270. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01614-2_24
10. Niemann, J., Pisla, A.: Teamcenter data management. In: Life-Cycle Management of Machines and Mechanisms. MMS, vol. 90, pp. 381–396. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-56449-0_19
11. Pfenning, M.: Breaking Down the Silos - How Systems Architecture Enables Interdisciplinary Collaboration. Aras Corporation, White Paper (2020)
12. ASCon Digital Twin Product Suite - Application Engine. https://ascon-systems.de/product-suite/application-engine. Accessed 27 May 2022
13. NetAllied Systems GmbH.https://netallied.de/. Accessed 17 May 2022
14. b.ReX GmbH. https://b-rex.de. Accessed 17 May 2022
15. Kisters AG.https://viewer.kisters.de, Accessed 17 May 2022
16. Shapira, G.: Kafka - The Definitive Guide: Real-Time Data And Stream Processing at Scale, 2nd edn., O´Reilly, Sebastopol (2021)
17. Fielding, R.: Architectural Styles and the Design of Network-Based Software Architectures. University of California, Irvine (2000)
18. Wikipedia. https://en.wikipedia.org/wiki/Representational_state_transfer, Accessed 17 May 2022
19. Hagemann, S., Stark, R.,An optimal algorithm for the robotic assembly system design problem: an industrial case study. CIRP J. Manuf. Sci. Technol. **31,** 500–513, Elsevier (2022)
20. Stark, R.: Virtual Product Creation in Industry, 1st edn. Springer, Berlin, Heidelberg (2022)

21. Hooshmand, Y., Resch, J., Wischnewski, P., Patil, P: From a monolithic PLM landscape to a federated domain and data mesh. In: Proceedings of the Design Society, vol. 2, pp. 713–722 (2022)
22. Bleisinger, O., et al.: Killing the PLM Monolith-the Emergence of Cloud-Native System Lifecycle Management (SysLM) (2022)
23. Aheleroff, S., Xu, X., Zhong, R., Lu, Y.: Digital twin as a service (DTaaS) in industry 4.0: an architecture reference model. Adv. Eng. Informat. **47** (2021)
24. Talkhestani, A., et al.: An architecture of an Intelligent digital twin in a cyber-physical production system. At - Automatisierungstechnik **67**(9), 762–782 (2019)
25. Eric Harper, K., Ganz, C., Malakuti, S.: Digital twin architecture and standards, IIC J. Innov. **12**, 72–83 (2019)

Security Analysis of a Blockchain Based Data Collection Method for Cross Company Information Sharing

Tobias Bux[(✉)], Oliver Riedel, and Armin Lechler

Institute for Control Engineering of Machine Tools and Manufacturing Units,
University of Stuttgart, Seidenstr. 36, 70174 Stuttgart, Germany
`tobias.bux@isw.uni-stuttgart.de`

Abstract. Digitization within medium-sized enterprises advanced in the last years. Collecting and analyzing data for optimizing internal production processes therefor is the current state of many companies. The next step of digitization is using this collected data not only for internal processes but for cross company business models along the value network. This step brings new requirements for how data is collected, stored and shared. In this paper those requirements are listed and explained. Afterwards, an implemented solution for data collection fulfilling the requirements is analyzed. The focus of the analysis lies on security issues within the data flow between data creation and cross-company usage. Therefore, the timespan between data creation on a sensor, processing the data within local IT-systems and reliably storing data within a blockchain is considered. A threat modeling approach considering attack vectors along the described data flow is used to quantitatively compare the proposed solution to regular industrial solutions. The analysis will highlight the differences of the compared solutions on different topics like data integrity and immutability. Lastly, an outlook on industrial usage of the analyzed solution is given.

Keywords: Blockchain · Data Integrity · Cross-Company Data Exchange

1 Introduction

Several IT systems are usually involved in the collection and processing of industrial data. These systems are located in differently secured networks, have different operating systems and each use their own software. It therefore becomes more difficult to make statements about data security in industrial data acquisition the more systems are involved. For in-house data usage, the problem is limited because the standard data acquisition process, shown in Fig. 1, is usually covered by the security policies of the company's IT department [1]. The requirements for internal company use of collected data are usually met here. However, if external partners within the value network are also involved in the processing of the data, the requirements change. In addition to data security, data integrity plays a role in this use case because there is no inherent trust between the companies. This means that when data is exchanged, not only must security

be guaranteed, but it must also be ensured that the data to be exchanged is unchanged and legitimate. Accordingly, the threat analysis in this paper does not focus exclusively on data and transport security, but also on the risks for preserving integrity.
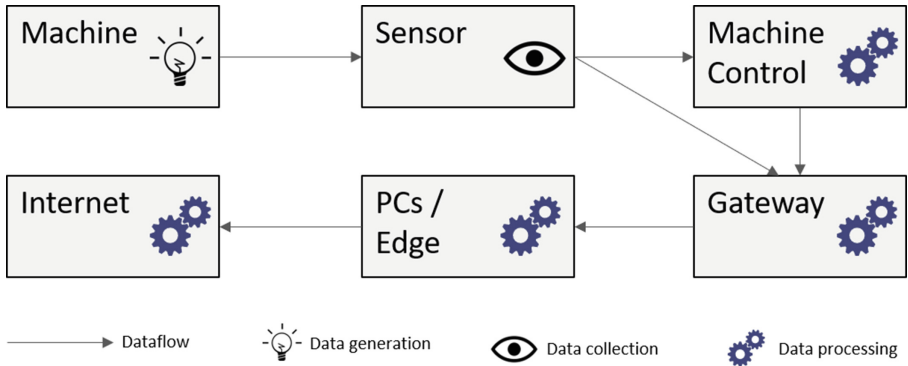
Fig. 1. Common Data Flow within a production environment [1]

2 Related Work

Before discussing the results of this work, the state of the art is considered in two basic areas. Firstly, the solutions that currently exist for cross-company data exchange and how their implementation is structured for security and integrity will be presented. Then, threat modeling techniques for industrial data exchange are examined.

2.1 Cross Company Data Exchange

A cross-company data exchange has requirements for the data to be exchanged that do not exist in an internal company environment. This is the conclusion reached by Uygun [2], who has therefore defined these requirements in more detail. In addition to structural and organizational requirements, which are not considered in this paper, he has established technical requirements. These include data security, communication security and consistency of data. A multi-agent tool was developed for validation, which was used by over 40 companies to verify the statements of Uygun. However, a model-based analysis of the technical requirements was not performed.

Ruf et al. [3] did exactly this. They looked at a framework for industrial data exchange called KOSMoS [4] and analyzed its security and integrity using threat model analysis. The considered framework uses a blockchain-based data exchange to guarantee the integrity of data. However, the results of Ruf et al. show that even in this environment, risks such as faulty smart contracts or incorrect interface usage must be considered [3]. In addition, communication between the data origin and the blockchain needs to be critically considered. The threat analysis by Ruf et al. mentions the risks of man-in-the-middle, identity spoofing and inside attacker [3].

The risks described by Ruf et al., were considered in more detail in a paper by Korb et al. [1] and a solution was proposed. The basic idea behind Korb et al.'s solution is to create and sign blockchain transactions close to the sensor. This way, data can be verified before it arrives in a Blockchain. The entry of false data, through manipulation or spoofing, into a blockchain is thus prevented. The implementation of Korb et al. is realized by using an ESP32 microcontroller that is directly linked to the data generating sensor via GPIO connection. The data is signed on the microcontroller before it passes through larger technical systems of internal IT. Although this solution offers increased security, the use of the Ethereum blockchain is not optimal for an industrial deployment, as quantitatively demonstrated by Polge et al. [5].

In summary, it can be stated that new requirements must be observed in the case of cross-company data exchange. There are already solutions for meeting the technical requirements, but they are not suitable for industrial use. Bux et al. have therefore developed a new technical solution based on Hyperledger Fabric [6], which is used as the basis for the threat analysis in this paper.

2.2   Threat Modelling for Production

There are several ways to analyze and assess the threats in an IT system. Shevchenko et al. [7] compared twelve currently used models in a report for the Defense Technical Information Center. As a result of their work, features were assigned to each model to help select an appropriate model. An overview of the four most common models can be seen in Table 1. According to Shevchenko, STRIDE [8] is the most widely used and mature model. In the described work by Ruf et al. [3], they also performed a STRIDE based risk analysis. Since STRIDE is the leading threat model, software tools were created for STRIDE to assist in performing the analysis. As STRIDE was invented by Microsoft, they offer software built around STRIDE called Thread Modelling Tool [9] which can be used within Microsoft Azure. However, besides paid software, there are also free alternatives. Threat Dragon, for example, is an open-source tool under Apache 2.0 license that is constantly maintained [10] and implemented STRIDE functionality.

To ensure the most representative analysis of the architecture evaluated in this thesis, STRIDE is used as a model for risk analysis. Threat Dragon is used to perform the analysis and graphically record the results.

Table 1. Threat modelling methods features based on Shevchenko et al. [8]

Threat Modeling Method	Features
STRIDE	• Helps identify mitigating techniques • Most mature • Easy to use but time consuming
PASTA	• Contributes to risk management • Rich documentation • Stakeholder collaboration intended • […]
LINDDUN	• Helps identify mitigating techniques • Built in prioritization of threat mitigation
CVSS	• Consistent results when repeated • Automated components • Not transparent score calculations

3 Proposed Solution Architecture

To understand the threat analysis, it is first necessary to present an overview of the software system under consideration. For this purpose, a UML-based architecture is presented in Fig. 2. A detailed description of the solution, an explanation of individual components, and the exact data flow within the solution are explained in detail by Bux et al. in [6]. The flow of this architecture is explained below:

Four hardware components exist in the architecture under study. A sensor (1) is seen as a data generation component and is directly linked via GPIO to the second hardware component, a microcontroller (2). On the microcontroller C++ code is executed, which is used for an information exchange with an industrial PC (3). Within this IPC, different services of the Hyperledger Fabric architecture communicate. A single service, called Client (Proxy) in Fig. 2, forms the interface between the local Hyperledger Fabric services and the microcontroller logic. This structure ensures that data collected by the sensor is converted into a Hyperledger Fabric compliant transaction under the supervision of the microcontroller. The transaction is signed on the microcontroller and then sent by the Committing Peer to a cloud instance (4) running a Hyperledger Fabric node. The data is stored there and can then be queried via REST interface.
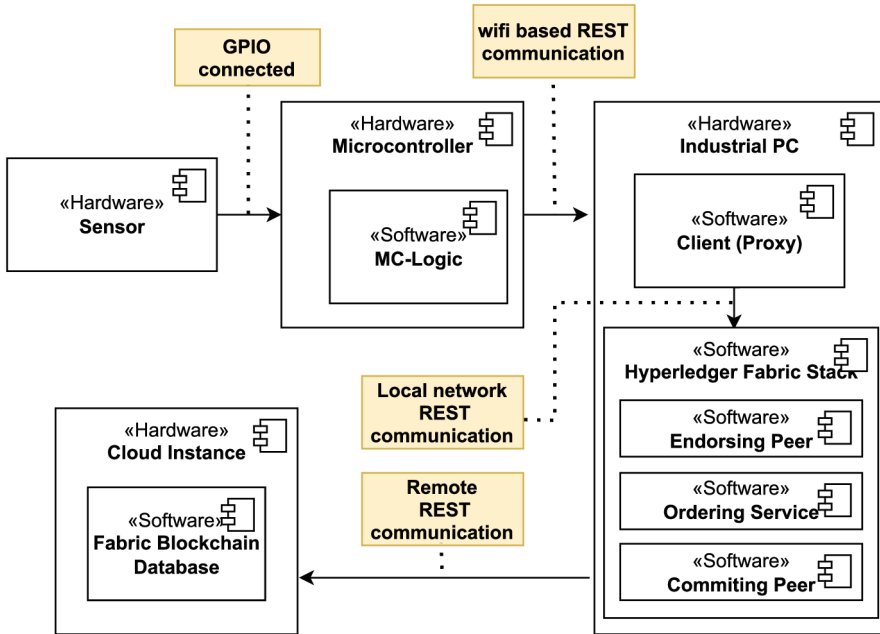
**Fig. 2.** UML overview on the solution that is analyzed within this paper [6]

4   System Analysis

At the beginning of the analysis, some assumptions were made about the system state. The sensor, which provides data, functions without restrictions. According to its definition, the blockchain is an immutable storage. Furthermore, it is assumed that no faulty functionality is contained in the third-party services used, such as the software development kit for Hyperledger Fabric. With these assumptions, a STRIDE analysis was performed of the system described above. A visual overview of the results is displayed in Fig. 3. Meanings of the different colors, shapes and connectors are according to STRIDE standards.

During the analysis, different threats have been found. These are summarized and explained in the following:

Information Disclosure: The data flow between the components in Fig. 3 contains confidential data. All communication is therefore encrypted. The communication between sensor and microcontroller is an exception since the transmission of digital or analog sensor data is realized directly via GPIO. This security gap can be solved by sealing the sensor and the microcontroller into a Blackbox.

Denial of Service (DoS): In the architecture under consideration, several services exist that are vulnerable to a DoS attack. The largest hardware component is the microcontroller. If this component no longer functions, the entire system is disabled. This is
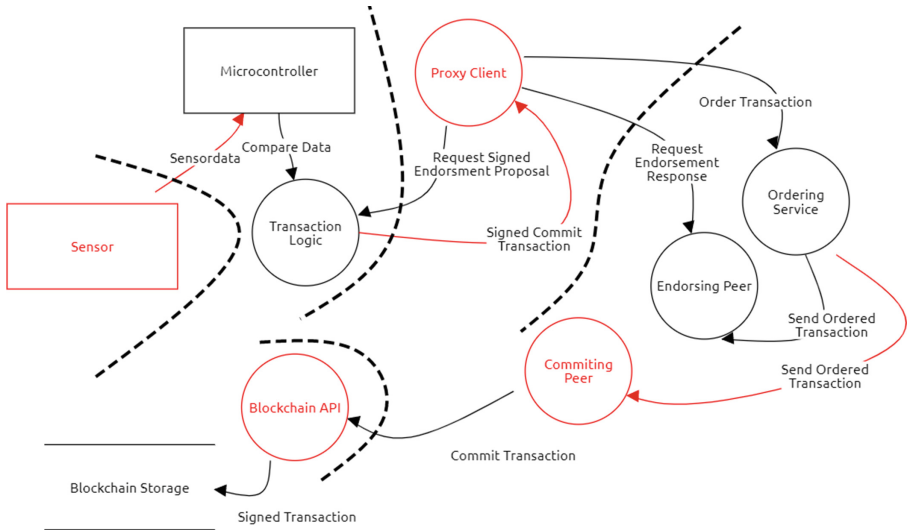
Fig. 3. Threats to using the proposed architecture for blockchain based data storage

possible, for example, by exhausting the CPU or completely filling the internal memory. However, since no write access to the microcontroller is allowed in the proposed solution, this threat is not considered further.

The proxy client, however, is much more susceptible to a DoS attack. It runs on an IPC with a conventional operating system and has a connection to the Internet. Accordingly, it depends on the security of the respective company IT whether the proxy client is sufficiently protected. Redundant systems and life-cycle management can minimize a service outage due to DoS attacks in this case.

The last component to be protected against a DoS attack is the blockchain interface. For this component, too, protection depends heavily on the respective implementation. It is worth mentioning that a temporary DoS attack on the blockchain interface has no impact, as the signed transactions are kept locally until they arrive in the blockchain storage.

Tampering: Using a signed blockchain transaction is a direct measure against tampering. Nevertheless, this transaction must first be created and signed. Here, Hyperledger Fabric has a disadvantage compared to other blockchains, such as Ethereum. Several components are involved in the signing process. These components cannot all be executed on one microcontroller. This architecture makes Hyperledger Fabric susceptible to tampering. For this reason, the protocol for signing transactions in this solution has been adapted. Each step of the protocol, described in more detail in [6], is checked by the microcontroller. In addition, the final signing process is performed directly on the microcontroller. This adaptation results in the signing process being tamper proof. However, this does not apply to the connection between the microcontroller and the

sensor. The already proposed blackboxing solution is the only approach for this, where no adjustments must be made on the sensor itself.

Escalation of Privilege: The use of a microcontroller-based solution, which is as close to the hardware as possible, ensures difficult access and modification of execution logic. However, this does not apply to the IPC on which the second part of the architecture is executed. For example, to prevent an inside attack by an employee who has gained access to the IPC, a user management and access control system is necessary. Similar to DoS prevention, threat assessment depends on the internal IT systems.

5   Findings

In the following, the threats found are assigned to different systems. The systems are technical (Infrastructure), person-related (Employee) and business model (Business) related. Each threat is classified systematically. For this purpose, the impact is defined on the one hand and the risk of the threat occurring is assessed on the other. Impact can be categorized into critical (C), medium (M) and low (L). Risk is expressed in high (H), medium (M) and low (L). In addition, common methods are listed that can be used for threat mitigation. Table 2 displays an overview of the risk assessment results.

Table 2. Risk Assessment based on Exposures

Exposure	Threat	Impact	Risk	Mitigation
Infrastructure	Malware	M	M	Malware detection
	DoS	C	M	Design, Firewall
	Information Disclosure	C	L	Encryption
Employee	Misconfiguration	C	L	Audits, Tests
	Inside Attack	C	L	Logs, Privilege management
Business	Tampering	M	L	Sealing, Transaction signature
	Man in the middle	L	L	Signature, Data buffering
	Data manipulation	C	L	Anomaly detection

5.1   Infrastructure

Malware is an often-used technique to corrupt processes or data storages. In this architecture, it is mainly the IPC that is vulnerable, since the microcontroller without a proprietary operating system is a difficult target for malware. If malware were to influence the Hyperledger Fabric components of the architecture, the integrity of the subsequent blockchain could potentially no longer be guaranteed. An up-to-date operating system and malware detection provide the necessary protection.

Denial of Service (DoS) attacks, as the name suggests, are mainly used to prevent systems from working. Exposed interfaces are usually used for this purpose. In the case of this architecture, such an interface is provided for the interaction with a blockchain. However, since the internal service only addresses a REST interface and cannot be operated itself, the risk of a DoS attack is reduced. Redundancy of systems and an up-to-date firewall can protect against this risk.

Information Disclosure is a serious threat once unencrypted data is accessible. Permanent encryption and a user-based access system is sufficient in most cases to prevent this risk.

5.2   Employee

Misconfiguration is a common error as soon as people are involved in processes. In the case of this architecture, it is necessary to configure both individual services and the communication between them. It is therefore necessary to take precautions. Unit and integration tests should be performed before rolling out the system. Finally, an audit of the running system helps to prevent incorrect configuration.

Inside attacks are very difficult to prevent. Despite rights management, access to data and services is necessary for selected personnel. The only effective protection is to know your personnel and to recognize changes in behavior. Logging systems help to detect inside attacks. Like surveillance cameras, logging systems can have a disabling effect on offenses.

5.3   Business

Avoiding **tampering** within this architecture is necessary mainly between the sensor technology and the microcontroller. Even if subsequent attacks are possible, the only effect is that individual data records are not saved or are saved late. However, the integrity of the stored data is not violated, due to the system sorting out malicious data. Meaningful protection between sensor and microcontroller can only be achieved by locking and sealing the two components.

Man in the middle attacks are not a major threat to this architecture. All data is encrypted and therefore cannot be viewed. In addition, the signature of each recipient is checked, which means that messages that have been smuggled in or changed are not accepted. In addition, local buffering of data ensures that data is not lost if it does not end up in the blockchain.

Sustained **data manipulation** is no longer possible after sensor data has entered the microcontroller. This means that either false data must be injected between sensor and microcontroller or the sensor itself must be manipulated. Both can be prevented by sealing the two components.

6   Summary

In this paper, a Hyperledger Fabric-based approach to data provisioning was tested for security and immutability. For this purpose, it was shown which gaps the proposed architecture closes. Then, based on state-of-the-art techniques, an analysis of remaining risks was performed, classified and explained.

In conclusion, it can be said that even the proposed architecture cannot exclude all threats, but the risk of occurrence is significantly lower than with commercially available solutions. The hardware-based signature of the blockchain transactions prevents manipulation of data within the downstream systems and thus offers a clear advantage over the rest of the solutions.

Nevertheless, a security gap of this approach has become clearly visible. The connection of sensor and microcontroller is the weak point of the system. However, without developing sensor specifics or having to customize each sensor, there are only physical ways to improve the connection of the two components. Using the technique of sealing both components in a Blackbox, an industrial usage is not yet given. By not using this technique, the system can be used with every sensor that can deliver its data via GPIO.

References

1. Korb, T., Michel, D., Riedel, O., Lechler, A.: Securing the data flow for blockchain technology in a production environment. IFAC-PapersOnLine **52**(10), 125–130 (2019). ISSN: 2405-8963. https://doi.org/10.1016/j.ifacol.2019.10.012
2. Uygun, Y.: A multi-agent-based communication prototype for cross-company capacity exchange in manufacturing networks, 1 April 2016. https://doi.org/10.2139/ssrn.3909338
3. Ruf, P., Stodt, J., Reich, C.: Security threats of a blockchain-based platform for industry ecosystems in the cloud. In: 2021 Fifth World Conference on Smart Trends in Systems Security and Sustainability (WorldS4), pp. 192–199 (2021). https://doi.org/10.1109/WorldS451998.2021.9514058
4. KOSMoS. https://www.kosmos-bmbf.de/kosmos-system/. Accessed 04 July 2022
5. Polge, J., Robert, J., Le Traon, Y.: Permissioned blockchain frameworks in the industry: a comparison. ICT Express **7**(2), 229–233 (2021). ISSN: 2405-9595. https://doi.org/10.1016/j.icte.2020.09.002
6. Bux, T., Riedel, O., Lechler, A.: Blockchain based approach on gathering manufacturing information focused on data integrity. In: Liewald, M., Verl, A., Bauernhansl, T., Möhring, H.C. (eds.) WGP 2022. LNPE, pp. 473–483. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-18318-8_48
7. Shevchenko, N., Chick, T.A., O'Riordan, P., Scanlon, T.P., Woody, C.: Threat modeling: a summary of available methods. Technical report 2018-07-01, Defense Technical Information Center. https://apps.dtic.mil/sti/citations/AD1084024. Accessed 04 July 2022
8. Kohnfelder, L., Garg, P.: The threats to our producs. Microsoft Security Blog. https://adam.shostack.org/microsoft/The-Threats-To-Our-Products.docx. Accessed 04 July 2022

9. Microsoft Threat Modelling Tool. https://docs.microsoft.com/de-de/azure/security/develop/threat-modeling-tool-threats. Accessed 04 July 2022
10. Threat Dragon Github. https://github.com/OWASP/threat-dragon/. Accessed 04 July 2022

Part C: Advanced Manufacturing and Sustainability

Combination of the Scanning and the Polar Measuring Method for the Defect Detection in Dry Fibre Layups Using an Eddy Current Sensor

Christopher Mimra[1,2]([✉]), Boris Eisenbart[1], Stefan Carosella[2],
and Peter Middendorf[2]

[1] Swinburne University of Technology, John Street, Hawthorn 3122, Australia
`cmimra@swin.edu.au`
[2] Institut für Flugzeugbau, Universität Stuttgart, Pfaffenwaldring 31,
70569 Stuttgart, Germany

Abstract. Gaps, fibre misalignment and foreign objects in dry fibre tape layups are detrimental towards the final mechanical properties of carbon fibre composite parts. Eddy current sensors can be used to inspect layups for the named defects. In comparison to optical inspection methods, eddy current measurements are capable of inspecting not only the top layer but multiple plies at once. In the literature, two different methods have been proposed for this inspection. One is the scanning method, where the sensor is moved over the specimen and creates a grey-scale image. The pixel brightness correlates with the layup's local electrical conductivity. The other method for the local inspection is the polar method, where the sensor is rotated on a single spot to detect the fibre orientation. In this work, both methods were conducted with the same inspection system, which is based on a collaborative robotic arm. It has been shown that the scanning method can identify gaps and foreign objects out of metal. The fibre orientation can be identified by the rotational method with a precision of $0.16°$. The exact positioning of the probe remains the main challenge for this inspection system. However, a combination of both methods promises to provide a reliable inspection method for the most common defect in a dry fibre layup.

Keywords: Eddy Current Inspection · Carbon Composites · Quality Inspection · Dry Fibre Layup

1 Introduction

Carbon composites are high-strength, low-weight materials which find increasing use in the aviation and automobile industry. Their mechanical properties can exceed those of metallic materials while being much lighter. However, even small defects, which might be within the material and invisible from the outside,

can weaken composite parts significantly. Therefore, there is a need for reliable inspection methods that can detect defects within the material without destroying the specimen. Eddy current sensors have been proven to be able to detect various defects in composites. This work proposes an integration of such a sensor in an inspection system based on a 6-axis robotic arm. This setup allows for performing two different inspection methods: the polar method, where the eddy current probe inspects a single position while rotating, and the scanning method, where the sensor scans the whole specimen. The research question is: *Can the polar and the scanning eddy current inspection methods be combined in a single system to detect the most common defects in dry fibre layups?* Furthermore, the challenges during the implementation are discussed in this paper.

## 2   Literature Review

2.1   Carbon Composite Manufacturing

Production processes for composite materials can be divided into two categories: prepreg (short for pre-impregnated) methods, where the fibres have been mixed with the resin before they have been brought into the final shape, and dry fibre processes, where the resin is added at the end of the process in an infusion step. Prepreg processes, such as Automated Fibre Placement, require expensive material handling under freezing temperatures. Dry fibre materials can be stored and processed at room temperature [13]. However, even small defects in the dry fibre layups can have significant effects on the final part's properties. Some of the most common defects in dry fibre processes and their influence on the final material's properties have been investigated in the literature. These are: Misaligned fibres [12], gaps in between the fibres [5], and foreign objects [10].
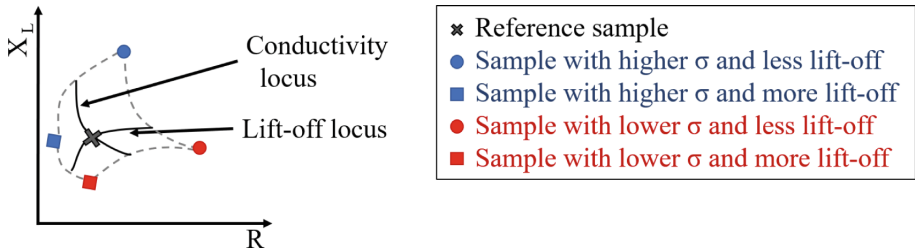
2.2   Eddy Current Sensors

On the left-hand side, Fig. 1 illustrates the basic principle of an eddy current sensor on a conductive specimen. An emitting coil creates an alternating primary magnetic field, which penetrates the specimen. According to Faraday's law of inductivity, a changing magnetic field induces an eddy current in the specimen. This creates a secondary magnetic field which can be picked up by either a second coil or by measuring the change of impedance in the emitting coil. The sensor's response is a signal in the complex plane with the coil's resistance on one axis and the inductive reactance on the other.

The result depends on different aspects - two of the most sensitive are the lift-off (distance between the probe and the specimen) and the specimen's electrical resistance. Figure 2 shows the loci of the results when these two variables are changed. The exact shape of the loci depends on the used probe and the material of the specimen [7]. In addition, tilting of the probe influences the results significantly [4].
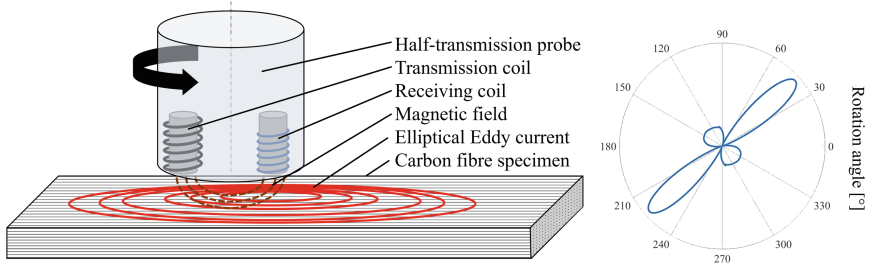
Fig. 1. Principle of the polar measuring method of uni-directional carbon fibres with a half-transmission Eddy current probe

Fig. 2. Eddy current sensor responses for specimen with different conductivity σ and different lift-offs (adapted from [7]). The shape of the loci depend on the material of the specimen and the specific probe.

2.3 Measuring Methods

Polar Method

The polar measurement is a method to determine the local fibre orientation at a specific location. It makes use of the elliptical shape of the eddy current, which is caused by the non-isotropic electrical properties of carbon fibre material. This aspect can be measured by a probe with spatially separated transmission and receiving coils, such as the half-transmission design. While this end-effector is rotated over the point of interest, the measured current in the receiving coil changes in correspondence to the shape of the eddy current. The principle is illustrated in Fig. 1 on the left. The complex response in a 2D plane first needs to be reduced to a 1D signal and then plotted against the rotation angle of the probe (illustrated in Fig. 1 on the right-hand side). The peaks of the lobes in this graph align with the fibre orientation [3]. The signal is symmetric. Therefore, it is sufficient to evaluate only 180° of the whole rotation [8].

The exact response of the coil is highly dependent on the exact probe design and the material of the specimen. Therefore, there are different approaches to reducing the complex signal to a 1D signal. Lange and Mook assessed the magnitude of the 2D signal [8]. Another method is to apply a vector rotation of the complex signal around the origin. Subsequently, the x-value of the data can be evaluated [9].

Using the polar method, Lange and Mook determined the fibre orientation of a single ply with an accuracy of less than 1°. However, they found that the results are less accurate for two plies orientated at a small angle relative to each other. Therefore, they proposed a reverse engineering approach based on a synthesis of two signals followed by a correlation. This approach showed an improved accuracy [8].

All the systems used in the literature to perform a polar measurement had the probe fixed to one position - either in a tool or as a rig set-up that always measures the same spot. For the inspection of complex 3D parts, a system is needed that is able to approach different positions for inspection.

Scanning Method

While the polar method can be used to inspect a single spot, the scanning method is able to provide a global view of the whole specimen. The analysis of this image can reveal various defects, such as the fibre angle deviations, gaps, overlaps, or foreign objects [11]. For this, the probe is guided along a grid that covers the specimen's entire surface. Along the path, at certain intervals, the sensor's response is measured. These data points are again reduced to a 1D signal and mapped on a grey scale. The lateral information and the grey scale value can be combined into a grey scale image of the entire specimen.

The resolution of the image depends on the size of the grid. The smaller the grid cells, the higher the resolution of the image. For the evaluation of the fibre angle, the texture of the textile must be visible. Even with a very small grid size, the measurement will not resolve the individual fibres but rather a macroscopic pattern of the fibre ply [11]. A woven material has a very distinct pattern. In non-crimp fabrics, using a very small grid size, the stitching pattern can be visualised. Unidirectional tapes, which are placed without gaps, lack a regular pattern and appear as a homogeneous surface in the grey-scale images. For this reason, Bühlow states that this method is unusable for unidirectional tapes [2].

With a visible pattern, image processing algorithms can be applied to the image to detect the fibre orientation. Bühlow determined it with a precision of 0.1° [2]. Heuer et al. demonstrated how to identify defects, such as foreign objects, in-plane, and out-of-plane waviness in cured composite parts. Copper injections were detected under up to 12 layers. The detection depth of the tape, polyethylene foil, and paper was 9 layers [6]. Mook et al. scanned a specimen with changing fibre orientation in a stair-like manner and were able to illustrate the discontinuities in the results [9]. Gaps in dry fibre layups have not been investigated with an eddy current sensor.

3   Method

3.1   Hardware Selection

In order to realise a system which is capable of performing both measuring methods, the eddy current probe is attached to a 6-axis robot arm. CIKONI GmbH supplied the system. This allows not only the necessary movements for the measurements of a flat specimen but can also follow the surface of a draped, 3D specimen. A similar set-up has been demonstrated by Bardl et al. who investigated a woven material while draping it into the third dimension. They, however, only applied the scanning method [1].

The eddy current sensor system is based on Suragus' EddyCus Sensor Kit. It can be equipped with different probes. Bühlow found that differential probes are not suitable for carbon fibre inspection [2]. For the scanning method, the absolute probe design (one coil design - response is measured as a change in impedance in the excitation coil) is chosen because it is the smallest possible design, improving the resolution of the measurement. For the polar method, a half-transmission probe (separate excitation and receiving coils) is used because a spatial separation between the transmission and receiving coil is needed.

As mentioned above, the lift-off is a relevant factor during the measurement and should be kept constant. For the polar method, a constant distance can be achieved by rotating the probe perpendicularly to the surface. A perfect normal alignment is also important to reduce tilt effects. During the scanning method, a constant lift-off is harder to maintain, as dry fibre layups do not have a uniform surface but deviate in height. This out-of-plane waviness leads to bright spots on the image where the fibres were closer to the probe. These effects dominate the results and make it impossible to evaluate the other areas. To tackle this issue, a non-conductive, incompressible foil of constant thickness was placed on the fibres. The robot was programmed to put the sensor in light contact with the additional layer - this way, a constant distance was maintained during the measurement.

3.2   Specimen and Experimental Plan

For all experiments, a unidirectional non-crimp fabric out of dry carbon fibres, with an areal weight of $80 \mathrm{g/m^2}$ is used. On the left-hand side, Fig. 3 shows the back of the fabric with a check-pattern of scrim fibres that keeps the carbon fibres in place. The fabric was produced by epo GmbH.

A separate piece of this fabric is prepared for gap examinations. Gaps of different sizes are cut into the fabric. In addition, some fibres are displaced to create a different kind of gap. Figure 3 shows this sample on the right-hand side.

For the scanning method, two setups are measured. For the first setup, the fabric with the induced gaps is covered by an defect-free layer and scanned. For the other specimen, different foreign objects of different sizes (1–5 mm) are placed under one defect-free layer of the fabric. This specimen is scanned to determine whether the particles can be detected.

For the polar measuring method, a device that rotates the fabric is built. It allows to set different fibre orientations with a tolerance of <0.5°. To determine the accuracy and precision of the eddy current measurement, two layers were stacked with an angle of 90° and placed in the rotation device. It is set to 11 different angles between 0 and 90°. For each angle, 3 measurements are conducted.
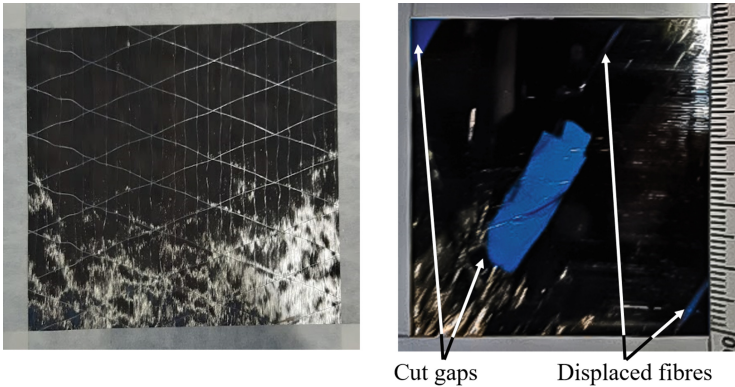

Fig. 3. Left: The non-crimp fabric material which is used for the specimen. Right: Specimen with introduced gaps.

4   Results and Discussion

4.1   Fibre Orientation

As an example of the polar measurement, Fig. 4 (left) illustrates the results of the 3 repetitions where the fixture was set to 45°. The 3 polar plots show a good agreement, and the red lines that indicate the maxima of the graph (which correlate with the detected fibre orientations) are well aligned. In the entire experiment, the angle difference between the two layers was measured to an accuracy of 0.9°. It should be mentioned that the specimen was handled manually. Therefore, the accuracy might be limited by slight deviations introduced by the operator. To eliminate this factor, the standard deviation of all measurements was calculated to be 0.16°. This is well below the precision of the fixture.
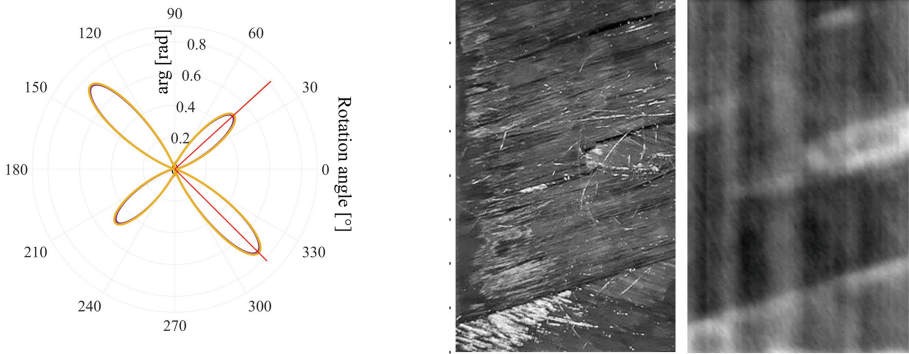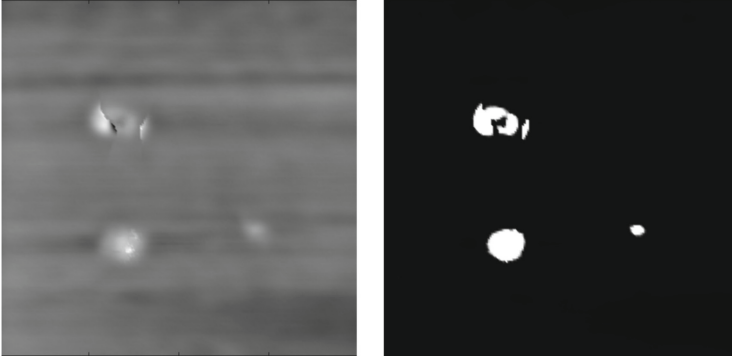
Fig. 4. Results of the eddy current measurement. Left: Result of the polar measurement. Middle: Photo of the specimen before it is covered with an additional ply. Right: Scan of the same specimen.

4.2 Gap Detection

The results of the scanning measurement with the gap specimen are illustrated in Fig. 4 on the right-hand side. The image of the specimen before it is covered with an additional ply is compared with the scan result. The gaps are clearly visible as brighter spots on the image. Even the small gaps, which are just displaced fibres, show up on the measurement. The vertical lines in the image are along the scanning direction of the sensor. The used eddy current instrument tends to drift over time, even if it is not moved. This effect is compensated by an automatic drift compensation which is applied frequently. However, the drift can still be seen in the image. The lines which run diagonally from the top left corner to the bottom right are the fibres of the layer covering the defective layer.

4.3 Foreign Objects

In Fig. 5, the fibre orientation of the covering layer and the scanning direction are both horizontal. The three white spots on the grey scale image are metal pieces of different sizes that have been induced. Even the small piece of $1 \times 1\text{mm}^2$ can be identified. The signals are clear enough to be easily thresholded for image segmentation. Objects made out of plastic do not show up in the scan. This result is reasonable, as they are not conductive and can therefore not create an eddy current.

Fig. 5. Results of the eddy current measurement of a specimen with different inclusions. Left: grey scale image. Right: Segmented result with a threshold applied to the gery scale. (Color figure online)

4.4   Discussion

The set-up of an eddy-current sensor attached to a 6-axis robotic arm allows for the execution of both measuring methods - the polar and the scanning method. The main challenge is to position the sensor precisely. For the polar method, the probe must be kept perpendicular to the specimen's surface. Otherwise, tilting effects might falsify the results. Therefore, polar measurements on a flat specimen can easily be performed. However, for 3D measurements the surface must be known in detail in advance to calculate the surface normals. For the scanning method, it is very critical to keep a constant distance between the probe and the fibres. The surface of dry fibre layups tends to be uneven, which leads to disturbing spots on the scan due to the lift-off effect. For this investigation, the issue was solved by flattening the fabrics with a non-conductive foil in between the probe and the fibres. However, this procedure turns this originally non-contact inspection into a contact method.

The data captured by the rotating method can be evaluated by finding the maximum of the response. A specimen with fibres in 0 and 90° orientation can be measured with a precision of 0.16°. Smaller angle differences between the layers require a more complex evaluation, as described in the literature [8].

The scanning method can detect defects like gaps and conductive foreign objects. Pieces from non-conductive materials, such as plastic foils, cannot be seen on the scans. Heuer et al. showed that even non-conductive insertions in cured carbon composite parts can be visualised [6]. The reason why these object showed up in infused parts, but not in dry layups is that during infusion the fibres drape tightly around the insertion. Therefore, the eddy current path in the fibres is influenced by the object which is detectable in the scan. On the other hand, the fibres of dry layups rest loosely on thin objects. This does not change the eddy current path in a detectable manner.

The research question can be answered positively. However, different probes are required for the two methods, and the limitations, which are mentioned above, must be addressed.

5   Conclusion

This research demonstrated that the most common defects in a dry fibre layup (fibre angle deviations, gaps and foreign objects) can be detected with a single eddy current inspection system. To do so, a 6-axis robot was programmed to position the eddy current sensor. Two measuring methods were investigated: the polar and the scanning method. The former is capable of detecting fibre angles with a precision of $0.16°$. This was proven for a 2 ply layup with perpendicular fibre orientation. For this method, a probe with a half-transmission set-up was used. The scanning approach, using an absolute probe design, proved to be useful for inspecting gaps and foreign objects. Metal insertions and gaps were easily detected, while non-conductive objects could not be identified.

The main challenge was to position the probe, as it needs to be perpendicular and within a small, constant distance to the fibres. Dry fibre layups tend to have an inhomogeneous surface which makes a contact-less scanning inspection very difficult. In the future, this issue could be solved by either a spring-loaded mechanism that keeps the probe at a constant distance from the fibres or a prior laser scan inspection to determine the exact surface of the specimen.

Furthermore, the reliability of these measuring methods should be investigated for different materials and multiple layers.

References

1. Bardl, G., Nocke, A., Hübner, M., Gereke, T., Pooch, M., Schulze, M., et al.: Analysis of the 3D draping behavior of carbon fiber non-crimp fabrics with eddy current technique. Compos. Part B Eng. **132**, 49–60 (2018). https://doi.org/10.1016/j.compositesb.2017.08.007
2. Bühlow, C.: Dynamische Toleranzkontrolle im automatisierten Carbonfaser-Preforming. Dissertation. Technische Universität Braunschweig, Braunschweig (2019)
3. De Goeje, M.P., Wapenaar, K.E.D.: Non-destructive inspection of carbon Fibre-reinforced plastics using eddy current methods. Composites **23**(3), 147–157 (1992). https://doi.org/10.1016/0010-4361(92)90435-w
4. Grahner, P.: Implementierung eines Wirbelstromsensors in eine vollautomatisierte Prozesskette zur endkonturnahen Fertigung komplexer RTM-Bauteile. Master Thesis. Fachhochschule Aachen, Aachen. Lehrstuhl für Flugsystemtechnik und Flugführung (2014)
5. Frieder, H.: Qualitätsbewertung von Faser-Kunststoff-Verbunden mittels optischer Texturanalyse auf 3D-Preformoberflächen. Dissertation. Universität Stuttgart, Stuttgart (2019)

6. Heuer, H., Schulze, M., Pooch, M., Gäbler, S., Nocke, A., Bardl, G., et al.: Review on quality assurance along the CFRP value chain - non-destructive testing of fabrics, preforms and CFRP by HF radio wave techniques. Compos. Part B Eng. **77**, 494–501 (2015). https://doi.org/10.1016/j.compositesb.2015.03.022

7. Hillmann, S., Schulze, M.H., Heuer, H.: High-frequency eddy current techniques. In: Ida, N., Meyendorf, N. (eds.) Handbook of Advanced Non-Destructive Evaluation, pp. 1–28. Springer International Publishing, Cham (2018). https://doi.org/10.1007/978-3-319-30050-4_49-1

8. Lange, R., Mook, G.: Structural analysis of CFRP using eddy current methods. NDT & E Int. **27**(5), 241–248 (1994). https://doi.org/10.1016/0963-8695(94)90128-7

9. Mook, G., Lange, R., Koeser, O.: Non-destructive characterisation of carbon-Fibre-reinforced plastics by means of eddy-currents. Compos. Sci. Technol. **61**(6), 865–873 (2001). https://doi.org/10.1016/S0266-3538(00)00164-0

10. Oromiehie, E., Prusty, B.G., Compston, P., Rajan, G.: Automated Fibre placement based composite structures: review on the defects, impacts and inspections techniques. Compos. Struct. **224**, 110987 (2019). https://doi.org/10.1016/j.compstruct.2019.110987

11. Schmidt, C., Schultz, C., Weber, P., Denkena, B.: Evaluation of eddy current testing for quality assurance and process monitoring of automated fiber placement. Compos. B Eng. **56**, 109–116 (2014). https://doi.org/10.1016/j.compositesb.2013.08.061

12. Tsai, S.W., Hahn, H.T.: Introduction to Composite Materials. Technomic Pub, Westport Conn. (1980)

13. Veldenz, L.: Automated dry Fibre placement and infusion process development for complex geometries. Doctoral. University of Bristol, Bristol (2019). https://research-information.bris.ac.uk/en/studentTheses/automated-dry-fibre-placement-and-infusion-process-development-fo. Accessed 11 Sep 2021

Detection of Gaps and Overlaps in Laser Line Triangulation Data of Dry Fibre Tape Layups Using Image Projection

Christopher Mimra[1,2]([✉]), Vincent Krein[2], Racim Radjef[1], Bronwyn Fox[3], and Peter Middendorf[2]

[1] Swinburne University of Technology, Hawthorn 3122, Australia
cmimra@swin.edu.au
[2] IFB Universität Stuttgart, 70569 Stuttgart, Germany
[3] Commonwealth Scientific and Industrial Research Organisation, Clayton 3168, Australia

Abstract. A major cost driver in the production of carbon composite parts is the quality inspection, which to this day relies on manual investigation by a trained worker. Gaps and overlaps in the layups are to be detected because they are proven to be detrimental towards the mechanical properties of the final part. In recent works, laser line triangulation sensors have been applied to inspect layups of prepreg tapes and non-crimp fabric material. These sensors create a 3D point cloud of the specimens surface. This is then evaluated by conversion into a greyscale image and a subsequent image processing algorithm. However, the most commonly used algorithms fail to differentiate between defects and small, acceptable irregularities, such as welding spots, slits and single fibres which stick out.

The aim of this research is to develop a reliable evaluation method for scans of dry fibre tape layups. An overview over the different groups of algorithms is provided, image projection is selected and compared to algorithms which have been proven to work best on pre-pregs. While the common algorithms fail to classify a test set of dry fibre specimen, image projection can reach a true positive rate of 100% and a false positive rate of 19%. The proposed setup can be a centrepiece of a future in-line quality inspection system for dry fibre layups which has potential for a significant decrease of the manufacturing costs.

Keywords: Dry Fibre Placement · Laser Triangulation Sensor · Image Projection · Carbon Composites · Quality Inspection

1 Introduction

1.1 Background

Composite materials are used in various industrial applications and for light high-performance structures, for example in the automotive industry. Main reasons are the advanced specific mechanical properties compared to metals. For

example, Kim et al. have designed a automotive lower arm which is 50% lighter and much stiffer than the conventional steel part [9]. However, the production of composite materials is still rather expensive. A significant part of the production costs are associated with quality inspection, as this is still conducted as manual labour [13]. An automated inspection system therefore carries a big potential to drastically reduce the manufacturing costs of composite parts.

In this work an inspection system for the Dry Fibre Tape Placement process is proposed. It is based on a laser line triangulation sensor which captures the surface of the fibre layup in-line and outputs a point cloud. A challenge of this application is to process the data in order to detect and classify defects in the fabric. In the literature, multiple different approaches from the domain of machine vision are considered. Multiple authors summarised and compared the methods, for example in the field of industrial fabric production [6,10,14], the medical domain [11,15], or the Automated Fibre Placement process [16]. However, as Mahajan et al. stated, there is not one best approach for all applications. Rather, for each use-case a suitable method must be selected, considering the specific process characteristics [14]. The inspection of dry fibres bear the challenge that the surface in not homogeneous. A layup without resin contains small out-of-plane waviness and some fibres sticking out of the tape. Furthermore, some dry fibre tapes already contain slits to enhance the resin flow during the infusion step. These kinds of irregularities are tolerable. On the other hand, gaps and overlaps must be detected as they influence the local fibre volume fraction and consequently the final mechanical properties [2]. So far, there is no report about an implementation of a method, that reliably differentiates defects (such as gaps and overlaps) from acceptable irregularities in dry fibre layups. The aim of this paper is to answer the following research question: *Which evaluation method can reliably detect gaps and overlaps in a dry fibre tape layup, which contains strong irregularities (such as slits, waviness etc.) and accurately classify them as defective?*

1.2   Dry Fibre Tape Placement

Production processes for composite materials can be divided into two main categories: pre-preg (short for: pre-impregnated) methods, where the fibres have been mixed with the resin before they have been brought into the final shape, and dry fibre processes, where the resin is added at the end of the process in an infusion step. While some pre-preg processes, such as Automated Fibre Placement, are established in the industry, they also require an expensive material handling under freezing temperatures [22]. Dry Fibre Tape Placement on the other hand promises to be an inexpensive production approach with little scrap material and low cycle times [5]. In this work, the FILL Multilayer (see Fig. 1 left) is used to deposit straight strips of bindered dry fibre tape onto a vacuum table with 360 degrees of freedom. It uses multiple laying heads simultaneously to lay the tapes. Each laying head performs 1-dimensional movements on linear rails. This process creates a near-net-shape layer of carbon fibre. After each layer the table can turn the layup and another layer with a different fibre orientation

can be added. The tapes of the different layers are held together by thermoplastic binder that is activated by local ultrasonic welding.
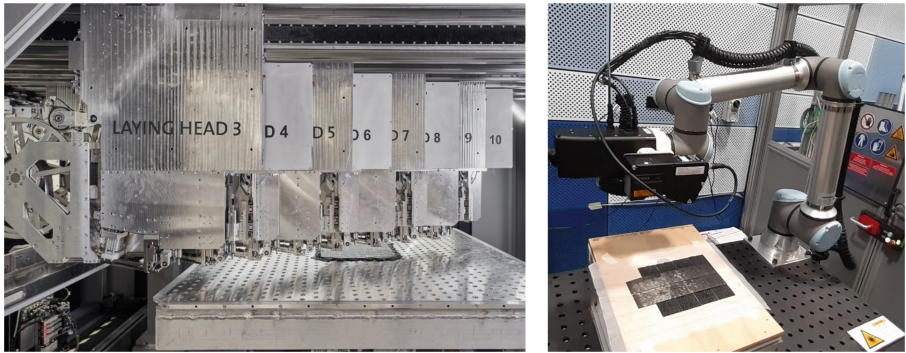

Fig. 1. Left: Multilayer with multiple laying heads running on rails. Right: Setup of the laser line triangulation sensor on a 6-axis robot arm.

1.3   Defects and Irregularities

At the stage of the 2D layup, many different irregularities can be observed. On the grey scale image of a defect-free specimen (compare to Fig. 2 left), an overall gradient in height from left to right is clearly visible. This is caused by the stiffness of the tape which was previously wound on a spool and tends to spring back out of plane after it was laid on the table. The oval shaped darker spot towards the right edge of the image is the dent, which the welding probe punches. To melt the binder, the tip of the ultrasonic device is pushed into the layup. Left of this dent a slit can be seen which was introduced by the tape's manufacturer in a regular pattern to enhance the resin flow during the infusion step. All over the image slight lines running from left to right indicate fibres that slightly stick out compared to their neighbours. All these irregularities are inherent in the process and acceptable towards the quality of the final part.

In between neighbouring tapes however, gaps and overlaps can be observed in this process. These can occur if the tape is locally too narrow (which leads to gaps) or too wide (which leads to overlaps). Furthermore, the tape usually has some lateral clearance on both sides in the laying head to prevent it from getting stuck. This might also lead to slight deviations in the positioning of the tape, resulting in gaps or overlaps. Both of these defects are detrimental towards the final part's mechanical properties, because they change the local fibre volume fraction [2].

Further defects that have been reported to occur in dry fibre tape placement are fold-ups, waviness, and early/late cuts [25]. These however are not considered in this work as they have not been observed regularly in the FILL Multilayer process or can easily be compensated by trimming.
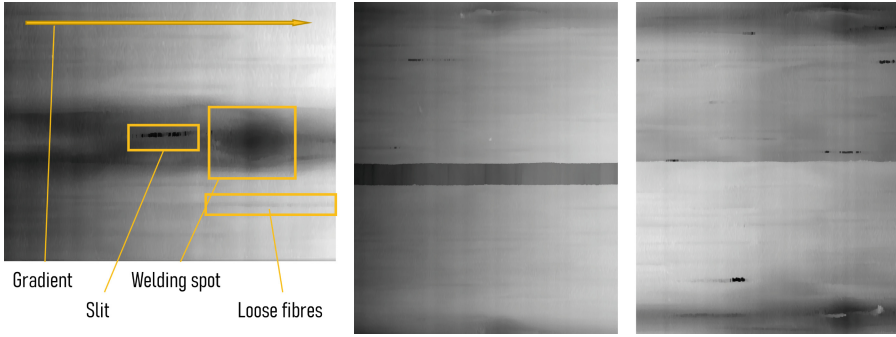
Fig. 2. Grey scale image of three specimen: defect-free (left), with a gap (middle), with an overlap (right). The grey scale represents the height value.

1.4   Laser Line Triangulation

A laser line triangulation sensor projects a laser line vertically onto the specimen's surface. In an angle to this beam, a camera records the projection. Therefore, height differences in the specimen can be identified as a lateral shift on the taken image. The sensor's software calculates a height profile of this image. As the sensor is linearly moved over an area of interest, it captures multiple profiles. These individual data sets can be reassembled to create a 3 dimensional representation of the specimen's surface.

In research laser line triangulation sensors are commonly used for the quality inspection of composite parts. Unlike other inspection technologies, such as CT or optical systems, it can perform an in-situ inspection without adding an additional process step. The method has been verified with different carbon fibre materials, such as dry non-crimp fabrics [21], dry tows [26] and pre-preg tapes [1,16,20].

1.5   In-Line Implementation

In order to implement an in-line quality inspection of the layup during the production, a laser line triangulation sensor will be attached to a laying head and travels with it on the linear rail. This enables in-situ data collection during tape placement and avoids disruptions in the production process. Another advantage is, that possible defects are always oriented in the same direction relative to the sensor. In Fig. 2 the scanning direction will always be horizontal.

2   Literature and Methods

2.1   Evaluation Methods in the Literature

A commonly used inspection method for fibre layups is the optical inspection, where a camera takes a high resolution image of the fibre layup. This image

is then evaluated. Focke et al. proposed to define a gap as an area where the fibre orientation is near to perpendicular to the dominating fibre orientation in the image [4]. The fibre's direction can be determined by for example a fast-Fourier-transformation [18] or direct tracking [17]. Both methods require an image quality which is sufficient to identify the individual fibres. For this, the camera can only take a small segment of the layup at once to provide a sufficiently high resolution [7]. Furthermore, the camera must stand completely still and vibration free or the image might be blurred. Hence, an optical system is not suitable for an in-line inspection on a moving laying head.

Laser line triangulation sensors, on the other hand, have been proven to provide good data while they are moved over the specimen. Some commercially available laser line triangulation systems take the positions of the measured points and apply an one- or two-sided threshold to the absolute z-value [8,23]. Some other systems can compare the measured points to CAD data and threshold the distance between the two data sets [3]. However, these methods proved to be inadequate to detect small defects in carbon fibre layups. Therefore, researchers have adopted algorithms from other domains. Most commonly, the 3 dimensional point cloud is converted into a 2 dimensional grey-scale image, where the z-coordinate is represented as brightness of the individual pixels. This bears the advantage that image processing algorithms from different domains can be applied.

Hanbay et al. investigated various image processing methods for fabric inspection and classified them into five categories: "structural", "statistical", "spectral", "model-based", "learning" and "hybrid models" [6].

Structural approaches require a definition of an arrangement of different primitive texture elements. This can be lines or dots on the image which should appear in a regular pattern. The algorithm checks whether these patterns are broken. Hanbay et al. state, their "reliability [...] is low. Structural approaches are only reliable in segmenting fabric defects from texture whose pattern is very regular" [6, p. 11963]. Laser scan images of dry fibre tapes, unlike woven materials for example, usually don't have a regular pattern, as can be seen in Fig. 2. This is why this group of algorithms is not applicable to dry fibre tape laying. **Statistical approaches** are based on statistical evaluation of the image data to extract features. The absolute threshold method is one of them. Furthermore, many edge-detection methods are part of this group. To mention are the morphological edge detection [16,24], gradient thresholding methods using the Sobel, Laplace and Scharr algorithms [12,16], cell-wise thresholding [16] and adaptive thresholding [10,14,16]. Another statistical approach is the image projection where the brightness values of the image are reduced to a graph. This can then be evaluated with basic analytical functions. Hanbay et al. showed that the position of an elongated defect in woven material can be determined in one dimension of the grid [6]. Lastly, a method using co-occurrence matrices must be mentioned in this group. This approach detects recurring patterns in the image. **Spectral approaches** intend to detect regular or homogeneous patterns in the image. The earlier mentioned fast-Fourier-transformation is one of them. Other

methods in this group are wavelet- or Gabor-transformations. The resolution of the laser line triangulation data is not sufficient to identify individual fibres. For this reason, spectral approaches cannot detect the fibre directions or defects. **Model-based approaches** require a model, which describes and constructs the expected data. This model is then compared to the measured data to detect unwanted areas [6]. Römer states that these "control systems are very complex to develop because many different effects have to be considered" [19]. Especially in the dry fibre tape placement, where the width deviations of the tape depend on the manufacturer's process, it is extremely difficult to create a model that predicts gaps or overlaps. Irregularities such as slits and individual fibres sticking out are close to randomly distributed. Therefore, creating a model is infeasible. **Learning approaches** are methods, where networks are taught to detect defects. They require a great amount of data. Creating this labelled training data experimentally is very expensive and time-consuming. For this reason, Zambal et al. chose to generate artificial data using a probabilistic model. However, they found that the network performed worse when it was trained on artificial data [26]. The approach, which this project follows is to find a suitable statistical classification algorithm which can then label sets of training data automatically. Then, at a later stage, learning approaches will be considered and included in the research.

Meister et al. compared different algorithms for their use-case of the automatic fibre placement of pre-preg fibres. They assessed 29 commonly used evaluation methods to find a suitable algorithm. 5 options were investigated in-depth. Finally, they concluded that adaptive thresholding and cell wise standard deviation thresholding suit the requirements best [16].

2.2 Research Method for the Algorithm Assessment

In this research, the projection method was chosen, as it was proven to detect long defects [6]. Gaps and overlaps span along the whole width of the specimen. Therefore, it is assumed that this method can be suitable. However, in this study it is compared to the most suitable methods from Meister et al. paper. For this, 63 images of scanned dry fibre tape layups were fed into the algorithms. These images contained either no defects (21 samples), a gap (21 samples) or an overlap (21 samples). It is assessed whether the algorithms detect the flaws correctly. The measures were the true positive rate (defect found correctly), and the false positive rate (defect found on a defect-free sample). Furthermore, it was investigated whether the size of the defect can be assessed.

2.3 Specimen, Hardware and Preprocessing

The specimen were produced by the FILL multilayer. The material used for the specimen is a dry fibre, uni-directional carbon fibre tape by MTorres. Its areal weight is $450\,\mathrm{g/m^2}$ and its thickness is on average $0.52\,\mathrm{mm}$. For the defect free samples, an area in the middle of one tape was taken. This way it was made sure, that it did not contain any gaps or overlaps but only irregularities, which

are acceptable. For the gaps and overlaps two tapes where placed next to each other with a defect size of 1, 3 or 5 mm (7 samples each). The defect was placed roughly in the middle of the scan.

To acquire the data a preliminary hardware set-up is used, which imitates the movements of the laying heads. A commercial laser line triangulation sensor (Keyence LJ-X8200) scanned the specimen. This device works at a distance of 245 mm from the specimen with a 72 mm field of view. Each profile provides 3200 points, which is equivalent to a resolution of 44 dots/mm in x direction. For the measurement, a 6-axis robot arm performed a linear, homogeneous movement at a speed of 10 mm/s. At a measurement frequency of 1 kHz this results in a resolution of 100 dots/mm in y direction. During the measurement the sensor head provides the data to an evaluation unit (Keyence LJ-X8000A). The set-up is illustrated in Fig. 1 right.

On the software side, the following process is used to pre-process the acquired test data: As a first step, the data is collected using Keyence' software LJ-X Navigator in order to receive a point cloud for further evaluation. This point cloud is cropped in x/y direction to a size of 50×50 mm^2. Furthermore, all outliers in z-direction which might be caused by reflections are removed. In a next step, the points are projected on a 2000×2000 pixel grid. If multiple points fall on the area of a pixel, the average z-value is taken. If, on the other hand no point falls into the pixel, the average of the surrounding pixels is calculated. To convert the point cloud into a grey scale image, the z- value is mapped on a 8-bit grey scale. The contrast equalisation is performed with the CLAHE algorithm to compensate for uneven intensity distributions. Figure 2 shows the pre-processed scans.

3   Results

3.1   Image Projection

The image projection algorithm was programmed to calculate the average grey scale value of every row of the image $R_{\mathrm{r}}(i) = \sum_{j=0}^{n}(I(i,j))$. This graph can be seen in Fig. 3. If this graph showed an average slope of more than 2 Intensity steps per pixel over at least 14 pixels ($\frac{\sum_{i=k-6}^{k+7} R_{\mathrm{r}}'(i)}{14} > 2$), this area was marked as a suspicious area.[1] If there is a pair of suspicious areas, which have opposite-signed slopes, they are identified as a gap. All single suspicious areas are defined as overlap. One example result for a specimen with a gap can be seen in Fig. 3. If an image contains either a gap or an overlap, it is classified as defective.

---

[1] The criterion to take the average over multiple pixels was added, because it naturally filters some laser reflections, which affected only one single pixel. Before, these single pixels which are far off the values of their neighbours, led to many false positive results.
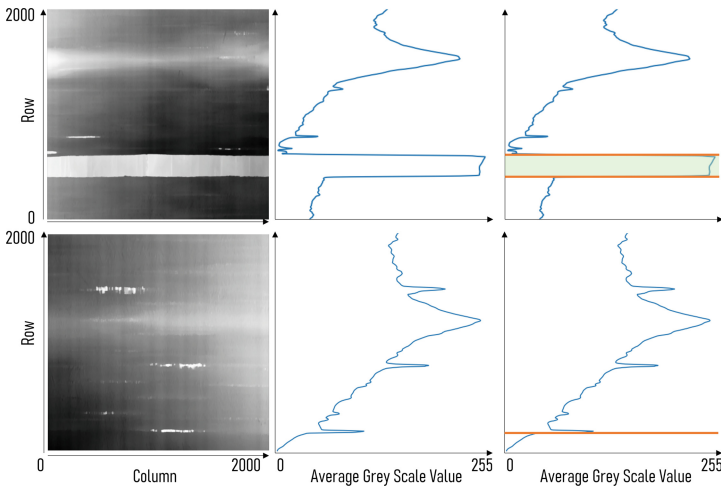
Fig. 3. Top: Successful defect detection using Image Projection. Left: grey scale image, middle: average grey scale values for each row, right: detected gap. Bottom: False detection of an overlap

Out of the 63 samples, all defective samples were classified as such. The true positive rate is 100%. However, on 7 samples an additional defect was detected, where none was expected. 3 of these were on samples that had a defect at another position. The remaining 4 were on defect free sample. Therefore, the false positive rate is 19%. The confusion matrix can be seen in Table 1. All of the 7 false reports can be explained with a slit, which was extraordinary deep at this position.

Table 1. Confusion matrix for defect classification using image projection

		Evaluation outcome		
		Defect	**No Defect**	**total**
	Defect	True Positive 42	False Negative 0	42
Ground Truth	**No Def.**	False Positive 4	True Negative 17	21
	total	46	17	

In addition to this classification, the algorithm also calculates the width of a gap once detected. The standard deviation between the produced gap width and the measured gap width is 0.12 mm. This accuracy includes some manufacturing tolerances, which are in the same range. This value has not been cross-checked

with another measurement system to determine the true size of the gap. However, this feature to measure the gap width is useful if the requirements are changed. If, for example, overlaps and gaps smaller 2 mm are allowed, the true positive rate is 100% and the false positive rate 0%. This means that all previously falsely detected defects are smaller than 2 mm.

3.2 Adaptive and Cell-Wise Thresholding

Adaptive Thresholding is an approach, where no global threshold value is set. Instead the threshold for each position on the picture is determined by the neighbouring pixels. A mean (or Gaussian-weighted mean) of all pixels in a predefined radius is calculated and the threshold is defined based on the result. Cell-wise thresholding works similarly, with the difference that calculation is faster due to the predefined grid. Furthermore, the standard deviation of the pixel values is incorporated, which makes the method more accurate in specific use-cases [16].
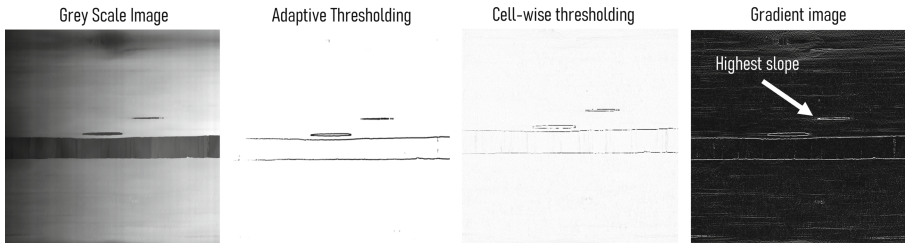
Fig. 4. Processed grey scale image (left) with different evaluation methods: Adaptive thresholding, cell-wise thresholding, gradient analysis (left to right)

Both methods were implemented and applied to the test data. It was found that all defects can be detected if the thresholds are selected sufficiently low. However, many small irregularities lead to false classifications. The algorithm marks all pixels that deviate from their neighbours, without differentiating whether this pixel is part of a major defect or a local deviation, such as a fibre sticking out of plane. This can be seen in Fig. 4 where, even in defect-free areas many pixels exceed the threshold and are thus marked as defects. In fact, all images of the data set where marked defective.

In order to lower the false positive rate, the threshold was increased and the image filtered and smoothed to reduce the severity of irregularities. However, no set of parameters was found that leads to a true positive rate of 100% and a false positive rate of <100%. In other words, either the algorithm misses multiple defects, or it is marks every image as defective. This paradox can be explained by analysing an image that shows the local gradients of the data set (see Fig. 4 right): Some irregularities have a higher gradient than the defects. Therefore, any set of parameters that detects the defect, also marks these irregularities as defective.

4   Discussion

In the pre-processed images, the inhomogeneity of the surface of dry tape layups can clearly be seen. Depending on the requirements, these irregularities (out-of-plane spring back, waviness and slits) might not be considered defects even if their local magnitude is similar to critical defects, such as gaps and overlaps. However, they are the reason why the algorithms that work best on pre-preg tapes (compare to [16]) fail on dry fibres. Above all, slits, which were introduced by the tape manufacturer, make a correct classification difficult if algorithms are used that threshold the slope of the edges or the magnitude of local pixels.

Spectral approaches have been proven to be applicable on high-resolution optical inspection systems [4,18]. They are based on the difference of fibre orientation between stacked layers. However, a sufficiently sharp optical image cannot be taken while the laying head is moving. Furthermore, the resolution of laser line triangulation sensors is too low to identify individual fibres. This is why the commonly used fast-Fourier transformation cannot be applied.

The proposed method of image projection makes use of the fact, that the relevant gaps span the whole image from left to right and are always parallel to the image's edges. Using these properties a classification with a true positive rate of 100% and a false positive rate of 19% can be reached. This accuracy exceeds the results from the literature in both, the false negative rate and the false positive rate (number based) [16]. However, these results also mean, that every fifth dataset is falsely classified defective. Therefore, a further inspection step is needed to reassess the results - this could either be a manual or an optical investigation. However, the proposed method is still useful, as the localisation of possibly defective areas accelerates the second inspection step drastically. It can be concluded, that this approach is a step towards a faster and cheaper inspection of dry fibre layups but does not yet replace other inspection steps.

The application of the image projection method is not limited to the FILL multilayer, but can be considered for all use-cases with dry fibres, where the laying heads run on straight paths. The irregularities, which prevent other evaluation methods from a proper classification, occur in most dry fibre tape processes.

5   Conclusion and Outlook

The aim of this research was to identify an evaluation algorithm that could detect gaps and overlaps in dry fibre tape layups. These contain irregularities such as an overall slope, waviness, slits or single fibres sticking out of plane. These irregularities make the application of adaptive and cell-wise thresholding methods impossible, even though they have been proven to perform well on pre-preg tape layups. Instead the image projection method is proposed which makes use of the uniform shape and orientation of the defects. When applied to

a test set of specimen that contain gaps, overlaps and defect-free layups, the true positive rate is found to be 100%, while the false positive rate is 19%. Hence, this method cannot yet replace an additional process step for the inspection, but make it faster by preselecting possibly defective areas. This can be beneficial towards the process time and cost.

In future works, the parameter will be optimised further to improve the false positive rate. This algorithm will then be used to create labelled training data for a machine learning approach. Furthermore, the hardware system will be transferred from the test set-up (mounted on a robot) and be implemented in a Dry Fibre Tape Placement machine. The goal is to scan and evaluate the layup in real time to provide an in-line inspection system for this manufacturing process.

References

1. Cemenska, J., Rudberg, T., Henscheid, M.: Automated in-process inspection system for AFP machines. SAE Int. J. Aerosp. **8**(2), 303–309 (2015). https://doi.org/10.4271/2015-01-2608
2. Croft, K., Lessard, L., Pasini, D., Hojjati, M., Chen, J., Yousefpour, A.: Experimental study of the effect of automated fiber placement induced defects on performance of composite laminates. Compos. Part A Appl. Sci. Manuf. **42**(5), 484–491 (2011). https://doi.org/10.1016/j.compositesa.2011.01.007
3. FARO (2021): FARO Quantum Max. The Next Generation ScanArm. https://media.faro.com/-/media/Project/FARO/FARO/FARO/Resources/2021/06/30/22/08/Brochure_QuantumMax_3DM_LTR_ENG.pdf. Accessed 15 Jan 2022
4. Focke, O., Miene, A., Koerdt, M.: Automated quality inspection in a continuous preforming process for Non Crimped Fabrics (NCF). In: The 11th International Symposium of Measurement Technology and Intelligent Instruments, Aachen, 1–5 July 2013
5. Fox, B.: Presentation Swinburne University of Technology (2019). https://worldmaterialsforum.com/files/Presentations2019/WMF2019-PS5-University-of-Swinburne-Prof-Bronwyn-Fox.pdf. Accessed 19 Jan 2022
6. Hanbay, K., Talu, M.F., Özgüven, Ö.F.: Fabric defect detection systems and methods-a systematic literature review. Optik **127**(24), 11960–11973 (2016). https://doi.org/10.1016/j.ijleo.2016.09.110
7. Heieck, F.: Qualitätsbewertung von Faser-Kunststoff-Verbunden mittels optischer Texturanalyse auf 3D-Preformoberflächen. Dissertation, Universität Stuttgart, Stuttgart (2019)
8. Keyence Cooperation: 2D/3D Laser Profiler - LJ-X8000 Series. Edited by Keyence Cooperation (2021). https://www.keyence.com/products/measure/laser-2d/lj-x8000/downloads/. Updated on 26 Jan 2022. Accessed 31 May 2022
9. Kim, D.-H., Choi, D.-H., Kim, H.-S.: Design optimization of a carbon fiber reinforced composite automotive lower arm. Compos. Part B Eng. **58**, 400–407 (2014). https://doi.org/10.1016/j.compositesb.2013.10.067

10. Kumar, A.: Computer-vision-based fabric defect detection: a survey. IEEE Trans. Ind. Electron. **55**(1), 348–363 (2008). https://doi.org/10.1109/TIE.1930.896476
11. Lee, L.K., Liew, S.C., Thong, W.J.: A review of image segmentation methodologies in medical image. In: Sulaiman, H.A., Othman, M.A., Othman, M.F.I., Rahim, Y.A., Pee, N.C. (eds.) Advanced Computer and Communication Engineering Technology. LNEE, vol. 315, pp. 1069–1080. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-07674-4_99
12. Levkine, G.: Sobel and Scharr 3 × 3 and 5 × 5 convolution kernels for image gradient calculations (2020). https://www.hlevkin.com/hlevkin/47articles/SobelScharrGradients5x5.pdf. Accessed 6 May 2022
13. Maass, D.: Progress in automated ply inspection of AFP layups. Reinf. Plast. **59**(5), 242–245 (2015). https://doi.org/10.1016/j.repl.2015.05.002
14. Mahajan, P.M., Kolhe, S.R., Patil, P.M.: A review of automatic fabric defect detection techniques. Adv. Comput. Res. **1**(2), 18–29 (2009)
15. Masood, S., Sharif, M., Masood, A., Yasmin, M., Raza, M.: A survey on medical image segmentation. CMIR **11**(1), 3–14 (2015). https://doi.org/10.2174/157340561101150423103441
16. Meister, S., Wermes, M.A.M., Stueve, J., Groves, R.M.: Algorithm assessment for layup defect segmentation from laser line scan sensor based image data. In: Huang, H. (ed.) Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems, 27 April–9 May 2020, p. 45. SPIE (2020). https://doi.org/10.1117/12.2558434
17. Pourdeyhimi, B., Ramanathan, R., Dent, R.: Measuring fiber orientation in nonwovens. Text. Res. J. **66**(12), 747–753 (1996). https://doi.org/10.1177/004051759606601202
18. Pourdeyhimi, B., Dent, R., Davis, H.: Measuring fiber orientation in nonwovens part III: Fourier transform. Text. Res. J. **67**(2), 143–151 (1997). https://doi.org/10.1177/004051759706700211
19. Römer, M., Bergers, J., Gabriel, F., Dröder, K.: Temperature control for automated tape laying with infrared heaters based on reinforcement learning. Machines **10**(3), 164 (2022). https://doi.org/10.3390/machines10030164
20. Sacco, C., Radwan, A.B., Harik, R.F., van Tooren, M.: Automated fiber placement defects: automated inspection and characterization. In: McClinton, I. (ed.) Proceedings SAMPE 2018. SAMPE 2018 Conference and Exhibition, Long Beach, 21 May 2018
21. Schmitt, R., Niggemann, C., Mersmann, C.: Contour scanning of textile preforms using a light-section sensor for the automated manufacturing of fibre-reinforced plastics. In: Berghmans, F., Mignani, A.G., Cutolo, A., Meyrueis, P.P., Pearsall, T.P. (eds.) Optical Sensors 2008. Photonics Europe, Strasbourg, France, 7 April 2008. SPIE (SPIE Proceedings), vol. 7003 (2008)
22. Veldenz, L.: Automated dry fibre placement and infusion process development for complex geometries. Doctoral, University of Bristol, Bristol (2019)
23. Wehbe, R., Sacco, C., Radwan, A.B., Albazzan, M., Harik, R.: Influence of process parameters in AFP fiber steering on cylinders: constant curvature paths. Compos. Part C Open Access **2**, 100036 (2020). https://doi.org/10.1016/j.jcomc.2020.100036
24. Zhao, Y.-Q., Gui, W.-H., Chen, Z.-C., Tang, J.-T., Li, L.-Y.: Medical images edge detection based on mathematical morphology. In: Conference Proceedings: 2005 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Annual Conference. IEEE Engineering in Medicine and Biology Society, pp. 6492–6495 (2005). https://doi.org/10.1109/IEMBS.2005.1615986

25. Zambal, S., Eitzinger, C., Clarke, M., Klintworth, J., Mechin, P.-Y.: A digital twin for composite parts manufacturing: effects of defects analysis based on manufacturing data. In: 2018 IEEE 16th International Conference on Industrial Informatics (INDIN), Porto, 18 July–20 July 2018, pp. 803–808. IEEE (2018)

26. Zambal, S., Heindl, C., Eitzinger, C.: Machine learning for CFRP quality control. In: SAMPE Europe Conference. Zenodo, Nantes (2019). https://doi.org/10.5281/zenodo.3381930

Parameter Study and Optimization of Forming Simulations for Tape-Based Fiber Layups

Muhammad S. Saeed[1,2]([✉]), Jakob M. J. Gugliuzza[1], Michael Liebl[1], Boris Eisenbart[2], Racim Radjef[2], Peter Middendorf[1], and Matthias Kreimeyer[1]

[1] University of Stuttgart, 70569 Stuttgart, Germany
msaeed@swin.edu.au
[2] Swinburne University of Technology, John Street, Hawthorn, VIC 3122, Australia

Abstract. This work presents a model of a parametrized fiber-reinforced tapes layup for simulations of stamp forming. BETA CAE Systems' ANSA v22.1.0 produced the forming model. The initial configurations of the forming process are simulated using LS-DYNA R12 from the Livermore Software Technology Corporation (LSTC) on the High-Performance Computing Center Stuttgart (HLRS). A doubledome geometry was used as a use case in the presented study. An optimal solution is found by predicting and analyzing the forming defects, both quantitatively and qualitatively. Typical defects induced by the forming process include wrinkles, bridging, waviness, misalignment of tapes, instabilities due to tape overlaps and gaps, etc. Adjustments to the forming process are made by varying model parameters such as stacking sequence, fiber angle, material properties, tape orientation, or the position of ultrasonic thermal bonds between tapes. In addition, process parameters such as tool velocity, temperature, and gripping force were considered. Different algorithms for the sample creation of the designed experiments are examined, and the influence of various parameters on the final part is discussed. The limitations and requirements of the employed software are examined for additional studies. Furthermore, the machine learning-based neural network algorithms discussed, such as particle swarm optimization (PSO), gradient-based optimization (GBO), response surface method (RSM), and genetic algorithms (GA), may be used in future research to determine an ideal tape arrangement for various geometries based on nodal locations and the parameters as mentioned above.

Keywords: Design of Experiment · Parametrization · Fiber Reinforced Thermoplastic · Draping Simulation · LS-Dyna
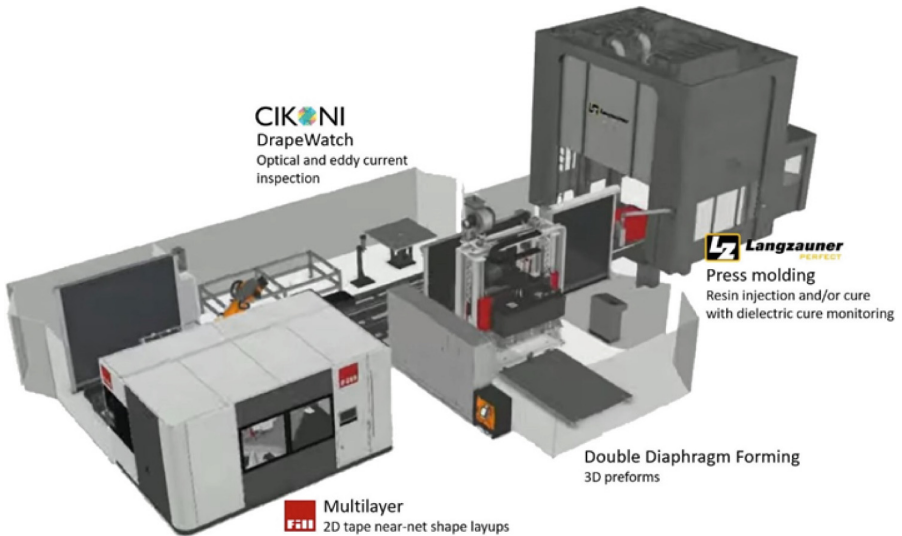
1 Introduction

Carbon fiber-reinforced plastics (CFRPs) play an increasingly important role in the lightweight manufacturing of structural aviation and automotive components [1]. This leads to the growing need for rapid and reproducible mass production of CFRP parts. One way of scaling manufacturing quantity is fully automated production processes. The production line shown in Fig. 1 presented in this research is driven by Swinburne-CSIRO

National Industry 4.0 Testlab as part of a Global Innovation Linkage (GIL) project funded by the Australian Department of Industry. The process chain includes an automated tape laying machine to produce 2D composite laminates, a double diaphragm former (DDF), a 300t press with resin injection capabilities, and a KUKA robot to shuttle the preform between the stations.

Fig. 1. Composite production chain at Swinburne/CSIRO Industry 4.0 Testlab [2]

The industry-scale production line combines a novel multilayer system from FILL for automated tape laying (ATL). The multilayer is relevant to this research project because it offers unprecedented freedom in designing, deposition speed and creating the initial 2D composites laminates. The individual layers are ultrasonically spot welded together, allowing shuttling from the Multilayer to the forming stations. Multilayers can use materials ranging from binder dry-fiber tapes to thermoplastics and thermosets. The spools of the Multilayer can hold tapes of 49.5 mm width [3]. It cuts material waste and manufacturing time [2].

Accurate simulation of the SF process is essential for this work to implement parameterization in forming models. Such simulations require much time and can therefore be very costly. It is also where existing simulation environments are still lacking because of their direct relationship with the manufacturing process. The 2D tape stacks created in the multilayer are the basis for the forming simulations shown here. The SF process is typically used on thermoplastic prepregs and involves a blank holder that presses a flat workpiece against a metal tool. The stamp is then progressively lowered into the mold while being able to be heated during the forming process [4]. The given method uses the model to perform simulations [5–7]. This early work takes SF into account.

The SF process on tape-based layups induces defects such as wrinkles, layer separation, overlaps, and gaps between tapes, especially on thermoplastic tape-based sheets [8–10]. Therefore, optimizing the geometry, stacks of tapes, and process parameters are

essential before forming. Optimization helps assess defects such as wrinkles, bridges, gaps, and overlaps. The optimization process repeats iteratively to ensure optimum structural performance with local drape effects. The iterative process is expensive, which includes material waste or virtual analysis, which causes computational time and accuracy. The finite element method (FEM) applied during the process uses data from actual situations to develop, mimic, and predict forming behavior [11]. In order to support FEM, different Design of Experiment (DoE) algorithms can be used to generate parametrized model simulations automatically [12]. These algorithms, such as Latin hypercube (LHC), full factorial, central composite design, box Behnken, etc., are commonly used in engineering and science to create samples to represent a much greater space of possibilities [13]. This work applies the LHC method to generate sets of parameters for different FE models. LHC aims to minimize the variety of parameter combinations needed to optimize the defect formation during forming.

In order to investigate the deformation behavior during the forming of tape laminates, it is necessary to categorize the deformation mechanisms. As a result, key deformation mechanisms are seen at the ply and laminate levels, which are often connected to mesoscopic and macroscopic approaches. The mechanism of Intra-ply occurs on a single tape of a multilayer laminate during the forming behavior. In comparison, the mechanism of inter-ply represents the deformation at the interfaces between the single tape of the stacked laminate [8–10]. It is necessary to evaluate formability that invariably depends on material properties such as Young's modulus, Poisson ratio, friction coefficient, bending stiffness, etc., and process parameters such as tool velocity, temperature, gripper force, pressure, etc.

2   State of the Art

State of the art in automated CFRP manufacturing uses tape-laying machines, such as the Multilayer in this research, that can build customized 2D tape layups using up to sixteen spools at a time and a rotatable ground plate to allow for different orientations of the tape stripes.

Various solvers exist to solve a FEM model to compute the forming modeled parts' stress, strain, and other responses. This research study is focused on four stagess, as shown in Fig. 2 below. The first stage includes preparing an initial FE model with material properties, initial conditions, and boundary conditions using Ansa v22.1.0 as a preprocessor. The second stage is followed by converting the FE model into the parametrized model by defining the parameters using Ansa inbuilt optimizer. The parameters are thickness, orientation, and position of the tapes, diameter, position, number of spot welds, tool velocity, material type, and gripping force. In the third stage, LS-DYNA R12 from the Livermore Software Technology Corporation (LSTC) is used to solve the FEM models. It is most suited solver for explicit and implicit simulations with many elements and contact conditions. An explicit solver is used in this research because it can solve the complex FEM model simultaneously. In the last stage, the data science approach concludes the results in the final section. It offers a variety of tools that can be applied to structure, cluster, and interpret the data based on mathematics. They range from simple linear regressions over advanced stochastics to machine learning.
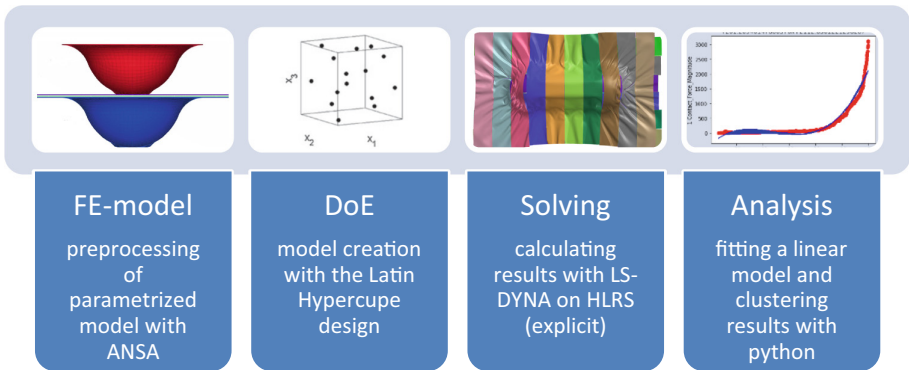
Fig. 2. Scheme of the experiment process

The following section describes the methodologies to build and execute the forming simulations in detail. Section 3 provides information on the setup of the FE models and how they are solved. In Sect. 4, the results are analyzed based on described methods. Section 5 includes the discussion of the results obtained on forming behavior. From that, Sect. 6 presents a conclusion and possible improvements.

3 Methodology

The research aims to enhance a stamp-forming process using simulations. The formed sheet is a tape-based layup as produced by Multilayer. It consists of multiple unidirectional (UD) tapes, making up four layers in total. 2D tapes preform positions between the mold and the blank holder. The blank holder has an opening for the forming tool. As seen in Fig. 3, the forming tool, which serves as the stamp, is positioned above the preform. The blank holder forces the preform against the mold during the forming process with a specific amount of pressure. Then the stamp is heated to the working temperature of the thermoplastic prepreg and lowered into the mold, forming the preform into the desired shape. The mold has a depth of 60 mm, and the blank holder has a size of 340 mm × 520 mm.

The tape design fits the mold's dimensions. Each tape is 500 mm long (longitudinal), 300 mm long (horizontal), and 50 mm wide. The total number of tapes is 32 (20 longitudinal) and (12 horizontal).

There are two basic categories for the evaluation of the simulations. The first is the quality, efficiency, and accuracy of the simulation itself, and the second is the quality of the resultant part of the forming process. Few wrinkles and gaps indicate that the overall forming outcome is satisfactory. In order to define quantitative criteria for evaluating the consequent part, it considers stress responses, displacement, internal energy, curvature, and other results. Thereby the result can be categorized for each criterion in different levels, for example: low, medium, and high. Table 1 gives the list of all requirements. All the qualitative measures must be as low as possible for an excellent forming process to mitigate the defects. It is necessary to run numerous simulations with various
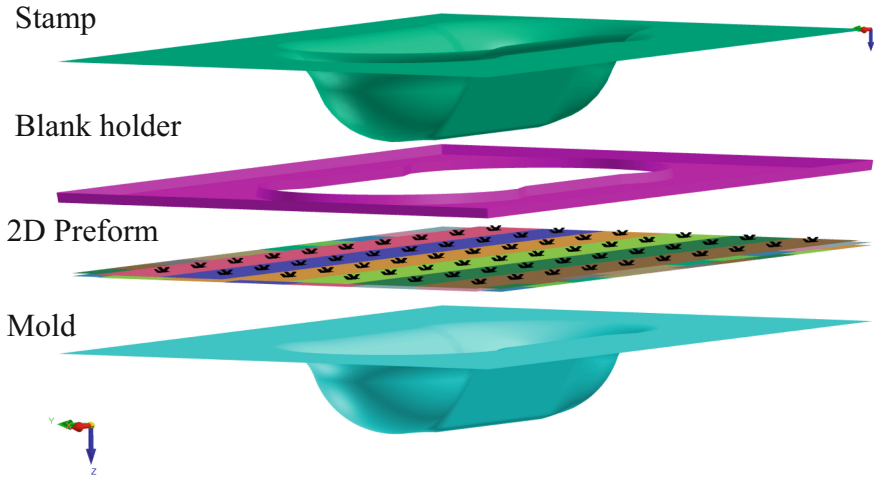
Fig. 3. FE-model of forming process

initial conditions, including material and process characteristics. The results of the corresponding simulations are then analyzed using statistical methods such as correlation and regression. Machine learning techniques assist in the discovery of a suitable SF model.

3.1 Initial Finite Element Model

A dynamic model for an SF process is produced to consider the material's behavior and the formation of defects under predetermined boundary conditions [15]. Forming simulations frequently use flexible thin shell parts [16–19]. Thije and Akkerman found that fully integrated standard elements should align with the fiber direction to avoid contact problems and intra-ply locking [20]. In order to generate a realistic simulation of plastic deformation under stresses beyond the yield stress, detailed information about the plastic behavior, such as effective plastic strain or hardening curves, is needed [21–24]. Explicit solvers are preferable in this sector because formation simulations inherit contact criteria [25–27].

As already seen in Fig. 3, the 2D preform is composed of four layers of tapes. All modeled parts are rigid bodies in this model except for the 2D preform tapes. While the blank holder and the stamp are free to move in the direction of the stamp's motion, the position of the mold is fixed. The rigid body force imparted to the blank holder parametrizes the forming model. The stamp is given a prescribed motion in the form of a displacement that depends on the termination time. Only the contact circumstances can limit the tapes. An anisotropic thermoplastic composite material uses (MAT 249 in LS-DYNA) to model them. The idea of using MAT 249 is to determine the bending stiffness by local integration points. It also defines the anisotropic hyperelastic material with up to three fiber directions. It is practical for this research since it utilizes the UD

Table 1. Criteria for the assessment of the simulation and the results

Type	Criterion	Value for optimal result
Qualitative assessment with two or three levels	Number of gaps	Low
	Size of gaps	Low
	Number of wrinkles	Low
	Size of wrinkles	Low
	Number of overlaps	Low
	Size of overlaps	Low
Quantitative assessment with absolute values	Maximum Stress	Low
	Maximum Strain	Low
	Displacement	Low
	Weight	Low
	Strength	High
	Accuracy	High
	Computational cost	Low

material model and matrix using a straightforward formulation of thermal elastic-plastic material.

Factors: Tools speed, temperature, and the blank holder's gripping force are process parameters. The tooling speed links with the simulation time since the displacement of the stamp are always the same. The geometrical parameters of the model are the position, thickness, and orientation of the tapes and the position and diameter of spot welds. Property-related parameters are the properties of the used materials. In order to keep things simple, the model considers only symmetrical cross-plies at various angles and symmetrically balanced angle plies concerning the longitudinal symmetry of the mold.

Design of Experiment: The tapes mesh as shell elements using an automated approach. Their geometric parameters are considered controllable variables for the DoE. They include the initial gaps between tapes, their thickness, orientation, and the tapes' horizontal position relative to the symmetry center. The width of the tape is kept constant at 50 mm to explore the limitation of the multilayer. Furthermore, the position and quantity of the spot welds are considered controlled variables, and beam elements model the spot welds. Morphing parameters are applied to parametrize the position, thickness, and orientation of the tapes and the spot welds for the DoE. Controllable process parameters are the selection of the material type for the tapes, the gripping force of the blank holder, and the force for the static test after the forming process. Simulation parameters are the simulation's termination time and the tapes' mesh size. The optimization tool used

during the DoE utilizes polyamide (PA), polycarbonate (PC), and polyetheretherketone (PEEK) tape materials.

The DoE approach uses the controlled design parameters for sampling as in Table 2 [28]. The derived approach of LHC determines the most influential factors on the output response. Some of the factors occur multiple times for different layers of tapes. LHC minimizes the number of experiments needed while covering most inference space [29]. It enhances the possibility of not only post-process the geometries but also reverse-engineering the parameters defined.

Table 2. Design variables for the DoE as factors with corresponding levels

	Factors	Levels	Unit
Geometrical	Thickness of tapes	0.14, 0.17, 0.3	mm
	4 × Position of tapes	−12.5, 0, 12.5	mm
	Type of layup	Angle-ply, Cross-ply	–
	Orientation of tapes	0, 15, 30, 45, 60, 75, 90	°
	15 × Number of spot welds	2, 4, 10	–
	3 × Position of spot welds	−9, −6, −3, 0, 3, 6, 9	mm
	Spotweld diameter	2.25, 2.5, 2,75	mm
Process	Material ID	1, 2, 3	–
	Gripping force	5, 10, 15	N
Simulation	Termination time	0.001, 0.002, 0.003	s
	Mesh size	2, 3, 4	mm

3.2 Parametrization Setup

The materials used for the multilayer determine the tapes' thickness (d). The chosen tapes are 0.14 mm, 0.17 mm, and 0.3 mm thick, as seen in Fig. 4.

d = 0.14 mm   d = 0.17 mm   d = 0.3 mm
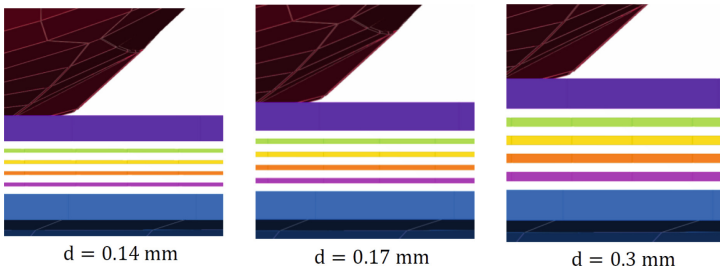
Fig. 4. Position of different tape thicknesses

Moving many layers of tape in the transverse direction is accomplished using the parameters for the tape's position, as demonstrated in Fig. 5. Each layer has its defined parameter. For layer orientation, one parameter sets the angle of the outer layers, and one sets the angle of the interior layers. In a cross-ply layup, the internal layers have an angle 90° different from the outer layers. Suppose the layup is a cross-ply. The inner layers have an angle shifted by 90° for the outer layers. If the layup is an angle-ply, the inner layers' orientation differs from the outer layer's, leading to a balanced layup. The angles 0°, 45°, and 90° are not included in angle-plies since they would produce another cross-ply layup.

Before morphing   Cross-ply   Angle-ply at 30°
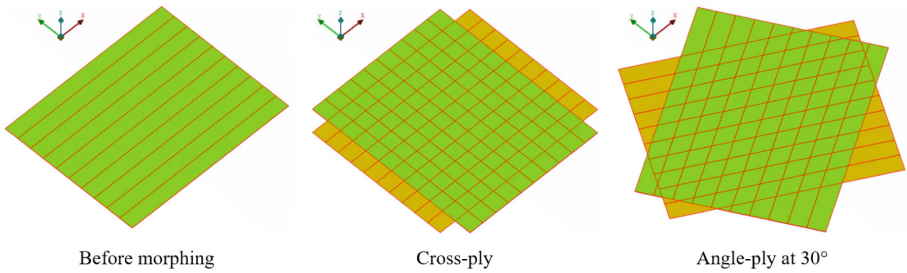
Fig. 5. 2D preform with different tape angles

The number of spot welds is defined for each layer of the tapes, as shown in Fig. 6. The number of spot welds per tape is determined separately for each symmetrical pair of tapes. Only one connection between each pair of tapes should be between the inner and outer layers. Tape length definition models the spot weld as stated in Sect. 3. Each tape initially has ten spot welds, modeled at a distance of 50 mm from each other. An absolute value of 3 mm, 6 mm, or 9 mm shift occurs in the position of the spot welds. The spotweld strength is affected by the use of cold forming. The initial weld strength has no bearing on hot forming.
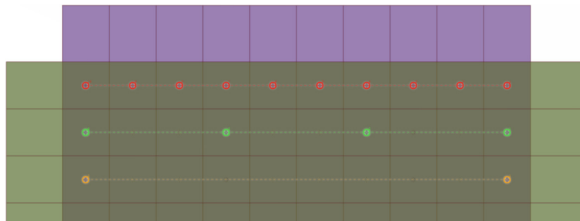
Fig. 6. Spotwelds modeling on tapes laminate

3.3   Analysis

We analyze the data with a Python script. The lasso. Dyna library reads the binout and d3plot files produced by LS-DYNA. Without loading the entire dataset, it is feasible

to extract arrays containing the results for each node, shell element, or beam element using this library. It is also possible to plot the results of the forming simulation. The output log contains detailed information about the execution of the simulation, such as the used computational resources and timing information for the individual tasks of the simulation, e.g., element processing, contact algorithm, etc. Typical forming defects, such as wrinkles and gaps between the tapes, are evident. Simulation results conclude the desired mean, variance, and maximum absolute values of shell and beam elements such as bending moment, stress, shear, and axial force, as listed in Table 3.

Table 3. Single value results

	Variance	Mean	Max. Absolute
Shell bending moment	$var(M_b)$	$\overline{M}_b$	$max(|M_b|)$
Shell shear force	$var(F_\tau)$	$\overline{F}_\tau$	$max(|F_\tau|)$
Shell stress	$var(\sigma)$	$\overline{\sigma}$	$max(|\sigma|)$
Beam axial force	$var(F_{ax})$	$\overline{F}_{ax}$	$max(|F_{ax}|)$
Beam bending moment	$var(M_{bb})$	$\overline{M}_{bb}$	$max(|M_{bb}|)$
Beam shear force	$var(F_{b\tau})$	$\overline{F}_{b\tau}$	$max(|F_{b\tau}|)$
Beam torsional moment	$var(M_t)$	$\overline{M}_t$	$max(|M_t|)$

The linearized model chosen to predict the experiments is the quadratic approximation of the form shown in Eq. 1. This model has been chosen over the linear approximation because most DVs have at least three levels, making the model more suitable than the linear approximation. In this case, the linear term and the square term led to linear dependent columns in the factor matrix, which is why the second order is the best guess:

$$Y = \beta_0 + DV\beta_1 + DV^2\beta_2 + \epsilon = \left(1|DV|DV^2\right)(\beta_0|\beta_1|\beta_2) + \epsilon = X\beta + \epsilon \quad (1)$$

The calculation of parameter matrix β minimizes the prediction error ϵ. It calculates the error using Eq. (2) as described by Webers [28]. The error is the difference between the simulated and the predicted results. In order to calculate the result matrix, we use the previously calculated single values in Table 3. There are 21 single values for every simulation, leading to a response matrix of size 500×21.

$$\Rightarrow \beta = \left(X^T X\right)^{-1} X^T Y \quad (2)$$

The K-Nearest Neighbors (KNN) approach is applied to the coordinates of each node at the most recent time step to analyze how similar the resulting geometry is to the original geometry. The nodes of each simulation are geometrically sorted, subsampled to the size of the smallest node array, then flattened to produce collections of the same size for each simulation [30]. In order to use a dimensionality reduction technique, the KNN method generates connectivity- and distance matrix that identifies groups of simulation results compared by Tenenbaum [31].

4   Results

The results shows that the ply-by-ply tape-based model, created by the combination of parameters, precisely replicates the multilayer ATL process. An analysis of relationships between the tape stack, forming models and the design factors shows a 3D scatter plot for each simulation. Figure 7 displays the maximum absolute values of the shell stress, shell shear force, and shell bending moment for the tape thickness. The majority of point cluster concludes that the thicker tape tends to have a lower maximum absolute stress and lower stress variance during forming. The conclusion is that choosing a thinner tape is an ideal option to reduce the shear forces and the bending moment.
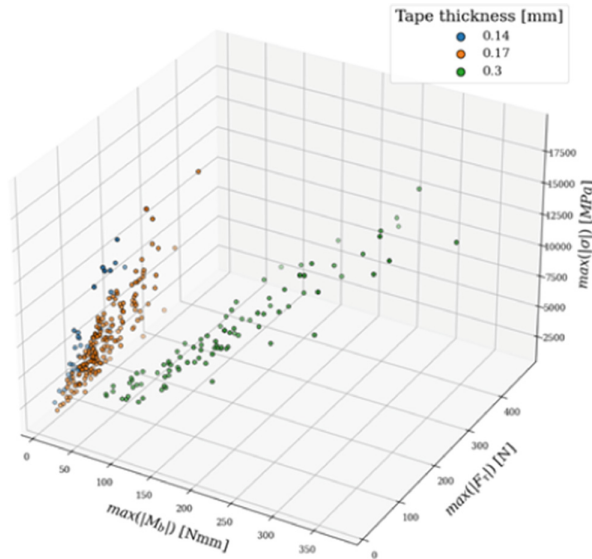
Fig. 7. Max. Abs. Shell responses for different tape thicknesses

Figure 8 compares a simulated geometry with similar design parameters to a created real geometry. Both parts display the recognizable wrinkles that develop during forming. This occurs because the spot welds link one tape on an upper layer to numerous tapes on the layer beneath. These tapes thus have more clearance as they are not perpendicular, resulting in the formation of gaps. The fringe pattern for the bending moment highlights the wrinkles in the simulation result. Both results also show a gap between the two middle tapes on the top layer. In particular, simulations of angle-plies with a 15° tape angle have shown severe gap creation, as shown in Fig. 9.

For both, more minor wrinkles appear along the double dome geometry's edge, while more prominent ones appear on the outside tapes.

The quadratic function discussed in Sect. 0 correlates with the maximum absolute cross, and angle plies values. A 45° angle lowers the maximum absolute values for cross-plies, and a 75° angle decreases the maximum values for angle-plies. The Pareto
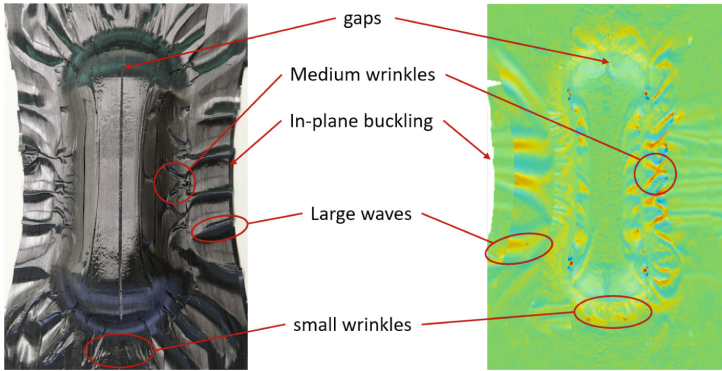
Fig. 8. SF rig results (left), simulation result (right)

plots created using the quadratic function, as shown in Fig. 10, corroborate the relationships between the tape thickness, the material ID, and the termination time. The relative root square mean error (rRMSE) for each value results assess the prediction function's accuracy. In order to compare multiple design parameters, each design variable is normalized. It shows the predicted effect of different design variables on the max. Abs. Bending moment.


Fig. 9. Gap formation on angle-ply with 15°

5   Discussion

The purpose of the research presented in this paper is to examine the thermoplastic UD tapes' limits. Due to the limitations of the available SF rig, the temperature is not considered during the simulations. The rig is capable of producing temperatures up to 160 °C. Utilizing a material with a low melting temperature is advised. Polycarbonate was the UD thermoplastic material employed in this study. The glass transition temperature of the tape is 140 °C. As a result, the SF process is carried out at a temperature above the glass transition but below the melting point. Implementing the parameterization concept in the same forming simulation framework brings novelty to this research.

Polycarbonate tapes give the optimal forming results among the simulated materials under defined current boundary conditions. Once the temperature is included and forming is done at or above melting temperature, the model will get different results. It means that the resin will not be able to flow through the tapes, and the model will have a variety of flaws such as wrinkles, gaps, buckling, etc. The forming of ultrasonically spot-welded tapes at room temperature causes the defects such as wrinkles to emerge.
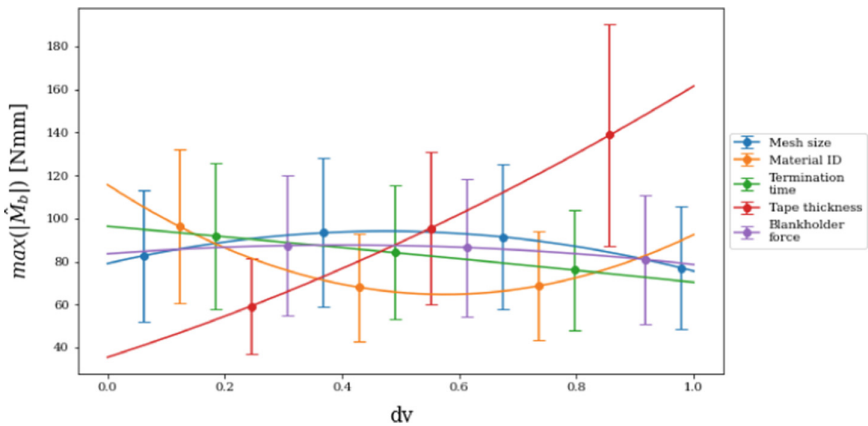
Fig. 10.   Pareto plot of max. Abs. Bending moment for different normalized design variables

If the forming occurs above the melting temperature, the weld spots melt. The spot welds are essential to this process and impact the wrinkles because of the SF rig's limitations. Angle-plies with a 15° tape angle are more likely to cause gaps due to the different arrangement of spot welds. Thinner tapes seem to mitigate the defects independent of the material due to the higher flexibility of the tapes, which leads to lower bending moments. Higher termination times and smaller mesh size come at higher computational costs but most likely lead to fewer defects. The moving parts, such as the stamp and the blank holder, have lower inertial forces.

The considerable variance between the model parameters leads to the high error of the quadratic estimation. It means that the function cannot be used to predict the results but only to get a tendency of the impact of the parameters on an outcome. Results with fewer defects due to the tape angle explain the doubledome geometry's curvature.

An optimal angle is achieved with double curvatures on most tapes, leading to lower intra-ply shear stress. It seems to be the case for the doubledome geometry at 45° for cross-plies and 75° for angle-plies. The analysis with the KNN method failed. It can result from the high variation of the simulated models, leading to many differences in the results. This theory needs to be disproved by additional investigation.

6   Conclusion and Outlook

The model has successfully implemented individual tapes and a tape stack. Models are also used to generate ultrasonically welded tapes that are not pre-consolidated. The double dome model successfully predicts the defects when forming below the melting temperature. Current results show that thermoplastics tapes that are only spot welded are not formable without creating defects. The model and the simulations need further development to include the temperature and material card. Designing and simulating parametrized models for forming simulations with ANSA and LS-DYNA is possible. The bending stiffness and spot weld strengths depend highly on the temperature, and the material formed at the appropriate temperature should behave very similarly. The conclusion is that the forming results mainly depend on the fiber angles. Wrinkles, gaps, and overlaps are predicted with the simulations as observed by Vanclooster [15] and by Hamila and Boisse [19]. Simulations with higher bending moments have prominent wrinkles.

Creating and analyzing the parameter setup for the ultrasonic spot welds on the tapes is challenging. It is improved by assessing the layup for possible spotweld positions and activating or deactivating each spotweld. Testing other connection kinds, such as cohesive bonding, is also possible. For improved simulation accuracy and to take temperature into account during the forming process, the model can be expanded to incorporate more complex material models and a thermal solver [23]. It will result in a better understanding of the forming defects. These defects can be analyzed in more detail to prove that the fiber angles and their respective spotweld positions are significant concerns. Additionally, it is possible to explore the forming behavior using various tape widths and geometries.

The linearized model is helpful in understanding trends in the correlations between the design factors and the results, but it does not correctly represent the simulated findings. Additionally, different DoE algorithms can be used and compared to the LHC [12]. The Isomap dimensionality reduction with the KNN algorithm was inconclusive. Various machine learning algorithms such as convolutional neural network (CNN) could be applied to images of the results or on geometrically sorted arrays of shell responses. A user interface on top of the python script can simplify the search for simulations with specific design variables and results. The design variables may be iteratively optimized using PSO, GBO, RSM, and GA algorithms.

References

1. m. s. f. De: Der Composites-Markt Europa: Marktentwicklungen, Herausforderungen und Chancen. https://composites-united.com/wp-content/uploads/2020/01/GER_CUeV_Marktbericht_2020_CFCap.pdf

2. Swinburne-CSIRO National Industry 4.0 Testlab. https://www.swinburne.edu.au/research/platforms-initiatives/industry-4-0-testlab/. Accessed 27 May 2022

3. Fill GmbH, Multilayer. https://www.fill.co.at/de/aerospace/composites/tailored-blanks/multilayer/pdf/1526p572i50.html. Accessed 14 Mar 2022

4. Hineno, S., et al.: Fiber deformation behavior during press forming of rectangle cup by using plane weave carbon fiber reinforced thermoplastic sheet. Procedia Eng. **81**, 1614–1619 (2014). https://doi.org/10.1016/j.proeng.2014.10.199

5. Potter, K., Ward, C.: Draping processes for composites manufacture. In: Advances in Composites Manufacturing and Process Design, pp. 93–109. Elsevier (2015)

6. Mallon, P.J., O'Brádaigh, C.M., Pipes, R.B.: Polymeric diaphragm forming of complex-curvature thermoplastic composite parts. Composites **20**(1), 48–56 (1989). https://doi.org/10.1016/0010-4361(89)90682-4

7. Chalivendra, V., Beese, A.M., Berke, R.B.: Mechanics of Composite, Hybrid and Multifunctional Materials, Fracture, Fatigue, Failure and Damage Evolution, Volume 3. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-86741-6

8. Sun, S., Han, Z., Fu, H., Jin, H., Dhupia, J.S., Wang, Y.: Defect characteristics and online detection techniques during manufacturing of FRPs using automated fiber placement: a review. Polymers **12**(6), 1337 (2020). https://doi.org/10.3390/polym12061337

9. Mazlan, N., Sapuan, S.M., Ilyas, R.A. (eds.): Advanced Composites in Aerospace Engineering Applications, 1st edn. Springer. Cham (2022). https://doi.org/10.1007/978-3-030-88192-4. https://link.springer.com/content/pdf/10.1007%2F978-3-030-88192-4.pdf. Accessed 29 Mar 2022

10. Akkerman, R., Haanappel, S.P.: Thermoplastic composites manufacturing by thermoforming. In: Boisse, P. (ed.) Advances in Composites Manufacturing and Process Design, pp. 111–129. Elsevier (2015)

11. Vanclooster, K., Lomov, S.V., Verpoest, I.: Simulating and validating the draping of woven fiber reinforced polymers. Int. J. Mater. Form. **1**(S1), 961–964 (2008). https://doi.org/10.1007/s12289-008-0217-7

12. Cimbala, J.M.: Taguchi orthogonal arrays (2014). https://www.me.psu.edu/cimbala/me345web_fall_2014/lectures/taguchi_orthogonal_arrays.pdf

13. Vořechovský, M.:Extension of sample size in Latin Hypercube Sampling with correlated variables (2010)

14. BwUniCluster 2.0 Hardware and Architecture - bwHPC Wiki. https://wiki.bwhpc.de/e/BwUniCluster_2.0_Hardware_and_Architecture. Accessed 26 May 2022

15. Vanclooster, K., Lomov, S.V., Verpoest, I.: Experimental validation of forming simulations of fabric reinforced polymers using an unsymmetrical mould configuration. Compos. Part A Appl. Sci. Manuf. **40**(4), 530–539 (2009). https://doi.org/10.1016/j.compositesa.2009.02.005

16. Chen, B., Govindaraj, M.: A parametric study of fabric drape. Department of Textiles and Apparel, Cornell University, Ithaca (1996)

17. Chen, B., Govindaraj, M.: A physically based model of fabric drape using flexible shell theory. Text. Res. J. **65**(6), 324–330 (1995). https://doi.org/10.1177/004051759506500603

18. Davey, S., Das, R., Cantwell, W.J., Kalyanasundaram, S.: Forming studies of carbon fibre composite sheets in dome forming processes. Compos. Struct. **97**, 310–316 (2013). https://doi.org/10.1016/j.compstruct.2012.10.026

19. Hamila, N., Boisse, P.: Simulations of textile composite reinforcement draping using a new semi-discrete three node finite element. Compos. Part B Eng. **39**(6), 999–1010 (2008). https://doi.org/10.1016/j.compositesb.2007.11.008

20. ten Thije, R., Akkerman, R.: Solutions to intra-ply shear locking in finite element analyses of fibre reinforced materials. Compos. Part A Appl. Sci. Manuf. **39**(7), 1167–1176 (2008). https://doi.org/10.1016/j.compositesa.2008.03.014

21. LSTC: Volume I Keyword. https://www.dynasupport.com/manuals/ls-dyna-manuals/ls-dyna_manual_volume_i_r12.pdf

22. LSTC: Volume II Materials. https://www.dynasupport.com/manuals/ls-dyna-manuals/ls-dyna_manual_volume_ii_r12.pdf

23. Stommel, M., Stojek, M., Korte, W.: FEM zur Berechnung von Kunststoff- und Elastomerbauteilen, 2nd edn. Hanser, München (2018). https://www.hanser-elibrary.com/doi/book/10.3139/9783446452831

24. Dangora, L.M., Mitchell, C., White, K.D., Sherwood, J.A., Parker, J.C.: Characterization of temperature-dependent tensile and flexural rigidities of a cross-ply thermoplastic lamina with implementation into a forming model. Int. J. Mater. Form. **11**(1), 43–52 (2016). https://doi.org/10.1007/s12289-016-1327-2

25. Dong, L., Lekakou, C., Bader, M.G.: Solid mechanics draping simulation of woven fabrics. School of Mechanical and Materials Engineering, The University of Surrey, Guildford (1999)

26. Boisse, P., Borr, M., Buet, K., Cherouat, A.: Finite element simulations of textile composite forming including the biaxial fabric behaviour. Compos. Part B Eng. **28**(4), 453–464 (1997). https://doi.org/10.1016/S1359-8368(96)00067-4

27. Boisse, P., Gasser, A., Hivet, G.: Analyses of fabric tensile behaviour: determination of the biaxial tension–strain surfaces and their use in forming simulations. Compos. Part A Appl. Sci. Manuf. **32**(10), 1395–1414 (2001). https://doi.org/10.1016/S1359-835X(01)00039-2

28. Weber, D., Skillings, J.H.: A First Course in the Design of Experiments: A Linear Models Approach. CRC Press, Boca Raton (2000)

29. Viana, F.A.C.: A tutorial on Latin hypercube design of experiments. Qual. Reliab. Eng. Int. **32**(5), 1975–1985 (2016). https://doi.org/10.1002/qre.1924

30. Vajda, S., Santosh, K.C.: A fast k-nearest neighbor classifier using unsupervised clustering. In: Santosh, K., Hangarge, M., Bevilacqua, V., Negi, A. (eds.) Recent Trends in Image Processing and Pattern Recognition, vol. 709, pp. 185–193. Springer, Singapore (2017). https://doi.org/10.1007/978-981-10-4859-3_17

31. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science **290**(5500), 2319–2323 (2000). https://doi.org/10.1126/science.290.5500.2319

Analysis of New Concepts for the Consolidation Roller in Laser-Assisted Automated Tape Placement Processes

Nils Widmaier[✉] and Lukas Raps

DLR, Institute of Structures and Design, Pfaffenwaldring 38-40, 70569 Stuttgart, Germany
`nwidmaier@swin.edu.au`

Abstract. The need for more sustainable production has led to an increased popularity of thermoplastic fibre reinforced composites over the last years. Typical production processes rely on autoclave and/or sequential processes, which lead to long cycle times, high energy consumption and high costs. One step in-situ production processes, like the thermoplastic laser assisted automated fibre placement, provide an ideal solution to these challenges while enabling the benefits of circular economy through improved recyclability. An essential part of the production process is to ensure full consolidation during layup. One of the main influences on consolidation is the applied pressure of the compaction roller. Increasingly complex part geometries with curved surfaces place special demands on the deformation and adaptability of these rollers. Here, current solutions quickly reach their limits. This paper investigates new concepts for the consolidation roller to enable successful use of in-situ placement technologies on complex part geometries. Different sheath thicknesses and materials were investigated in experiments, followed by simulative investigation of further compaction roller concepts.

Keywords: Automated fibre placement (AFP) · compaction roller · in-situ consolidation · compaction pressure

1 Introduction

1.1 Thermoplastic In-Situ AFP

Driven by the need for lighter and more fuel-efficient vehicles, more industries rely on the use of fibre composite materials with their excellent strength-to-weight ratio [1]. Other beneficial properties that make the use of these materials even more relevant are their high corrosion resistance, good fatigue properties and high damage tolerance [2]. Airbus and Boeing have implemented fibre composites in their next generation airplanes with a share of around 50% in the A350 XWB and B787 Dreamliner [3]. This increased demand for fibre composites is calling for new and improved manufacturing technologies from traditional hand-layup to fully automated and reproducible processes. Furthermore, traditionally used thermoset matrices often require autoclave cure with long cycle times and size limitations [2]. One production method that promises to solve these issues is

the laser-assisted automated fibre placement for the in-situ processing of thermoplastic composites [4]. This one-step production technique allows for an autoclave free process by using laser heating to achieve full consolidation of the matrix during tape layup. Full consolidation not only requires heat but also pressure, which is provided via a compaction roller.

1.2   Compaction Roller Overview

Compaction pressure is one of the most important process parameters regarding the final product quality [5]. Khan et al. found that higher compaction pressures lead to a better overall quality of the laminate but can also result in high pressure air inclusions that lead to deconsolidation if the material is not cooled below glass transition temperature under the compaction roller [6]. Raps et al. found that on a flat surface a reduction in compaction pressure leads to a reduction in interlaminar shear strength [7]. Chu et al. defined two requirements for successful interfacial bonding. Those are full contact between the compaction roller and tape and an even pressure distribution. It was shown that a lower Young's modulus and higher outer diameter of the compaction roller led to an increased deformability and therefore adaptability [8]. This is important as traditional compaction rollers out of metal are not flexible enough to conform to tool surfaces with radii smaller than the tool. This missing contact can lead to a reduction in mechanical properties [9]. Flexible compaction rollers further offer the benefit of compensating misalignment between the roller and the tool surface. Already at 1° misalignment, Schledjewski et al. found 63% higher peel forces for an adaptive roller compared to a non-adaptive roller [4]. Typical materials for flexible compaction rollers are Polytet-rafluorethylen [6] or Polyurethan [10]. Jiang et al. investigated silicone compaction rollers and found more even pressure distribution for lower hardness, higher outer diameters, and smaller widths. A porous structure was proposed to optimize flexibility and deformability [11]. This approach was investigated by Bakhshi et al. with different compaction rollers made from Polyurethan. It was found that holes through the compaction roller (over the width of the roller) lead to deviations of up to 50% in compaction pressure. A full material compaction roller with a Shore A hardness of 35 achieved the best results for different tool radii [12]. Lichtinger et al. investigated a compaction roller made from a thermoset foam material with a shrink tube sheathing and analysed pressure distribution on convex and concave tools. Full contact was not achieved for concave tools with a lateral orientation of the compaction roller. Pressure peaks were noted at the edges of the roller (concave tool) and in the middle (convex tool) [13]. This highlights that the compaction roller and tool selection must be coordinated. He et al. found slightly improved pressure uniformity for a compaction roller with 11 segments compared to a common roller [14].

This work aims to find new concepts for compaction roller designs to enable the use of the laser assisted thermoplastic in-situ tape placement processes on curved surfaces. The goal is to increase the understanding of how compaction rollers behave on curved surfaces by investigating pressure uniformity and compaction pressure. To achieve this goal, various compaction roller configurations with three different sheath thicknesses made from two silicones are analysed experimentally and perforated compaction rollers and compaction roller with increased width are investigated by means of simulation.

2   Methodology

2.1   Analytical Analysis of Compaction Roller Contact and Deformation

The deformation of the compaction roller plays a significant role for its pressure distribution. The necessary deformation for full contact depends on the distance between the roller and the tool. This distance is a function of the roller diameter, roller length, surface geometry and orientation. In this work, especially cylindrical surface geometries are of interest. A 0° orientation of the compaction roller to the cylindrical tool is defined with the two cylindrical axes perpendicular to each other (see Fig. 1, centre). The coordinate system rotates with the compaction roller and the X-axis parallel to the compaction roller axis. The origin is set in the middle of the roller width and length at the point of first contact with a convex tool for symmetrical contact. The turning angle α is defined as the angle between the Y-axis and the tool cylindrical axis (tool centre line).
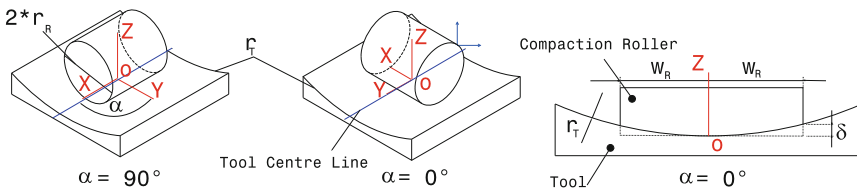
Fig. 1. Overview coordinate system, orientation, and analytical deformation

For two cylinders, the distance δ is equal to the necessary deformation for full contact between both, over the length of the compaction roller. The maximum distance for concave or convex cylinders is equal but can be found either on the side of the compaction roller (convex tool) or at the centre (concave tool) at a turning angle of 0°. The deformation at a specific location (x/y) can be described with Eq. (1).

$$\delta = r_R - \sqrt{r_R^2 - y^2} \pm (r_T - \sqrt{r_T^2 - (\cos(\alpha)^* x + \sin(\alpha) * y)^2}$$
$$\text{with } 0 < \alpha \le 90 \begin{cases} x : -w_R \le x \le w_R \\ y : -r_R \le y \le r_R \end{cases} \tag{1}$$

The tool radius is r_T, the compaction roller radius is r_R and its half width is w_R. The change of sign represents a convex tool for "+" and a concave tool for "−". In the case of a concave tool surface with a turning angle of alpha unequal 90°, negative distances will arise. These negative values represent the already required deformation of the roller or theoretical penetration into the tool surface.

Non-adhesive contact between solid, elastic bodies was first described by Heinrich Hertz [15]. In his considerations, Hertz mostly referred to elastic half-surfaces in contact with a spherical surface [16]. Hertz theory can, with some adjustments, be applied to the contact between two cylinders. The perpendicular contact between two convex cylinders can be represented by two ellipsoids and calculated with the two half-axes of the ellipsoid. Atanackovic et al. formulated the contact between two cylinders in their longitudinal axis according to Hertz (1892) and Johnson (1987). In these considerations it is assumed

that the pressure surfaces in contacted pressure areas are small compared to the bodies in contact. In addition, the contact is idealised as frictionless [17]. The maximum contact pressure p_0 can be described using Eq. (2) with "B" as the contact width, "F" the applied force, E^* the effective Youngs modulus, "r" the radii, "v" the Poisson ratios and "E" the Youngs modulus of cylinder 1 and 2 in contact.

$$p_0 = \frac{2}{\pi} * \frac{p}{b_0} \text{ with } b_0 = \sqrt{\frac{4}{\pi} * \frac{1}{E^*} * p * \frac{r_1 * r_2}{r_1 + r_2}} \text{ and } p = \frac{F}{B}$$
$$\frac{1}{E^*} = \frac{1 - v_1^2}{E_1} + \frac{1 - v_2^2}{E_2}$$

(2)

2.2 Compaction Rollers Fabrication and Experimental Test Setup

All new investigated roller concepts have an outer diameter of 80 mm and width of 30 mm. Three different rim diameters are used (30, 60 and 70 mm) to create sheath thicknesses of 25, 10 and 5 mm respectively. The material of choice was silicone, as it allowed for the necessary flexibility after curing, is heat resistant and easy to pour into the mould. Two silicones were used in this work. The first silicone was "SF45-RTV2" by the company "Silikonfabrik" with an uncured viscosity of 8500 mPas and a Shore A hardness of 45. The second silicone, "Smooth-Sil 960" has an uncured viscosity of 30000 mPas and a Shore A hardness of 60. In addition to pure silicone sheaths, compaction roller with 5wt% short glass fibres "Vetrotex" embedded in SF45 silicone were produced. Figure 2 shows three representative compaction rollers.

25 mm sheath	10 mm sheath	5 mm sheath
SF45	SS960	SF45 5wt% glass fibres

Fig. 2. Example pictures of compaction rollers with different materials and sheath thickness

All 9 different compaction rollers variants were quasi-statically tested in a universal testing machine from "Zwick" by pressing them against seven different tool surface geometries. The tools were chosen based on a fuselage half shell demonstrator and had a cross-section of 120×120 mm^2 and a surface finish of Rz = 4. An overview of the tools can be seen in Table 1. A test speed of 500 mm/min and a load of 547 N (current standard setting at the AFP machine) were chosen for the static tests. Before the experiments, each compaction roller was loaded with 547 N and rotated axially in 10° steps to account for the initial loss of stiffness in the silicone. All tests were performed with compaction roller orientations in 0°, 45° and 90° for each of the 7 tools with a half-inch-wide tape (Toray Cetex TC1225, Standard Modulus Carbon 145 gsm UD Tape) between compaction roller and tool surface, and without tape. A polyimide-film-sandwich based (KAPTON® - DuPont) electronic pressure mat "Tactilus Typ H,

$4'' \times 4'''$ with 32×32 quadratic piezoresistive elements from "Tiedemannn Instruments GmbH & Co.KG" was used to measure contact pressure one second after reaching 547 N.

Table 1. Overview tool geometries for static testing of compaction rollers

Geometry	Name	Geometry	Name
Cylindrical, concave, R = 250 mm	250	flat	flat
Cylindrical, concave, R = 500 mm	500	Double curved	dc
Cylindrical, concave, R = 1000 mm	100	Cylindrical, convex, R = 250 mm	konv
Cylindrical, concave, R = 2000 mm	2000		

The uniformity of the pressure distributions was evaluated by averaging the row of sensors in the middle of the contact length over the width of the compaction roller without a tape. If an equal number of sensors were contacted over the contact length, the two middle rows were averaged. The contact length is defined perpendicular to the roller axis (direction Y, see Fig. 1), the contact width parallel to the roller axis (direction X, see Fig. 1). The coefficient of variation was formed over the respective areas. To allow for an easier evaluation, the pressure uniformity u is displayed according to Jian et al. [11] by subtracting the coefficient of variation from "1". This means that values of "1" represent a perfect even pressure distribution and declining values present a decline in uniformity. p_i represents the pressure of the individual cell and p_{avg} the average pressure of all cells in the measuring area.

$$u = 1 - \sqrt{\frac{1}{2}\sum_{i=1}^{n}\left(\frac{p_i - p_{avg}}{p_{avg}}\right)^2} \tag{3}$$

The compaction rollers were also evaluated based on the magnitude of the compaction pressure in experiments with tape. The pressure was averaged over the contact area of the tape in the middle of the contact length.

2.3 Simulation Setup

A FEM model was setup in the software "Ansys Workbench v19.2" to replicate the deformation behaviour of the silicone roller and to explore more complex, perforated compaction roller geometries and a roller width extension from 30 mm to 60 mm. The compaction roller (silicone sheath and aluminium rim), aluminium tool and if applicable the carbon fibre UD tape were replicated in the simulation (See Fig. 8). The load of 547 N was introduced though the rim. Boundary conditions were set to allow only vertical movement of the compaction roller and to fix the tool in all degrees of freedom. The friction coefficients were set to 0.3 for silicone – tape, 0.8 for silicone – tool and 0.1 for silicone – rim. It was assumed that the tape is incompressible and is fixed on the tool. The material parameters of the carbon fibre tape were chosen based on available manufacturer data (Toray Cetex TC1225, Standard Modulus Carbon 145 gsm UD Tape) with a modulus

in 0° of 135 GPa (tension) and 124 GPa (compression) and a ply thickness of 0.14 mm. Solid185 elements were used to model the hyperelastic five parameter Mooney-Rivlin material model for the two silicones (parameters see Table 2). The required stress-strain material data for the SF45 and SS960 silicones was obtained using unidirectional tensile tests on dog-bone shaped specimens according to DIN 53504. The specimens were stretched three times to 100% strain. Both silicones showed a stiffness reduction after the first loading cycle with nearly constant stiffness afterwards (see Mullins effect [18]). The maximum strain observed in the simulation was around 60% with localised strains of up to 80%. This required scaling of the obtained experimental material data (acquired at 100% strain) for better agreement of the simulation results to the experiments. The material data was scaled to a maximum strain of 60%. Only one material dataset for each silicone was used for the three different sheath thicknesses which led to higher deviations for thin sheath thicknesses (See: **Validation of Simulation Against Experiments**). Deformation strain measurements of the compaction rollers in experiments were not performed but only estimated based on the vertical movement of the rim and compression of the compaction roller.

Table 2. Mooney-Rivlin material parametres for simulation model

	C10 [MPa]	C01 [MPa]	C20 [MPa]	C11 [MPa]	C02 [MPa]	D1 [1/MPa]
SF45	−0,034	0,361	0,8316	−1,672	0,7987	0
SS960	−1,239	1,83	6,005	−10,14	4,782	0

Validation of Simulation Against Analytical Model

The general setup of the simulation model was verified against the analytical calculations. To achieve this, the parallel contact between two convex cylinders was modelled in ANSYS with an isotropic material and compared to Hertz theory. As Table 3 shows, the theory of Hertz and the simulations show better agreements for rollers with thicker sheaths. This is in accordance with the expected results as the theory of Hertz assumes full material sheaths. The basic setup of the simulation can be seen as validated with respect to the analytical model for sheath thicknesses of 25 mm.

Table 3. Comparison analytical model to simulation

	$p_{0\_Ana}[MPa]$	$p_{0\_Sim}[MPa]$	$\Delta p_{0\_Sim\_to\_ana}[\%]$
5 mm Sheath	0,63	0,90	29,64
10 mm Sheath		0,74	14,85
25 mm Sheath		0,65	2,04

Validation of Simulation Against Experiments

The simulation model was validated for both material (SF45, SS960) models by comparing the contact length, contact width, vertical movement of the rim, pressure distribution and magnitude in the middle row of the contact length (below the compaction roller axis) over the whole contact width on the flat and concave tool with 250 mm radius in 0° orientation. Detachments from the rim and compression buckling on the side of the compaction roller could be replicated in the simulation. It was found that the simulation and material model for the SF45 silicone can predict the pressure distribution for all sheath thicknesses with a maximum deviation in compaction pressure of approx. 20% on the flat surface. Radii of 250 mm can be simulated within the same accuracy but show greater deviations for smaller sheath thicknesses. The simulation and material model for the SS960 showed deviations of up to 30% in compaction pressure for all compaction rollers. However, the 5 mm sheath showed higher deviations of up to 60% for the concave radii with pressure distributions that did not match the experiments. This is partly caused by an unexpected deformation behaviour of the 5 mm sheath in the experiments (see Sect. 3). The general shape of the pressure distribution was replicated for the 25 mm and 10 mm sheaths of both silicones and all tool geometries. The use of only one material dataset for each silicone for the three different sheath thicknesses led to deviations to the experimental results. Deviations in vertical compression for the SF45 (SS960) silicones were in the range of 22% (31%) for the 5 mm sheath, 9,7% (7,8%) for the 10 mm sheath and 1,2% (1,3%) for the 25 mm sheath on flat surfaces. The deviations in compaction pressure and deformation strain may have been caused by differences in the maximum strains the material models were set up for, different friction conditions as assumed and measurement errors in the experiments due to the presence of the pressure measuring foil and limited resolution of about 3 mm.

3   Results and Discussion

3.1   Experimental Results: Pressure Uniformity

The pressure uniformity of the compaction rollers on the different tool surfaces and orientations can be seen in Fig. 3. All materials show a similar trend with the overall most uniform pressure distribution for the compaction rollers with a 25 mm sheath and a decline in uniformity for thinner sheaths. The silicone choice, and in turn their different hardness, does not significantly influence the pressure uniformity.

Furthermore, the orientation of the compaction roller produces a nearly identical pressure uniformity value over the roller width. It must however be noted that the actual shape of the pressure distribution varies for the different tool orientations as the pressure uniformity only shows averaged values. Figure 4 shows the pressure uniformity of the 10 mm sheath in 0° orientation of all three materials. Here again, it is clearly visible that the difference between the materials is nearly negligible.
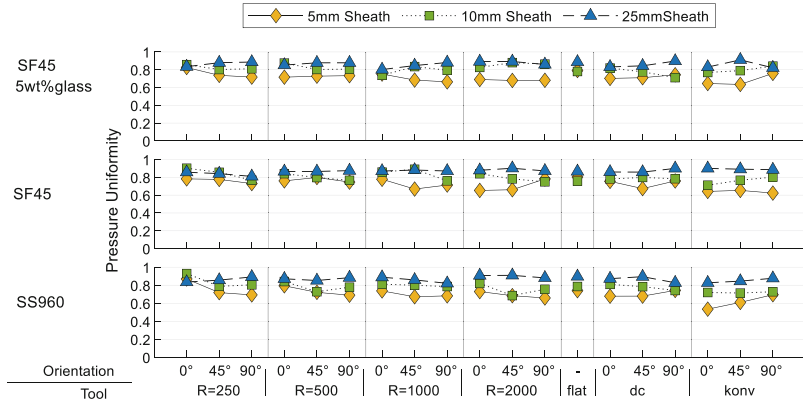
Fig. 3. Pressure uniformity on tool surfaces and compaction roller orientations in experiments
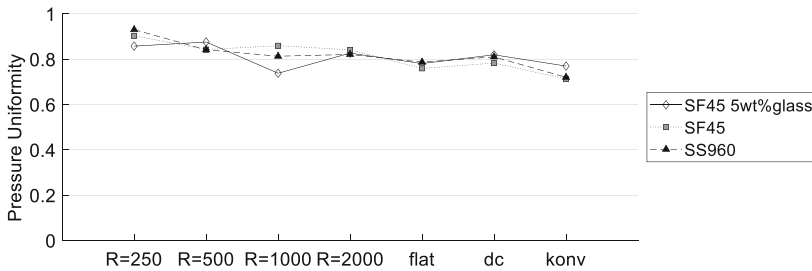
Fig. 4. Pressure uniformity, 10 mm sheath (0° orientation) in experiments

This direct comparison however highlights the trend of the compaction rollers to produce 10%–15% lower pressure uniformities on a flat surface than on a concave tool with 25 mm radius. This behaviour is exactly opposite of the expected results. An explanation can be found in a direct comparison of the overall pressure gradient over the roller width as seen in Fig. 5. The upper diagrams show the contact pressure over the whole roller width on a flat surface and the lower row on the concave tool with 250 mm radius in 0° orientation.

Especially the 5 mm and 10 mm sheath compaction rollers of each material show a strong decrease in pressure towards the edges of the roller on the flat surface. The 5 mm sheath shows up to 30% reduced contact pressure in the middle of the compaction roller, resulting in a "M"-shaped pressure distribution. The 25 mm sheath however shows an even distribution over the whole contact width with a small increase towards the edges. The reason for this behaviour could not be fully explained. It is however assumed, that the sheaths were able to shift on the rims since the position was only fixed by silicone filled groove in the rim. During the experiments, no sliding of the silicone on the tool surface or electronic pressure mat was observed. Under compaction, the silicone sheath will not only move towards the edges and create bulges, but also move towards the middle of the compaction roller. This may have resulted in buckling of the silicone between the rim and tool and thereby created deviations in contact pressure. The different sheath thicknesses
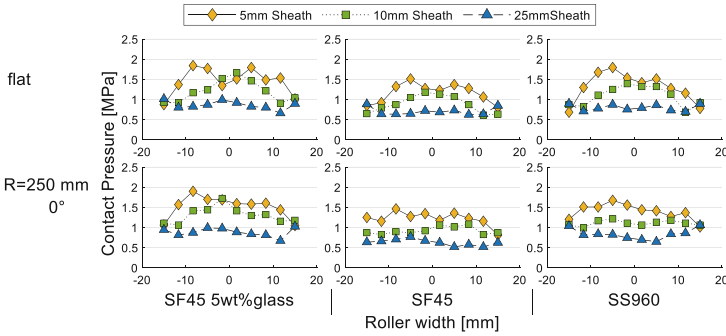
Fig. 5. Contact pressure in the middle of the contact length over width of compaction roller on flat and concave (R = 250 mm) tool, (0° Orientation) in experiments

will result in a different buckling behaviour. The lower row of diagrams in Fig. 5 with the pressure distributions on the concave 250 mm tool in 0° orientation shows an increase in pressure uniformity compared to the flat tool. Especially distinct are the effects on the 5 mm sheath. The pressure reduction in the middle of the compaction roller is no longer visible. The 10 mm sheaths show a similar distribution to the flat measurements. The SF45 material with 5wt% glass fibre shows similar pressure distributions to the SS960 silicone with higher local deviations. This can be explained by the additional short glass fibres resulting in local stiffness changes of the SF45 silicone.

3.2   Experimental Results: Compaction Pressure

The magnitude of contact pressure on the tape is a further important criterion as shown in the literature overview. Figure 6 presents the average compaction pressure as measured directly on a tape on the different tool surfaces and compaction roller orientations. All materials show a similar trend with the lowest compaction pressure on the tape for the 25 mm sheath and the highest for the 5 mm sheath. The compaction pressure of the 10 mm sheath is approximately 50–60% higher than the 25 mm sheath. The 5 mm sheath leads to a further increase in compaction pressure of 38–47%. This results in 2–2.5 times higher compaction pressures for the 5 mm sheath than the 25 mm sheath.

This is a result of a smaller overall contact area of the sheath with the tool surface. The silicone SS960 (Shore A60) achieves 20–30% higher compaction pressures than the SF45 (Shore A45) silicone in most experiments due to a shorter contact length. The SF45 silicone with 5wt% glass fibres reaches similar compaction pressure to the SS960 silicone. The average compaction pressure is independent from the orientation of the compaction roller to the tool surface. Figure 7 shows a direct comparison of the results for the 10 mm sheath in 0° orientation of the three materials. A trend towards higher compaction pressure for less curved tool geometries can be observed. This behaviour was expected as the required deformation for contact between the middle of the compaction roller and tape is higher on curved surfaces. The contact pressure on the tool with radius 250 mm is about 10%–20% lower than on flat surfaces. The measurement for the SF45 with 5wt% glass on the tool with radius 250 mm represents an outlier which could not be explained and is most likely caused by a measurement error.
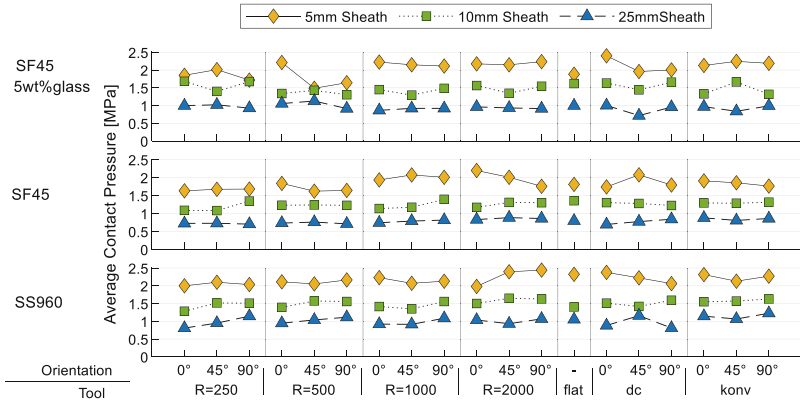
Fig. 6. Average contact pressure on tool surface and compaction roller orientation in experiments
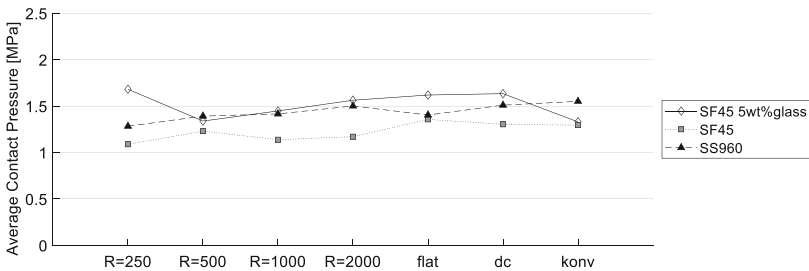

Fig. 7. Average Contact Pressure [MPa], 10 mm Sheath (0° Orientation) in experiments

This means that a higher compaction pressure produces a lower pressure uniformity (compare Fig. 4). Furthermore, higher compaction pressure at the expense of smaller contact areas reduces compaction time and in turn the available time for intimate contact development [19].

3.3 Simulation

Simulation of Perforated Compaction Roller Concepts
Different compaction roller designs were investigated with ANSYS to enable the in-situ laser assisted thermoplastic fibre placement on curved surfaces by improving pressure uniformity. A short overview on some concepts can be seen in Fig. 8. All designs are based on a 30 mm wide compaction roller with a 25 mm silicone sheath. A description of the simulation set up can be found in Sect. 2.3.

Simulation results showed that any change in the compaction roller geometry, which deviates from a full material sheath, will result in a change of pressure uniformity over the length of the tape. Areas directly under a perforation exhibit up to 3 times less compaction pressure than areas without perforation. These deviations would lead to varying consolidation pressure in a continuous process over the length of the tape.
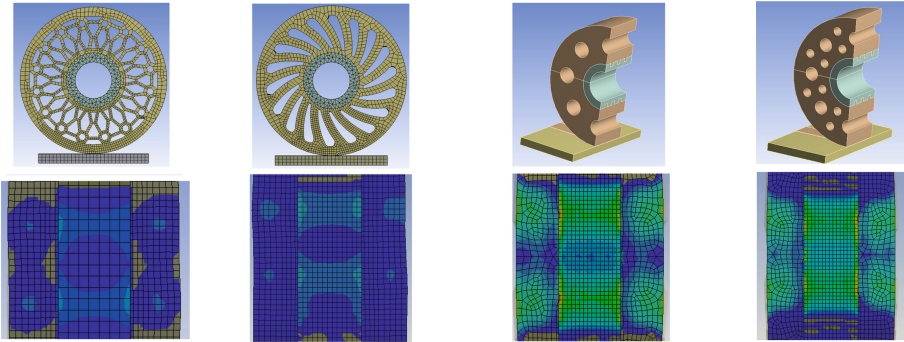
Fig. 8. Perforated compaction rollers and top-down view of pressure distribution on tool and tape

It was found that small perforations in the range of 2 mm can be added to the roller without significant impact on the pressure distribution over the length of the tape but also no improvement of the overall pressure uniformity over the width of the tape. These results are in accordance with the findings from Bakhshi et al. [12] who reported similar pressure distribution and differences of up to 50% in compaction pressure. It was found that silicones with a Shore hardness A45 and A60 cannot withstand the applied force of 547 N for larger perforations (Fig. 8 left and second left (adapted from [20])) and tend to collapse.

Simulation of Increase in Compaction Roller Width to 60 mm

All experimentally investigated compaction rollers have a width of 30 mm and only allow for the use of a single half-inch-wide tape. To increase productivity, multiple tapes need to be laid down at once. The pressure distribution of a 60 mm wide compaction roller on a flat and concave tool with a radius of 250 mm in 0° orientation was investigated using ANSYS. The simulations show that on a flat surface both silicones produce similar shaped pressure distributions (see Fig. 9). The SS960 silicone achieves 10%–30% higher compaction pressures than the SF45 silicone with greater differences towards thicker sheaths. The pressure distribution on a concave tool with 250 mm radius in 0° orientation is shown in the lower half of Fig. 9.
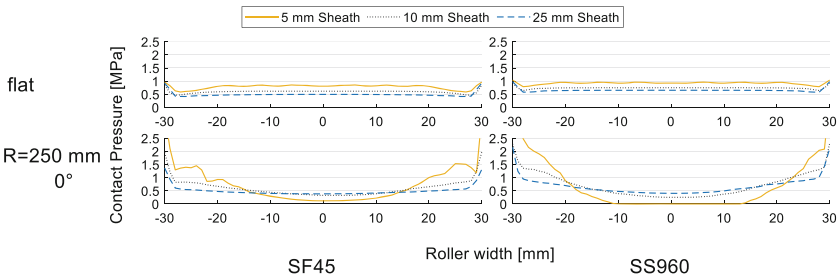
Fig. 9. Contact pressure distribution, simulation results for sheaths with 60 mm width

The 5 mm sheath of the SS960 silicone is not able to conform to the surface which results in a loss of contact in the middle of the compaction roller. It is however possible that the 5 mm sheath performs better in real applications as deviations between simulation and experiment are especially high for thin sheath thicknesses. The 10 mm and 25 mm sheaths show sufficient adaptability with similar pressure characteristics.

4 Conclusion

This work aimed to find new concepts for compaction roller designs to enable the use of the laser assisted thermoplastic in-situ tape placement process on curved surfaces.

Experimental and simulative methods were used to explore new concepts of compaction rollers to improve adaptability to curved surfaces, pressure uniformity and compaction pressure. Experiments revealed that 5 mm and 10 mm silicone sheaths exhibit a 10%–15% higher pressure uniformity on curved, concave surfaces with radius 250 mm than on a flat surface with a continuing trend towards reduced pressure uniformity on convex surfaces. This phenomenon is likely caused by internal deformation mechanisms that lead to buckling of the sheath between the rim and the tool surface. This buckling reduces pressure uniformity on flat and convex surfaces but increases pressure uniformity in superposition with concave radii as it counteracts the change in surface geometry. This effect is not noticeable with the 25 mm sheath. Experiments showed that the pressure uniformity over the tape width increases with increasing sheath thickness, with similar results for all materials. The compaction roller orientation on the tool does not significantly influence the pressure uniformity value. It is however important to note that the individual pressure distribution is different.

The contact pressure on a curved tool with radius 250 mm in 0° orientation is about 10%–20% lower than on a flat surface. In general, a clear correlation between increasing contact pressure and decreasing sheath thickness by a factor of 2 from a 5 mm sheath to a 25 mm sheath was found. A Shore A60 (SS960) silicone produces 20%–30% higher contact pressures than a Shore A45 (SF45) silicone, regardless of the sheath thickness. SF45 silicone with 5 wt% short glass fibres exhibits similar material behaviour to the SS960 silicone.

The experimental results showed that the best compromise for the counteracting mechanisms of higher compaction force at the expense of lower pressure uniformity can be achieved with a solid 10 mm Shore A60 (SS960) silicone sheath. Further experiments to validate the general work principle of this compaction roller configuration are currently ongoing with the objective of manufacturing a curved demonstrator structure.

Simulations showed that through thickness perforations in a 25 mm silicone sheath can lead to localised reduction of compaction pressure by a factor of 3 over the contact length on flat surfaces. Compaction roller designs which deviate from a pure solid material sheath are therefore not recommended. The simulations showed further that an increase in compaction roller width from 30 mm to 60 mm will likely result in a loss of contact for a 5 mm sheath on curved surfaces with radii smaller or equal than 250 mm. In future work, the simulation model needs to be adjusted to better reflect the behaviour of thin sheaths and to include temperature dependent silicone and tape material data.

A threshold value for the pressure uniformity could not be set in this work. Further investigations are necessary to understand the influence of compaction pressure on

mechanical properties and to analyse if findings on flat surfaces can be translated to curved surfaces. Here, pressure uniformities on curved surfaces could be replicated in 2D (for mechanical tests) with a compaction roller with varying elastic properties over its width.

References

1. Ahmad, H., Markina, A.A., Porotnikov, M.V., Ahmad, F.: A review of carbon fibre materials in automotive industry. In: IOP Conference Series: Materials Science and Engineering, vol. 971, no. 3, p. 32011 (2020). https://doi.org/10.1088/1757-899X/971/3/032011

2. Henning, F., Moeller, E.: Handbuch Leichtbau: Methoden, Werkstoffe, Fertigung. Hanser, München (2011)

3. Marsh, G.: Composites in commercial jets (2015). https://doi.org/10.1016/j.repl.2015.06.001

4. Schledjewski, R.: Thermoplastic tape placement process - in situ consolidation is reachable. Plast. Rubber Compos. **38**, 379–386 (2009). https://doi.org/10.1179/146580109X12540995045804

5. Li, J.F., Song, C., Wang, X.F., Xiao, J.: Study on pressure control of automated fibre placement process. In: 19th International Conference on Composite Materials, Montreal, Canada. http://www.iccm-central.org/Proceedings/ICCM19proceedings/papers/LIJ80033.pdf

6. Khan, M.A., Mitschang, P., Schledjewski, R.: Identification of some optimal parameters to achieve higher laminate quality through tape placement process. Adv. Polym. Technol. **29**(2), 98–111 (2010). https://doi.org/10.1002/adv.20177

7. Raps, L., Chadwick, A.R., Schiel, I., Schmidt, I.: CF/LM-PAEK: characterisation and sensitivity to critical process parameters for automated fibre placement. Compos. Struct. **284**, 115087 (2022). https://doi.org/10.1016/j.compstruct.2021.115087

8. Chu, Q., Li, Y., Xiao, J., Huan, D., Zhang, X.: Placeability restricted by in-complete contact between laying roller and mould in an automated fibre placement process. J. Reinf. Plast. Compos. **37**(7), 475–489 (2018). https://doi.org/10.1177/0731684417752871

9. Zhao, P., Shirinzadeh, B., Shi, Y., Cheuk, S., Clark, L.: Improved uniform degree of multi-layer interlaminar bonding strength for composite laminate. J. Reinf. Plast. Compos. **36**(17), 1211–1224 (2017). https://doi.org/10.1177/0731684417704075

10. Hélénon, F., Ivanov, D., Potter, K.: Modelling slit tape deposition during automated fibre placement. In: 19th International Conference on Composite Materials, Montreal, Canada (2013). http://www.iccm-central.org/Proceedings/ICCM19proceedings/ICCM19-eproceedings.pdf

11. Jiang, J., He, Y., Ke, Y.: Pressure distribution for automated fibre placement and design optimization of compaction rollers. J. Reinf. Plast. Compos. **38**(18), 860–870 (2019). https://doi.org/10.1177/0731684419850896

12. Bakhshi, N., Hojjati, M.: Effect of compaction roller on layup quality and defects formation in automated fibre placement. J. Reinf. Plast. Compos. **39**(1–2), 3–20 (2020). https://doi.org/10.1177/0731684419868845

13. Lichtinger, R., Lacalle, J., Hinterhölzl, R., Beier, U., Drechsler, K.: Simulation and experimental validation of gaps and bridging in the automated fibre placement process. Sci. Eng. Compos. Mater. **22**(2), 131–148 (2015). https://doi.org/10.1515/secm-2013-0158

14. He, Y., Jiang, J., Qu, W., Ke, Y.: Compaction pressure distribution and pressure uniformity of segmented rollers for automated fibre placement. J. Reinf. Plast. Compos. **41**(11–12), 427–443 (2022). https://doi.org/10.1177/07316844211054166

15. Hertz, H.: Ueber die Berührung fester elastischer Körper. pp. 156–171, (1881)

16. Popov, V.L.: Contact Mechanics and Friction. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-10803-7

17. Atanackovic, T.M., Guran, A.: Theory of Elasticity for Scientists and Engineers. Springer, Boston (2000). https://doi.org/10.1007/978-1-4612-1330-7
18. Roland, C.M.: The mullins effect in crosslinked rubber. J. Rheol. **33**(4), 659–670 (1989). https://doi.org/10.1122/1.550032
19. Stokes-Griffin, C.M., Compston, P.: Investigation of sub-melt temperature bonding of carbon-fibre/PEEK in an automated laser tape placement process. Compos. Part A Appl. Sci. Manuf. **84**, 17–25 (2016). https://doi.org/10.1016/j.compositesa.2015.12.019
20. Xiaohui, Z., Wenfeng, Wu., Yugang, D., Ti, Y., Chen, Li., Chenping, Z.: A kind of large deformation flexible pressing device for automatic fiber placement, Patent No. CN110356019A (2019)

Development, Implementation and Evaluation of a Prototype System for Data-Driven Optimization of a Preforming Process

Michael Liebl[1], Jonas Holder[2], Tobias Mohr[2], Albert Dorneich[2], Florian Liebgott[2(✉)], and Peter Middendorf[1]

[1] Institute of Aircraft Design, University of Stuttgart, 70569 Stuttgart, Germany
[2] Balluff GmbH, 73765 Neuhausen a.d.F, Germany
`florian.liebgott@balluff.de`

Abstract. Modern production of fiber reinforced composites via the preforming process is widely used in the industry. A common way to create dry, semi-finished fiber products is forming or draping a textile into a three-dimensional component geometry. The punch and die process is often used for resin transfer molding (RTM) composite manufacturing. Due to the major influence of the preforming step on the later mechanical performance of the component, a detailed knowledge of the fiber architecture is beneficial.

To enable in-situ monitoring of the specific deformation of a woven fabric, a novel kind of single-use two-dimensional strain sensors has already been developed and characterized. We show that by using industrial communication standards, data from various data sources can be consolidated in an edge computer and used to improve the process. To this end, we developed the hardware and firmware of a device that reads out the printed strain sensors and transfers the data to the edge device via IO-Link. In addition, the edge device collects data from a programmable logic controller and is capable of connecting further IO-Link sensors.

Our demonstrator is intended as a proof of concept for in-situ monitoring, data-driven analysis and improvement of the punch and die process and will be further developed. We propose a machine learning-based edge analytics approach for detecting defects and increasing the preforming quality during the draping process. Forming tests with the double-dome benchmark geometry and the carbon fabric which is suitable for industry have been carried out to validate our prototype system.

Keywords: Composite · Draping · Strain sensing · Edge computing · IO-Link

1   Introduction

In-line quality assessment is a standard for modern automotive series production processes. In the manufacturing of high-performance composite structures, a crucial process step is the preforming of dry semi-finished fiber products. For this draping process, to

date there are no established sensor systems for quality assurance. However, these are necessary for the production of components in order to guarantee the desired quality in an economical and resilient manufacturing process. The presented method is independent of the geometry of the component. We thus seek to present a universally applicable method for the optimization of the draping process.

1.1 Forming of Woven Fabrics

We present an approach for an enhanced textile forming process that uses previously introduced in-situ shear monitoring sensors to analyze the local fiber architecture [1]. Using highly anisotropic reinforcement materials such as carbon fibers, the quality of a component depends to a large extent on the local variations of the fiber orientation. During the forming process, the fabric is adapted to a three-dimensional geometry by relative rotational or lateral yarn movements. Thus, this adaption process has a major impact on the mechanical performance of the whole structure [2, 3]. To ensure good further processing of the dry preform in a subsequent infusion or resin transfer molding (RTM) step, wrinkling during forming has to be avoided by reducing the local shear angle. If this textile can be prevented from reaching its specific locking angle, wrinkling is unlikely to happen [3].

To document the current fiber architecture of a woven fabric optical methods are common and widely used, but these are not applicable in the closed mold process of stamp forming which is investigated here [4, 5]. The application of flat in-situ sensors on the textile and their use in the draping process is therefore beneficial.

In general, the correlation between the shear angle of a woven fabric and the strain sensors used has been analyzed in [1] with the picture frame test setup, where the sensors are orientated according to the crosshead movement of the frame. With given information about the strain ε of the sensor, the corresponding shear angle θ of the fabric can be calculated by

$$\theta = \frac{\pi}{2} - 2 \cdot \cos^{-1}\left(\frac{1 + \varepsilon}{\sqrt{2}}\right). \tag{1}$$

1.2 Sensors for Composite Structures and Manufacturing

The increasing need for lightweight fiber composite structures in the automotive, aerospace and energy industries gives rise to a demand for low-cost, high-quality and efficient manufacturing in this field. Automation and quality monitoring will help to reduce waste and resource consumption (e.g. energy, water and materials).

In addition, there is a clear trend toward sensor integration for structural health monitoring during the whole life cycle. The integration of sensors in carbon fiber-based tanks for high-pressure renewable hydrogen storage allows safe weight-optimized tanks due to the permanent monitoring (pressure, temperature, aging).

Work is reported on dielectric sensors, thermal flux sensors, fiber optic sensors and contactless sensors that use ultrasound and electromagnetic waves to provide in-situ production data [6].

A network of 74 sensors, including 57 ultrasonic sensors, is used in the T-RTM molding of a thermoplastic composite battery box cover demonstrator for the CosiMo project. Dielectric sensors detect the flow, cure and glass transition temperature by measuring the resin electrical resistivity. Electromagnetic methods are used to monitor flow, cure, viscosity, polymerization and other properties of liquid resin. Temperature strain is measured by fiber Bragg gratings integrated in the composite structure during the production process. The requirements for measuring quantity, range, accuracy and rate and for energy consumption as well as the type of interface, such as standardized, proprietary or wireless, depend to a large extent on the application.

To our knowledge, no sensors have been used for monitoring the preforming process to date. In our work, we demonstrate the development of a printed strain sensor prototype for industrial applications using IO-Link communication and an evaluation of the sensing signals using machine learning methods.

1.3 Strain Sensing

The resistance R of a conductor depends on the conductivity ρ, the length l and the cross-section A:

$$R = \rho \frac{l}{A}. \tag{2}$$

The elongation $\Delta l / l_0$ applied to the conductor therefore leads to a change in the resistance R which can be used to measure the strain. In commercial strain gauges the relative change in R is very small because of a typical elongation in the order of 10^{-3}. A Wheatstone bridge circuit is often used to increase the sensitivity [7].

For our application, a novel strain sensing element as well as an evaluation method had to be developed, which will be described in Sect. 2.

1.4 Edge Computing

In industrial automation, processes are usually controlled by programmable logic controllers (PLCs). A PLC controls its outputs based on its inputs using a custom program tailored to the process to be controlled. It is, however, not suitable for collecting large quantities of data and analyzing it using complex algorithms, for example from the field of machine learning. Therefore, these tasks are often executed in cloud environments with large amounts of computing power. However, transferring production data into a cloud environment, where the data and the data security are not under the direct control of the operator of the manufacturing process, is not always desired. Additionally, the availability of the data storage and analysis depends on the availability of both the internet connection of the manufacturing plant and the cloud operator's infrastructure.

It is therefore beneficial to implement the data collection and analysis on-site and close to the process. Using an edge computer, which can collect the data directly from the sensors and PLC used in a manufacturing process, eliminates availability issues and ensures low latencies. A drawback of edge computing, however, is the limited computing power. For complex tasks, it is therefore crucial to find solutions that are suitable for the computing resources available.

2   Experimentation

In this work, we measure local fabric deformation using a bindered carbon fiber weave and piezoresistive strain sensors. The following sections will illustrate the materials used and the setup of our prototype system for the draping process.

2.1   Basic Materials

To create the layout of the piezoresistive strain sensors, a screen-printing process is used with SunTronic PTF Silver (AST6400) conductive ink from SunChemical®. This ready-to-use ink is well suited for applications with high elongation, such as strain sensors. The sensors are printed on a thermoplastic polyurethane substrate, which is highly flexible as well. The strain sensors are applied to a Hexcel® HexForce® G0926 carbon fabric, which is a standard textile in the aerospace industry. It is bindered with 2.5% of an epoxy powder binder per side and, in the draping experiments, one layer of fabric with an areal weight of 375 g/m^2 is used.

2.2   Sensor Characterization

We are using the printed sensor layout shown in Fig. 1 to establish the designated strain measurement during the draping process. The basic sensor element is represented by a printed area of 20 x 4 mm, which is connected by silver ink lines with a width of 1 mm, which are also printed and therefore conductive.
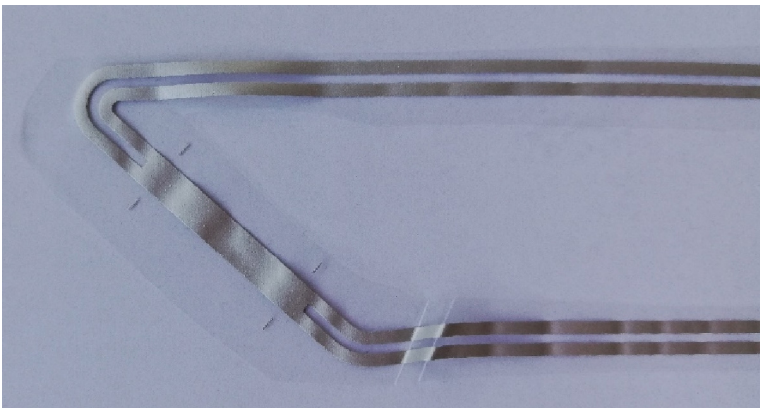
Fig. 1. Screen-printed silver ink strain sensor with marks for ultrasonic welding

Tensile tests are carried out to characterize the resistance over the physical deformation. A 3D-printed test rig is designed and produced for this purpose. With the test rig, tensile tests are carried out between 0 °C and 80 °C in 20 °C increments in the CTS T-40/25 temperature test chamber. The resistance is measured by imprinting 10 mA of current through the feed lines and measuring the voltage across the two sensing lines

using a Fluke 175 multimeter. This allows to neglect the resistances in the feed and sensing lines and their variations due to temperature or strain.

As shown in Fig. 2, the change in resistance at 0 °C and at 20 °C is significantly higher compared to the higher temperatures up to 80 °C. The change in resistance in the worst-case scenario at 80 °C is an increase from 0.5 Ω at zero elongation, over 2.4 Ω at 50% elongation, up to 8 Ω at 100% elongation.
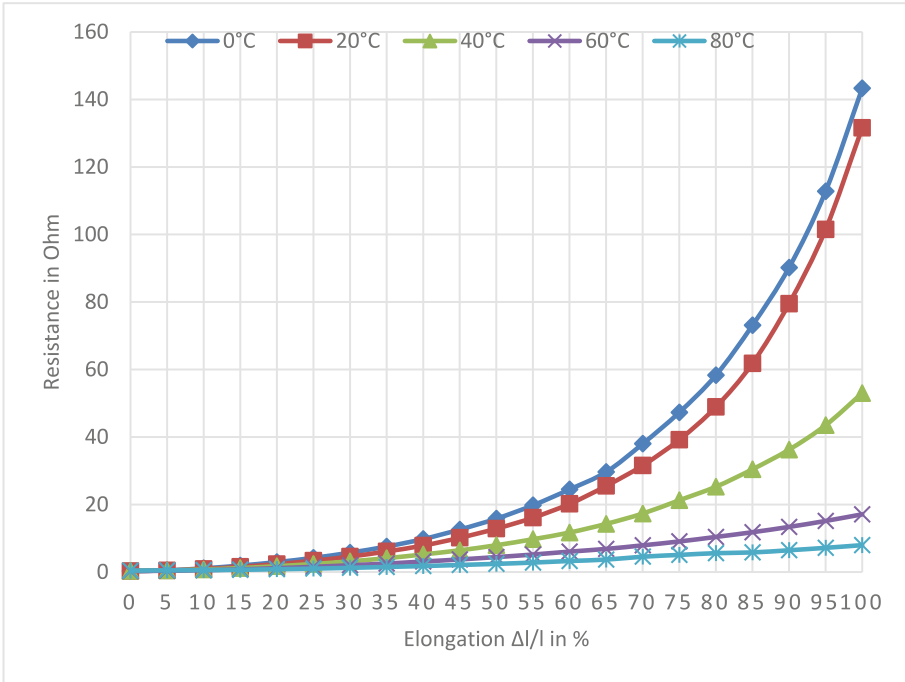
Fig. 2. Resistance over elongation at different temperatures

In [1], several calibration experiments have been carried out, where the printed strain sensors are applied to the Hexcel® carbon fabric in a picture frame test. In this test, the fabric shear is accurately defined and the strain is documented by the sensors. According to the optical evaluation of the shear angles, the strain sensors are well suited for reproducibly measuring the deformation of the fabric.

2.3 Strain Sensor System

A four-wire measurement setup is used to detect the change in resistance. Two wires are used to imprint a constant current into the system. The other two wires have a high input resistance to measure the voltage drop over the sensing element, without being jammed by a change in resistance in the feed line. Multiplexers switch the current and the sensing connections internally to measure up to eight strain sensors in a time-multiplexed manner. The imprinted current is in the range of 1–10 milliamps to keep

the voltage over the measuring system below 3.3 V. This low current only generates a small voltage drop across the sensing element. To amplify the voltage over the sensing element, an instrumentation amplifier is used. Its amplification factor can be modified by the resistors that are connected to it. In this work, four different resistors can be connected to the amplifier via dual in-line switches resulting in amplification factors of approximately 1, 10, 100 and 1000.

The constant current source is implemented using an LT3092IST. It is capable of switching between different loads within microseconds, enabling switching frequencies in the range of several kilohertz. Two resistors in parallel set the current to 1 mA. The current is switched between the channels using an SN74LV4051 eight-channel multiplexer. All sensors are connected to a common ground. The other two multiplexers are used to connect the measuring lines to an instrumentation amplifier's input. All multiplexers' address lines are connected in parallel, resulting in synchronized channels. The INA849 is an ultra-low-noise, high-bandwidth instrumentation amplifier, whose amplification can be set using resistors. The output of the amplifier is connected to the analog-to-digital converter (ADC) of the IO-Link board, which also controls the address lines of the multiplexers. Additional DC-DC converters are used to supply all parts with their correct supply voltage. The DC-DC converters are supplied with 24 V from the IO-Link master. The schematic of the constant current source and the multiplexers is shown in Fig. 3.
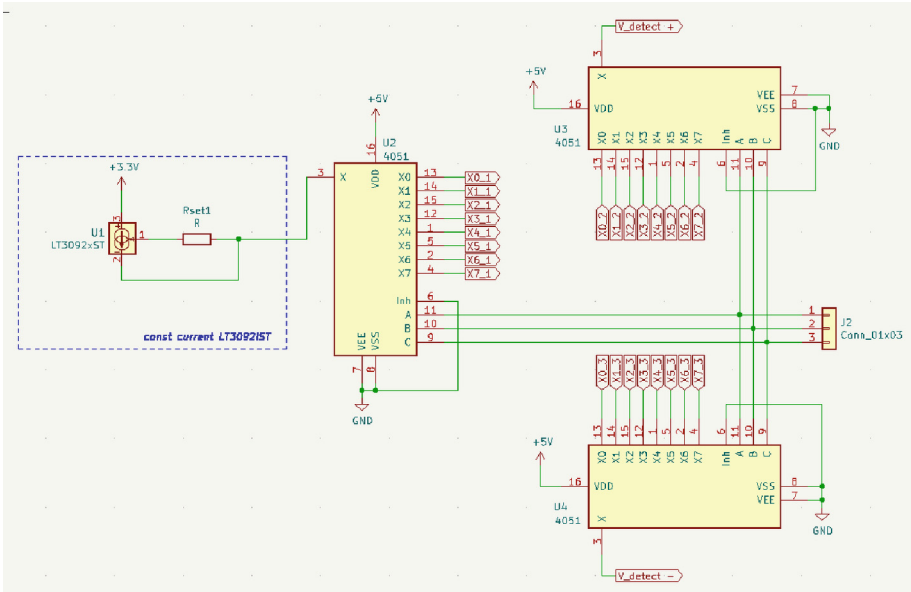
Fig. 3. Schematic including multiplexers and power supply

The IO-Link board runs custom code that switches the multiplexers' channels cyclically and reads the ADC's value in-between. The switching speed depends on the ADC's

averaging. With an averaging of 32 consecutive values, the addresses change every 0.66 ms whilst the ADC's values are saved in a ring buffer.

IO-Link is a master driven point-to-point communication protocol. Whenever the IO-Link master requests data, the current data in the ring buffer is transmitted via IO-Link process data.

To allow the sensor box to be recognized by any IO-Link master, we created a custom IO device description (IODD). The IODD specifies how to convert the raw ADC values to human-readable resistance values. In future implementations, it will also be possible to convert the raw ADC values directly into elongation values.

2.4   Integration into Edge Device

To evaluate the data, the sensor values of the PLC and IO-Link boards are collected. The Balluff Condition Monitoring Toolkit (CMTK), which is an edge device containing an IO-Link master, acts as the central device for linking all this measurement data.
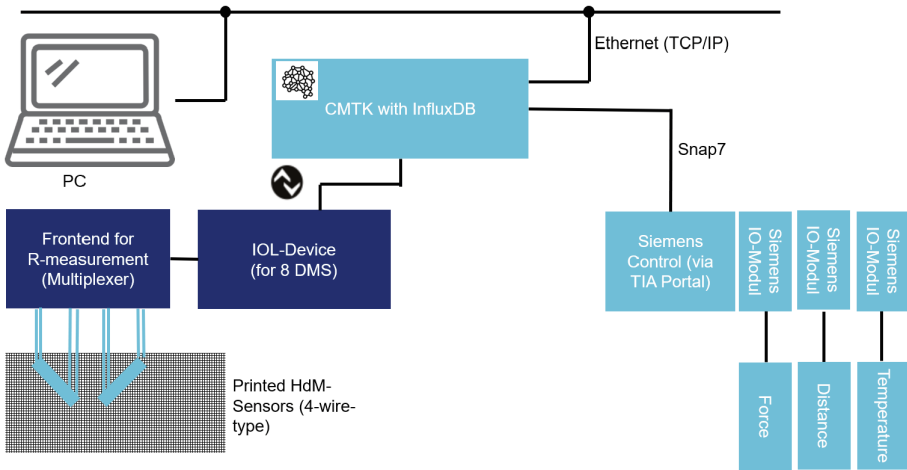
Fig. 4.   System architecture with interfaces

Figure 4 shows the complete system architecture with the different communication protocols. These are described below.

The strain sensor system is integrated via the IO-Link ports to the CMTK. This is done automatically via a custom IODD file, which contains the format of the data. The data is processed and sent to the broker in integer values via MQTT. Afterwards, the data is written to an influx database.

A SNAP7 python wrapper interface was used to implement the communication between the CMTK and the Siemens controller. This is an Ethernet communication protocol for native coupling with Siemens S7 PLCs. For this purpose, the data of the various IO modules is stored in a database on the Siemens controller. The data is read out via SNAP7 running inside a Docker container on the CMTK and stored in the influx database.

The data collected from the IO-Link ports and via SNAP7 can then be analyzed, for example with machine learning algorithms running directly on the edge device. Additionally, all of the data stored in the database is made available through a REST API endpoint via a CSV export. Via this CSV export, the data can be further processed on another device.

We implemented a trigger in the control system to prevent the unnecessary storage of data while the prototype system is idle. Only if this is active, will the data from the controller and the MQTT broker be stored in the influx database. This allows for more efficient use of the available memory and more cycles to be stored in the database.

2.5 Draping Process

The printed sensors are applied on the textile material by ultrasonic welding, due to its fast processing and minimal influence on the behavior of the textile during the forming process. The sensors are welded directly on the textile using a VE35 Compactline Solid STE 1200 (from Hermann Ultrasonic) without any additional adhesive and an energy of 20 J per spot weld. The main element of the sensor, which has a length of 20 mm, as well as the ink-printed conduction lines stay unfixed to avoid interference with the measuring signal.

The textile is placed inside the preforming station, which is equipped with an isolated metal double-dome geometry mold for stamp forming (see Fig. 5). The sensors are connected to the sensor box and data is recorded during the entire forming process. To ensure a good draping quality, blank holder plates are placed around the circumference of the textile with a down force of 10 N.
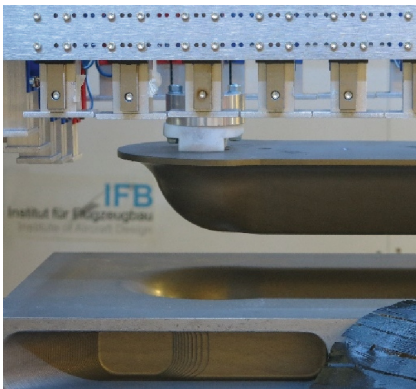
Fig. 5. Stamp mold preforming station
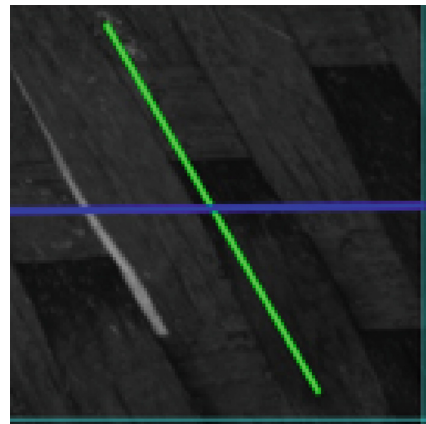
Fig. 6. Analysis of fiber orientation

An optical method for quality assessment similar to [8] is used to validate the shear information based on the strain of the printed sensors. The area where the sensors were placed on the textile is evaluated using grayscale image analysis to obtain information about the local shear angle (see Fig. 6).

3   Results

We show the evaluation regarding elongation values of the strain sensors as a function of the stamp movement during the textile forming process, as shown by way of example in Fig. 7. In total, four printed strain sensors are used on two woven fabrics in an area with a high estimated shear angle. The strain sensors show no elongation during the first half of the forming process. After the stamp moves further into the textile, an almost linear increase in strain is documented. At the final position of the stamp, the maximum elongation of the strain sensors is shown as a numerical value.
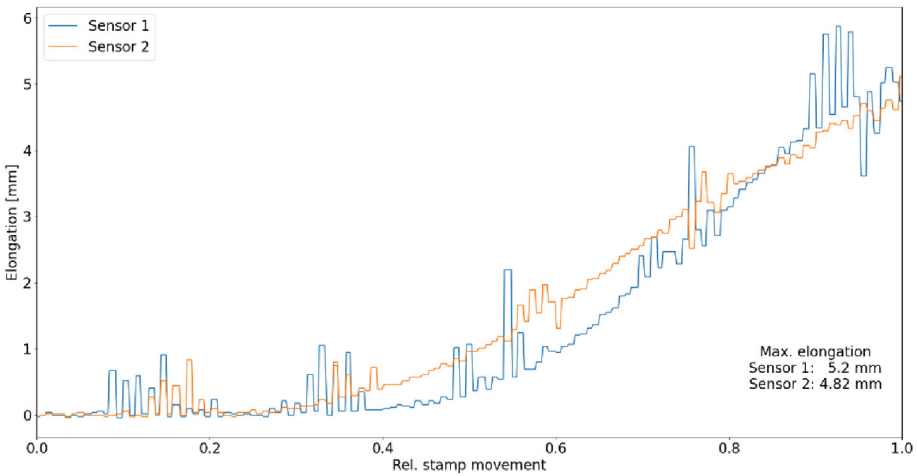
Fig. 7. Sensor elongation over stamp movement

Based on this value and Eq. (1), we estimate the local shear angle of the fabric in Table 1. Additionally, we also display the shear angle which is determined by the optical grayscale method as well as the deviation from the sensor-based data.

In general, good alignment of the shear angle evaluation with both methods can be shown. The small deviation of up to 11% can occur due to the manual positioning of the printed sensors for ultrasonic welding and the tolerances of the optical grayscale analysis. Furthermore, the printed strain sensors show good reproducibility as expected from [1].

Table 1. Shear angle evaluation

ID	Sensor strain ε [%]	Shear angle by sensor θ_s [°]	Shear angle, optical θ_o [°]	Deviation [%]
1	21,4	28,3	30,1	-6
2	20,1	26,3	29,6	-11
3	26,1	36,2	33,7	7
4	24,2	32,9	30,1	9

4 Conclusion

Our work shows that it is feasible to improve the quality of a draping process by con-solidating and using data from different sources in an edge device. We developed and implemented the hardware and firmware of an IO-Link-enabled sensor box to read out printed strain sensors. Furthermore, we implemented the communication via IO-Link, MQTT and SNAP7 on a commercially available edge device and successfully integrated the edge device in the draping station. As a first proof of concept, we used the strain sensor data to estimate the shear angles of the woven fabric as an indicator of wrinkling. Forming tests with the double-dome benchmark geometry showed promising results that can be used to improve the process quality.

In the next step, we want to improve the monitoring of the draping process. Instead of relying on the shear angle estimation as an indicator of possible wrinkling, we want to estimate the occurrence of wrinkling and other defects directly from the sensor data using machine learning. Defect detection will be implemented directly on the edge device, enabling the process quality to be improved without additional hardware or cloud computing.

References

1. Fial, J., et al.: Shear characterization of reinforcement fabrics using textile-applied printed strain sensors. Proc. Manuf. **47**, 65–70 (2020)
2. Boisse, P. (ed.): Advances in Composites Manufacturing and Process Design. Woodhead Publishing, Cambridge (2015)
3. Long, A.C. (ed.): Design and Manufacture of Textile Composites. Woodhead Publishing Limited, Cambridge (2005)

4. Harrison, P., Clifford, M.J., Long, A.C.: Shear characterisation of viscous woven textile composites: a comparison between picture frame and bias extension experiments. Compos. Sci. Technol. **64**, 1453–1465 (2004)
5. Christ, M.: Definition und Quantifizierung der Drapierbarkeit von mulitaxialen Gelegen durch die Vermessung von Einzeleffekten. Dissertation. Faserinstitut Bremen e.V. (2018)
6. Composites World Website. https://www.compositesworld.com/articles/sensors-data-for-next-gen-composites-manufacturing. Accessed 30 May 2022
7. Hering, E., Schönfelder, G. (eds.): Sensoren in Wissenschaft und Technik. Vieweg+Teubner Verlag, Wiesbaden (2012). https://doi.org/10.1007/978-3-8348-8635-4
8. Fuhr, J.-P., Böhler, P., Heieck, F., Miene, A., Middendorf, P.: Optische Preformanalyse zur 3D-Validierung der Drapier- und Flechtsimulation, Augsburg 2014

Resource Efficient Manufacturing of Complex Cooling Structures

Leonard John[1(✉)] , Dina Becker[2], Steffen Reuter[1] , Hans-Christian Möhring[2] ,
Martin Doppelbauer[3] , and Lars F. Berg[1]

[1] Fraunhofer Institute for Chemical Technology,, Pfinztal, Germany
`Leonard.John@ict.fraunhofer.de`
[2] Institute for Machine Tools, University of Stuttgart, Stuttgart, Germany
[3] Karlsruhe Institute of Technology, Karlsruhe, Germany

Abstract. The combination of additive manufacturing (LPBF) and plastic injection molding offers great benefits for functional integration and thereby the functionality of parts, components and systems. By manufacturing cooling channels as very thin-walled structures, short laser processing time is needed to produce these parts. The combination with plastic overmolding allows the mechanical loads to be absorbed, while at the same time utilizing the freedom of shape of the LPBF process. In order to prevent the structures from collapsing during overmolding or to ensure that they do not have to be very thick-walled in order to withstand the high injection pressures, the structures are closed, so that the base-powder is still contained in the structure during the overmolding. After opening the overmolded components, the powder can be extracted and reused. This technique offers great potential for high functional integration and new possibilities in terms of realizing internal cooling structures of electric traction machines. Other advantages that can be exploited are the thermally good connection due to the material bond and the possibility of local reinforcement due to the design freedom of the insert. Thanks to the injection-molded component, complex geometries can also be produced without additional unit costs.

Keywords: additive manufacturing · cooling channels · plastic overmolding · electric motor · compliant mechanism insert

1 Motivation

In the construction of electric devices like motors, there is a great potential to reduce the use of expensive and rare materials like neodymium, cobalt or copper. To achieve this goal the continuous power density needs to be increased. Many mechatronic power converters are limited in this regards by the cooling capacity. Here we see the possibility to increase the performance with improved internal cooling channels in the device's encapsulation. Were standard methods for cavities like cores are not applicable anymore inserts with cooling channels, can offer more flexibility. Therefore, this paper will emphasize on the possibilities and challenges of additively manufactured metal inserts in injection over molded devices.

2   State of the Art

Some of the most advanced electric motors are designed with internal cooling channels [1]. So the thermal path can be greatly shortened. In addition, the devices can be embedded, potted or molded in high thermal conductive materials.

An example for an electric motor with internal cooling channels through the groove is the DEmiL project [2]. Here, the channels were created directly in the potting process. Core bars with a triangular cross-section formed the channels between the windings after demolding. In bench tests, a 2.76-fold higher power density was demonstrated in comparison with a similar, conventional jacket-cooled machine. However, the techniques used in this project cannot be used for devices whose structure requires 2 or 3 dimensional internal channels. Such shaped cores could not be demolded due to the resulting undercuts.

The use of additive manufacturing for an improved cooling is often seen in (experimental) high-power heat sinks and exchanger in power electronics [3][4]. These parts benefit from flow optimized and narrow sized channels enabled by the design freedom of additive manufacturing.

Mechanical metal inserts in injection molded components are well known. Threaded inserts are for example widely used to form and reinforce the area around fixation points. The mechanical optimization of the plastic/metal contact of additively manufactured parts was investigated by Saurav Verma [5].

3   Proposed Advantages Through the Enhancing of Injection Molded Encapsulations with Additive-Built Inserts

The benefits of mixing additive manufacturing with injection molding becomes relevant when the advantages of one specific production technique is used to cancel out the disadvantages of the other. Additionally, the insert can and should combine other functions, which are enabled by (AM-) implants as shown in Table 1.

Additive manufactured cooling channels are reinforced by the over molded plastic. So, manufacturing costs can be kept low, as wall thicknesses and thus laser time and material consumption can be minimized. This can further be exploited when the processing metal powder of the LPBF process is kept inside the channels during the over molding. So, the channels are stabilized against indentation by the injection pressure.

By using the injection molding process for the encapsulation, not only the inserts can be designed as small as possible, but complex external elements can also be integrated in the plastic [6]. The very good flow properties of the material allows filling of even the thinnest gaps, so that a very good thermal interconnection can be realized while mechanical vibrations are damped. Injection molding processes are widely established in the industry and enable low unit costs with increasing quantities, since the high mold costs can be amortized by the high part volume.

Figure 1 shows a conceptual bearing-seat-insert in a plastic housing as an example for a well-designed insert, that utilizes many of the design directives, implicated by Table 1.

Table 1. Suspended disadvantages of plastic injection molding (IM) and metal additive manufacturing (AM) by the combination of both.

Disadvantage IM / Advantage AM	Functionalities and examples
Material strength	Mechanically loaded surface,Threaded insert
Inner undercuts	Cooling channels, threads
Electrical and thermal conductivity	Electronics
Disadvantage AM / Advantage IM	
Overmolding capability	Tolerance compensation, bonding, filling
Costs and production time for big sized parts	Housing parts
Material density	Housing parts
Surface finish	Sealing surfaces
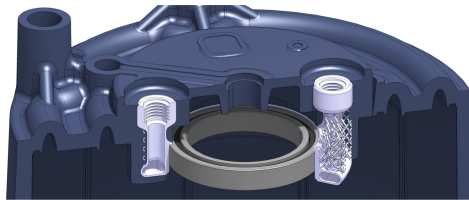
Fig. 1. Example for function combination in a cooled bearing seat insert

4 Motivation in the HEaK Project

In the research project "High-efficiency electric motor with additively manufactured cooling system in plastic overmold (HEaK)" [7], the power density of the widely used hairpin-wound electric traction motors is to be increased by a factor of 2 by inserting cooling channels inside the motor. Due to the twisted and welded and thus fanning winding head, no straight cores can be inserted in the mold. Therefore, additively manufactured channels are inserted, to allow the cooling medium to be directed around the undercut winding head (Fig. 2).
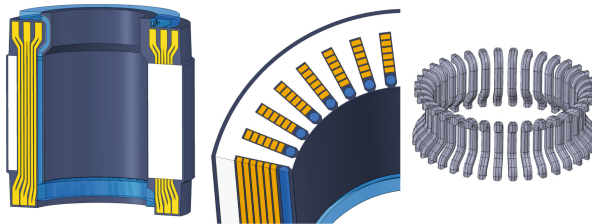
Fig. 2. Left and Middle: Hairpin winded HEaK Stator with internal cooling channels; **Right:** channel-inserts

The production process is as follows: The insert ring with all channels is manufactured from stainless steel as a tubular structure with closed ends using the LPBF (Laser Powder Bed Fusion) process. The surrounding powder can be directly reused according to known recycling processes, while a little powder initially remains in the tubes. After sawing from the building platform the ring can be inserted in the stator winding [8]. The stator is then overmolded with highly thermally conductive thermoset plastic. Thanks to the enclosed powder, the channels withstand the high injection pressures. Drilling and milling open the channel ends, and the powder can trickle out for further reuse (Fig. 3).
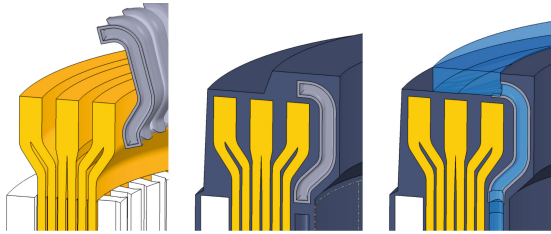
Fig. 3. Production steps (from left to right): Inserting the ring with all channels; overmolding with thermoset plastic; opening the ends of the channels

4.1   Design of the Insert Structure

Strut Design. In order to be able to mount the insert structure on the stator, it should allow a flexibility, which enables the channels to be guided past the winding head undercut. The design freedom through the additive manufacturing is further used to implement this assembly function. For this the all-channels-connecting struts, are designed as a compliant mechanism, which allows a bistable, inward orientated rotation. This is achieved by disc spring-like ring struts. For an improved manufacturing process, the spring is designed in a double-strut design in the later versions (Fig. 4).
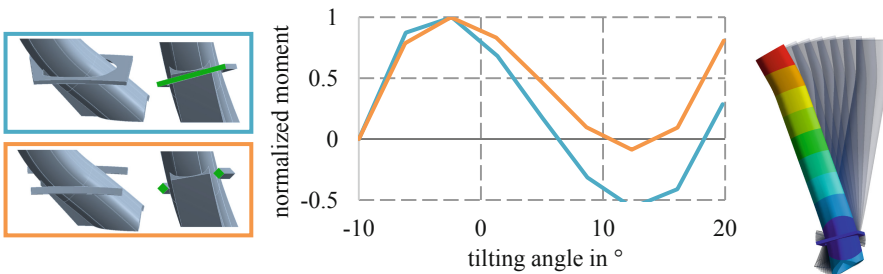
Fig. 4.   Left: Two disc spring based strut designs (cutout) Blue: disc spring with gap around the channel; orange: double-strut design. **Middle:** spring characteristic: **Right:** Simulated Movement of one channel in the ring (displacement to scale)

Design of the tubes. For the LPBF production-ready design of the tube structure, the aim is not to design any overhangs of more than 45° in order to avoid the need for support structures. In addition, a wall thickness of at least 0.35 mm is maintained. The cross-section of the tube element is implemented as a rounded square with a circular inner channel. This contour allows the minimum wall thickness in the direction of the winding head, so the channel to be run along it as narrowly as possible. At the same time, the thickened corners stiffen the tube for safer handling during installation. In the angled tube area, the inner contour changes into a "circle with roof". This ensures that the inner walls do not have too much overhang (Fig. 5).
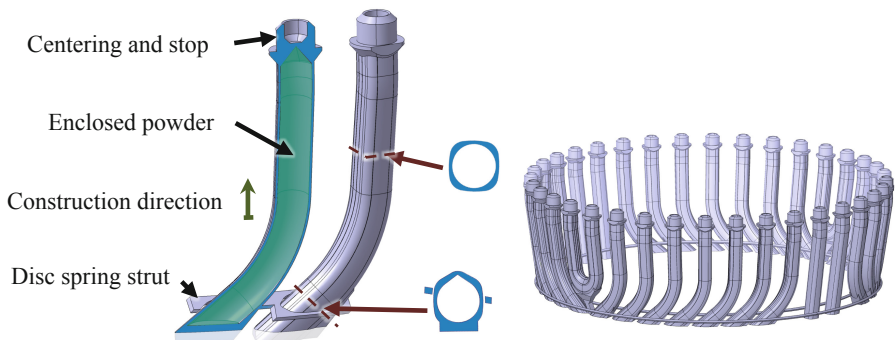
Fig. 5. Tube design of two cut out channels and final inserting ring
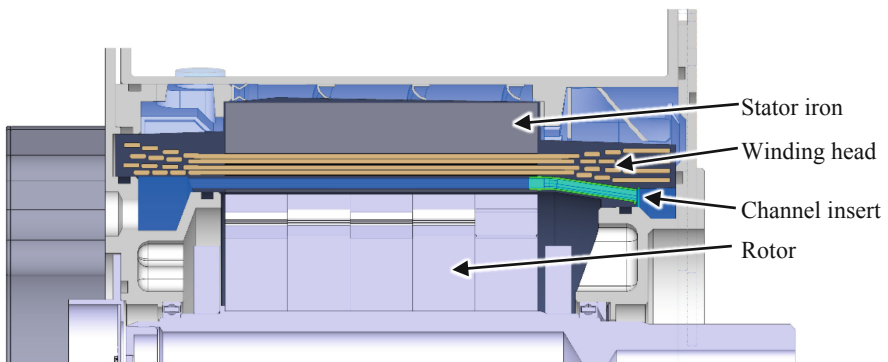
Fig. 6. Position (green) of the inserted channels in the HEaK e-motor design

5 Preliminary Investigations

5.1 Pressure Tests

To ensure that the additively manufactured tubular structures can withstand the static pressure during overmolding, a pressure test is performed. This test is also intended to ensure that the walls are printed tightly enough to prevent plastic from leaking in. For

this purpose, individual channels printed on a test basis from titanium were placed in a pressure vessel filled with water and subjected to the pressure expected from injection molding (1 × 30 bar, 1 × 60 bar). After one minute, the water was drained, the test specimens were removed, externally dried and opened with a cutting grinder. It was shown that the printed tubes withstood the pressure without any leakage and that the enclosed metal powder showed no signs of moisture.

5.2 Powder Investigations

To determine, if the Ti6Al4V metal powder from inside the channels cold be recycled and reused, it was analyzed under the microscope and compared with new powder and powder from the printing bed. The left Fig. 6 shows new powder. In contrast, the middle figure shows the already sieved powder from the area next to the print object. In the right figure, the powder recovered from inside the cooling tube is shown. It can be seen in this optical analysis that the quality of the powder from the inside of the tube is not distinguishably worse than that of the powder next to the object. So it a recycling of this powder can be seen as a realistic option. From one HEaK-tube, 2 grammes of powder could be recovered (72 grammes per motor insert structure). So, by weight, roughly half of the inserting structure is enclosed powder (Fig. 7).
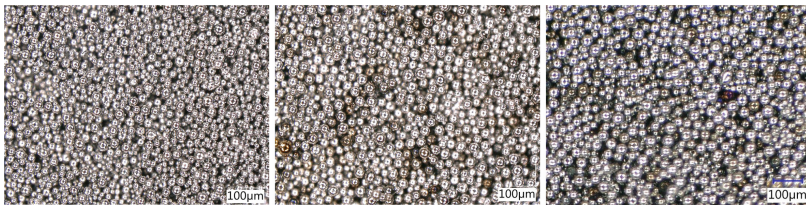
Fig. 7. **Left:** New powder; **Middle:** Powder from the printing bed; **Right:** Powder from the inside of the printed structure

6 Conclusion

In this paper the integration of additive manufactured cooling channel inserts in over-molded high-performance devices was proposed as a concept and designed out on the example of the HEaK e-motor project. It is suggested to concentrate multiple functions in these inserts, to make up for their production costs. Furter cost saving could be achieved with thin-walled structures, filled with remaining powder which was removed after the molding process. First pressure tests of channel inserts and a powder investigation showed positive results.

References

1. Rhebergen, C., Bilgin, B., Emadi, A., Rowan, E., Lo, J.: Enhancement of electric motor thermal management through axial cooling methods: a materials approach. IEEE Energy Conv. Congr. Expos. (ECCE) **2015**, 5682–5688 (2015). https://doi.org/10.1109/ECCE.2015.7310458
2. Reuter, S., Berg, L.F., Doppelbauer, M.: Performance evaluation of a high-performance motor with thermoset molded internal cooling. In: 2021 11th International Electric Drives Production Conference (EDPC), pp. 1–5 (2021). https://doi.org/10.1109/EDPC53547.2021.9684222
3. Chinthavali, M., Ayers, C., Campbell, S., Wiles, R., Ozpineci, B.: A 10-kW SiC inverter with a novel printed metal power module with integrated cooling using additive manufacturing. IEEE Workshop Wide Bandgap Power Dev. Appl. **2014**, 48–54 (2014). https://doi.org/10.1109/WiPDA.2014.6964622
4. Wrobel, R., Mecrow, B.: Additive manufacturing in construction of electrical machines – a review. In: 2019 IEEE Workshop on Electrical Machines Design Control and Diagnosis (WEMDCD), pp. 15–222019https://doi.org/10.1109/WEMDCD.2019.8887765
5. Verma, S., Yang, C.-K., Lin, C.-H., Jeng, J.Y.: Additive manufacturing of lattice structures for high strength mechanical interlocking of metal and resin during injection molding. Addit. Manuf. **49**, 102463 (2022). https://doi.org/10.1016/j.addma.2021.102463
6. Reuter, S., Sorg, T., Liebertseder, J., Doppelbauer, M.: Design and evaluation of a houseless high-performance machine with thermoset molded internal cooling. In: 2021 11th International Electric Drives Production Conference (EDPC), pp. 1–6 (2021). https://doi.org/10.1109/EDPC53547.2021.9684226
7. Innovations Campus, EM2 – HEaK – Hocheffizienter Elektromotor mit additiv gefertigtem Kühlsystem in Kunststoffumspritzung (2022). https://www.icm.kit.edu/downloads/ICM-Projekte/ICM_Kurzvorstellung-Projekt-EM2-HEaK.pdf
8. Möhring, H.-C., Becker, D., Maucher, C.: Spanntechnik beim Absägen von AM-Bauteilen, wt Werkstatttechnik. **112**, 61–66 (2022)

The Potential of High Speed Sintering for Small Series in the Automotive Industry

Timo Huse[1] and Laura Rehberg[2(✉)]

[1] German Aerospace Center (DLR), Pfaffenwaldring 38-40, 70569 Stuttgart, Germany
[2] University of Stuttgart, Pfaffenwaldring 19, 70569 Stuttgart, Germany
`laura.rehberg@eni.uni-stuttgart.de`

Abstract. Individualization in the high-performance segment within the automotive industry is becoming increasingly important. Especially when small volumes are required, conventional manufacturing processes often no longer prove profitable. The use of additive processes in general and high-speed sintering (HSS) in particular offer the freedom to produce complex organic shapes in a cost-effective and resource-saving manner from batch size one onwards. The HSS process is a powder bed-based additive manufacturing process in which thermoplastics are sintered at a constant layer time using an infrared lamp instead of a laser

For this reason, we shed light on the use of high-speed sintering specifically for small components in low-volume production. More precisely, we add the process-specific properties of high-speed sintering to Design for Additive Manufacturing (DfAM) rules. We propose an approach that also enables a time-saving alternative to conventional manufacturing processes and optimizes the design process for the use of high-speed sintering.

Keywords: Additive Manufacturing · Automotive · DfAM

1 Introduction

The challenges and associated pressures for companies are growing due to a lack of raw materials [1], collapsing sales markets [2], and the governmental push for ecological sustainability [3].

The use of additive manufacturing processes is driving the efficiency of the value chain and has thus experienced strong development in recent years. The successive addition of material allows complex geometries to be produced and eliminates many design and manufacturing constraints [4]. The use of additive manufacturing processes offers opportunities in pre-series production and prototyping and enables OEMs to test radically different design methodologies [5]. Especially in highly competitive sectors, product innovations are essential to survive as a company [6]. In particular, the topics of lightweight construction and individualization play an important role here [7].

For example, Seidel et al. [8] describes that the customized car will consist of mixing and matching standard components and that the customer can also incorporate individual wishes regarding the shape and style of the components [8].

The use of additive manufacturing, for example, enables personalized components based on customer wishes in a competitive lead time. In addition, individual spare parts can be manufactured, for example, if they are no longer available on the market [9]. While conventional manufacturing processes are limited in design freedom, the complexity of additive manufacturing is not a challenge [10].

This paper investigates the potential of high-speed sintering for the automotive industry for small-batch manufacturing. Using a speaker cover as an example, we develop design guidelines and show possibilities for the use of personalized interior components. The rest of this paper is organized as follows. The following section describes rapid manufacturing and the benefits as well as challenges. This is followed by a description of the High-Speed Sintering manufacturing process and the experimental setup for manufacturing a personalized speaker cover. Finally, recommendations for use in the automotive industry and HSS-specific design guidelines are derived.

This paper aims to demonstrate to design engineers the possibilities of High-Speed Sintering, especially for small personalized batches. Furthermore, we would like to contribute with this paper to the design guidelines in additive manufacturing and extend them with HSS-specific requirements.

2 Rapid Manufacturing vs. Rapid Prototyping

Rapid manufacturing refers to methods and manufacturing processes that produce components from CAD data with the aid of tool-free production. Part of this is the now mature rapid prototyping. This includes technologies such as stereolithography or laser sintering [11]. The objective here is the direct production of 3D products that are used either as parts of assemblies or as whole products by the end user [12].

The increasing acceptance is based on the savings in terms of tooling costs as well as shorter lead times. Furthermore, the use of rapid prototyping allows unique freedom in product design [12]. Due to automation, minimal human intervention is required. Thus, complex 3D geometries can be produced automatically without part-specific fixtures or tooling. In addition, components are manufactured directly as assemblies and do not require subsequent assembly. This makes rapid prototyping particularly suitable for small series [13]. However, the application of rapid manufacturing in general and rapid prototyping in particular is also subject to challenges. Market acceptance contributes to the extent to which proven molded parts can be replaced by layered versions. Furthermore, influencing factors such as different machines and working methods can lead to limited reproducibility of parts. In addition, the costs must be analyzed in order to identify suitable products that are suitable for rapid manufacturing [12].

3 Introduction to High Speed Sintering

The High-Speed Sintering (HSS) process is a powder-based additive manufacturing technology for thermoplastic materials in which components are selectively fused using black ink and an infrared lamp. First, a thin white powder layer is applied to a building platform. A print head then applies black ink locally to those areas of the powder layer where a component is later to be generated. With the use of an infrared lamp, heat is then

applied to the entire powder layer, causing the black areas to absorb it more strongly than the white areas. This causes the powder to exceed the melting temperature only at the desired black areas, where it melts and sinters [14]. This process is repeated layer by layer until the desired part has been generated. In this case, the Voxeljet VX200 HSS (see Fig. 1) system is used, which has a build volume of 290 x 140 x 180 mm.

Fig. 1. Voxeljet VX200 HSS

Figure 2 below shows the build platform during the sintering process through the infrared lamp passing over each layer. In the previous step, the print head colored the areas to be sintered black.
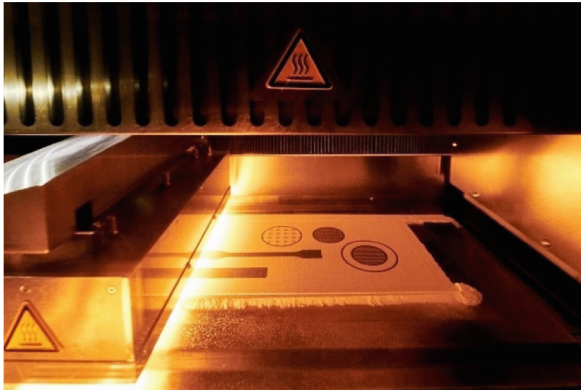
Fig. 2. Sintering process

4   Research Approach

The overall objective of the research project is the efficient, sustainable use of additive manufacturing processes in the automotive industry.

We present a new approach to reducing the environmental impact using waste powders. The open machine concept allows the influence on a multitude of process parameters. This is used in this paper and applied to a specific application in the automotive industry.

In addition, complex components with very high lightweight potential can be produced without a support structure while maintaining the same shift time.

In order to maximize the sustainability of the process and at the same time increase the economic efficiency, the aging of the powder is also investigated in addition to the structural optimization in order to be able to add as much recycled powder as possible to a new process. The powder used consists on average of 80% recycled powder.

The process was tested and optimized in an industrial environment as part of a longitudinal study over two years. Prototypes were produced for our research purposes and in cooperation with other institutes in topology optimization, function integration, composite materials, and powder recycling, among others.

Previous research has focused on applying high-speed sintering in general and industry. The present work provides new insights into the optimized design of components specifically for applying high-speed sintering processes with reused powder.

The development of the design guidelines is based on the literature on the one hand and the findings of two years of research on the other hand. In iterative loops, parameters were adjusted, and many prototypes were produced to validate the findings. These form the basis for the properties' weighting for the pairwise comparison application and the elaboration of the use case of the personalizable speaker cover.

5   Development of Design Guidelines for HSS

In the field of additive manufacturing, there are now a large number of design guidelines that apply to different additive processes across the board. These include, for example, that the build time of a job depends significantly on its Z-height or that anisotropic material properties are present in most components due to the process. Among other things, within powder-based additive manufacturing processes, components usually achieve the highest mechanical properties with a load direction parallel to the X-Y plane. However, if the component is loaded in the Z direction, the lowest mechanical properties are usually present here. This is due to the fact that the bond between two layers is weaker than the cohesive material within a layer. The same applies to dimensional accuracy. Circular holes or a groove should lie in the profile in the X-Y plane to obtain the target geometry with a high accuracy [15]. If they are not perpendicular to this plane, e.g. perpendicular to the X-Z plane, round holes tend to become oval and step along the circumference due to layer changes.

Beyond the partially general design guidelines that can be applied to this process, further HSS-specific design specifications should be considered. These include, if possible, designing components or aligning them in the build space so that individual layers form the smallest possible contiguous black areas on the X-Y plane. Since the black-colored areas absorb significantly more thermal energy, large contiguous black areas can overheat in some spots. This leads to an inhomogeneous temperature distribution and thus also to inhomogeneous cooling behavior, which can lead to warpage in the component. However, it must be considered that even with this method, depending on the material, there are usually anisotropic material properties. In general, the mechanical properties are at the highest in the X-Y plane and less in the Z direction [15]. Furthermore, it should be clear if the mechanical properties or the accuracy of the geometrics are prioritized and, if needed, adapted to the specific application of the component. In this process, the aim is to place as many components as possible on the building platform in the X-Y direction, since the layer time is always the same [15]. This results in optimized use of space and shortened manufacturing time. In contrast to the SLS process, HSS doesn't need a laser to scan each geometry individually; instead, the infrared lamp always covers the entire powder layer as it passes over it. Hence, the height in the Z direction is therefore decisive for the total manufacturing time of the job to be done (Table 1).

Table 1. Design guidelines for HSS

1	**Mechanical properties**	In a three-dimensional build volume (X-Y-Z), the mechanical properties of components with a loading direction in the X-Y plane are higher than in the Z direction
2	**Dimensional accuracy**	For example, the profile of circular holes should be in the X-Y plane to obtain the best possible accuracy
3	**Small areas per layer**	Large accumulations of black-colored areas in the X-Y plane should be avoided to prevent inhomogeneous temperature distribution
4	**Low z-height**	Components should be oriented with a low Z-height in the build volume to reduce manufacturing time

The developed design guidelines were applied to the example of a speaker cover and are presented below.

6   Use of HSS for the Manufacturing of a Speaker Cover

The use of additive manufacturing processes in the automotive industry has expanded from prototyping to production technology. Small batches are particularly suitable in terms of time and cost savings. In addition, components can be personalized. In the following, an application example for the use of high-speed sintering in interiors is

presented. The suitability of potential parts or components can be determined on the basis of the following criteria:

- no safety relevance
- can be personalized (in terms of branding and design)
- possibility of independent replacement by customers

It is important here that, although personalization is carried out by the customer, it is not tied to the customer over the lifecycle.

In the following, the application for HSS is presented using the example of a speaker cover. It is important to note that in most cases not all design guidelines can be considered at the same time. They must always be adapted to the specific target component and its desired properties. If, for example, particularly high mechanical properties are required of a component in a certain spatial direction, the orientation of the component must be adapted in such a way that this may be disadvantageous for another property. This could be, for example, the dimensional accuracy of certain areas. Thus, a prioritization of the component properties and thus also of the design guidelines must always be carried out in order to optimize a component for the required application. Using the speaker cover as an example, we will show below how the process-specific design guidelines of the high-speed sintering process, in particular, could be prioritized for this application.

In the pairwise comparison, the four selected parameters were compared and weighted to the specific use case of the speaker cover. A rating of 2 means that the row value is more important than the column value, 1 means that it is equally important and 0 means that it is less important. The line values per target value added together then result in the ranking.

The following Table 2 shows the pairwise comparison that was performed for the application regarding the speaker cover.

Table 2. Pairwise comparison

	Mechanical properties	Dimensional accuracy	Small areas per layer	Low Z-height	Total
Mechanical properties		0	2	2	4
Dimensional accuracy	2		2	2	**6**
Small areas per layer	0	0		2	2
Low Z-height	0	0	0		0

Since this component is not a structural part and therefore does not have to absorb any forces, the first design rule for optimizing the mechanical properties in a specific spatial direction is not prioritized. However, the second rule concerning dimensional accuracy is more important here, since it represents a design component in the interior that is visible to the end customer. Therefore, the geometry should be achieved as accurately

as possible in order to ensure that it can also be assembled. Applying this requirement, the profile of the circular holes should be in the X-Y plane to obtain the best possible accuracy (see Fig. 3).
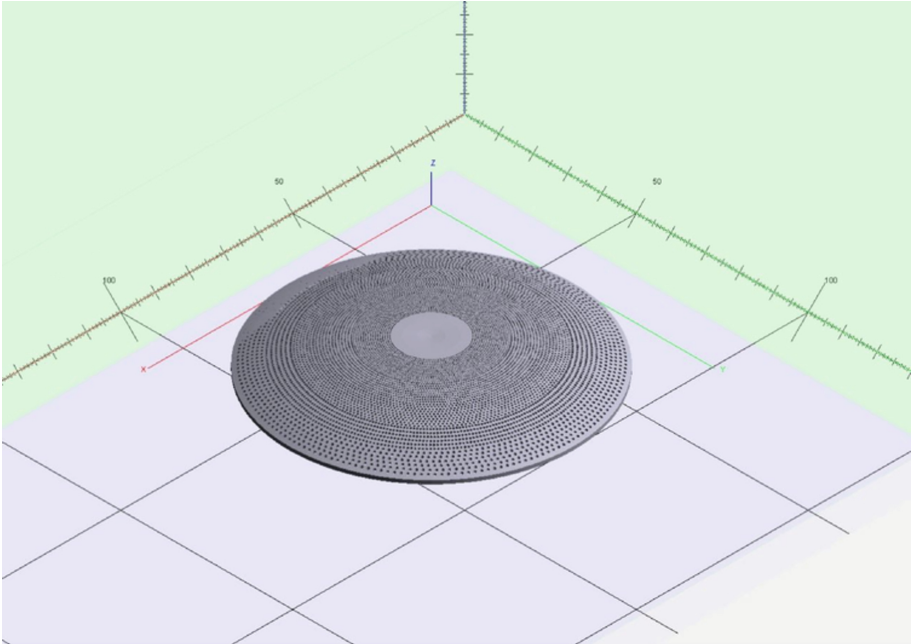
Fig. 3. Oriented CAD-Model for optimized accuracy

The process-specific third rule cannot be explicitly considered here because, first, the design is predetermined for this component and, second, dimensional accuracy is the requirement with the highest prioritization. However, this leads in this specific case to the presence of large black areas within a layer. Furthermore, it can lead to an inhomogeneous temperature distribution in the component and to warpage. To counteract this effect, the targeted introduction of grayscale into the component could compensate for this effect in the future. As a result of the described orientation, the shortest build time is automatically achieved, as long as the orientation has the lowest Z-height. However, it must be assumed that the build space is always fully utilized within this process.

7   Conclusion

In this work, we investigated the design guidelines for rapid manufacturing in general and for high-speed sintering in particular.

With the overall objective of enabling efficient and sustainable use of additive manufacturing processes in the automotive industry, waste powder was used as part of the research to conserve resources. In addition, the developed design guidelines enable a more efficient production by an optimized prioritization of the developed parameters.

The results were applied and validated with manufactured prototypes over a period of two years. They were transferred to the automotive industry using the example of the speaker cover. Hereby, we provide an important contribution to the application of customizable parts and components for small series in the automotive industry.

We found that four parameters need to be adjusted to meet the requirements of the part: *Mechanical properties, dimensional accuracy, small areas per layer, low z-height.* The prioritization of a parameter influences the other three parameters depending on the geometry of the component, as these are directly dependent on each other. In the case of design components (e.g., speaker cover) whose geometries are specified, priority is given to dimensional accuracy in order to be able to guarantee aesthetics and assemblability in the vehicle. In this example, this leads to a deterioration of the "small areas per layer", since the black-colored areas per layer are larger and thus lead to an inhomogeneous temperature distribution. At the same time, the parameters "mechanical properties" and "low z-height" are improved.

The geometry of the components is crucial for the possible degree of influence. For example, spherical components cannot be oriented in the build-volume and influenced regarding the improvement of the parameters.

The introduction of HSS for custom components has far-reaching implications for the development process.

- Impact on design CAD model
- Impact on manufacturing
- Impact on post-processing

We distinguish between components that are visible to the customer and those that are not. Visible parts, for example, require post-processing, e.g., to smooth or seal the surface.

In this research, the focus was on the design guidelines and the parameters described above. The results show that the design guidelines must always be adapted to the specific application. Here, these must be prioritized according to the specific requirements of the component.

8   Limitations and Further Research

The use of rapid manufacturing has several advantages. In addition to the design freedom and the possibility to print complex structures without tools, assemblies can be printed as a whole. Furthermore, the use of rapid manufacturing in general and high speed sintering, in particular, is especially suitable for personalized components.

Nevertheless, its use in industry is still limited. Although high speed sintering enables the realization of complex geometries, components that are visible to the customer still have to be smoothed and finished. In addition, there are still no rules governing the suitability of components in the vehicle in terms of market acceptance. HSS-specific design guidelines were developed as part of the publication. These are based on the usage of the Voxeljet and should be further tested and validated regarding the use cases in the automotive industry.

In addition, the reproducibility and degree of customization would need to be investigated from an economic perspective with respect to various components. The validity of

the rules is therefore limited to the context under consideration and the prevailing boundary conditions (e.g. materials). The application of different materials and the effects on the presented parameters provides a fundament for further research in this area.

References

1. Linz, S., Neumann, D., Abdalla, S., Gladisdörr, G.: AUSWIRKUNGEN DER CORONA PAN DEMIE: LIEFERENGPÄSSE BREMSEN INDUSTRIE UND TREIBEN PREISE. Statistisches Bundesamt | WISTA |, vol. 1, pp. 71–82 (2022)
2. Stepanov, A.A., Gorin, V.S., Tettsoeva, O.A., Merenkov, A.O.: Problems of the Automotive Business Ecosystem in the Context of a Pandemic (2021)
3. Lempp, M., Siegfried, P.: Characterization of the automotive industry. In: Lempp, M., Siegfried, P. (eds.) Automotive Disruption and the Urban Mobility Revolution: Rethinking the Business Model 2030, pp. 7–24. Springer International Publishing, Cham (2022). https://doi.org/10.1007/978-3-030-90036-6_2
4. Khorasani, M., Ghasemi, A., Rolfe, B., Gibson, I.: Additive manufacturing a powerful tool for the aerospace industry. Rapid Prototyping J. **28**(1), 87–100 (2022). https://doi.org/10.1108/RPJ-01-2021-0009
5. Hague, R.: Unlocking the Design Potential of Rapid Manufacturing. In: Rapid Manufacturing: An Industrial Revolution for the Digital Age, pp. 5–18. John Wiley \& Sons (2006)
6. Baregheh, A., Rowley, J., Sambrook, S.: Towards a multidisciplinary definition of innovation. Manag. Decis. **47**(8), 1323–1339 (2009). https://doi.org/10.1108/00251740910984578
7. Wiberg, A., Persson, J., Ölvander, J.: Design for additive manufacturing – a review of available design methods and software. Rapid Prototyping J. **25**(6), 1080–1094 (2019). https://doi.org/10.1108/RPJ-10-2018-0262
8. Seidel, M., Loch, C.H., Chahil, S.: Quo Vadis, automotive industry? A vision of possible industry transformations. Eur. Manage. J. **23**(4), 439–449 (2005). https://doi.org/10.1016/j.emj.2005.06.005
9. Heinen, J.J., Hoberg, K.: Assessing the potential of additive manufacturing for the provision of spare parts. J. Oper. Manag. **65**(8), 810–826 (2019). https://doi.org/10.1002/joom.1054
10. Seepersad, C.C.: Challenges and opportunities in design for additive manufacturing. 3D Print. Addit. Manuf. **1**(1), 10–13 (2014). https://doi.org/10.1089/3dp.2013.0006
11. Hague, R., Campbell, I., Dickens, P.: Implications on design of rapid manufacturing. Proc. Inst. Mech. Eng. C J. Mech. Eng. Sci. **217**(1), 25–30 (2003). https://doi.org/10.1243/095440603762554587
12. Hopkinson, N., Dickens, P.: Rapid prototyping for direct manufacture. Rapid Prototyping J. **7**(4), 197–202 (2001). https://doi.org/10.1108/EUM0000000005753
13. Ingole, D.S., Kuthe, A.M., Thakare, S.B., Talankar, A.S.: Rapid prototyping – a technology transfer approach for development of rapid tooling. Rapid Prototyping J. **15**(4), 280–290 (2009). https://doi.org/10.1108/13552540910979794
14. Hopkinson, N., Erasenthiran, P.: High Speed Sintering – Early Research into a New Rapid Manufacturing Process (2004)
15. Bikas, H., Lianos, A.K., Stavropoulos, P.: A design framework for additive manufacturing. Int. J. Adv. Manuf. Technol. **103**(9–12), 3769–3783 (2019). https://doi.org/10.1007/s00170-019-03627-z

Asset Administration Shell for the Wiring Harness System

Georg Schnauffer[(✉)], David Görzig, Christian Kosel, and Johannes Diemer

ARENA2036 E.V., Pfaffenwaldring 19, 70569 Stuttgart, Germany
info@arena2036.de

Abstract. The wiring harness is one of the most expensive components of the automobile. It is still a custom-made product despite the high-volume production in the automotive industry. However, various trends, such as increasing fragility of supply chains and functional safety demands for autonomous driving are demanding the automation of the value creation processes. But in mass customisation, automation needs the digitalisation of all information flows. Machines and processes must be able to automatically gather all required information. In the wiring harness industry, much of the information has been geared toward processing by humans and has hardly been digitised at all. One of the foundations for Industry 4.0-oriented automation is ergo the existence of complete digital information of each wiring harness. A digital twin of the wiring harness is required. The Asset Administration Shell (AAS) is the technical concept for transparently interoperating with the digital twin. The project "Asset Administration Shell for the Wiring Harness" or in German (Verwaltungsschale für den Leitungssatz - VWS4LS) creates the first complete digital description of the wiring harness based on the AAS. As a result, all actors in the value network can access complete and consistent data ("single point of truth"), enrich it, and control its application through authorisations. To achieve this goal, a consortium from all levels of the value network was assembled for VWS4LS starting with the OEM, through the suppliers at the various levels of the value chain, and to the software suppliers and machine manufacturers. This publication presents the project and its initial results.

Keywords: Digital Twin · wiring harness · Asset Administration Shell · VWS4LS

1 Introduction

The wiring harness is the electrical and communication backbone of the automobile. As a key element of the vehicle electrical system, it is fundamental to enabling the future trends of electromobility and autonomous mobility [1]. Automotive industry is characterised by its high-volume production of cars generally with the same equipment in a relatively repetitive manner; though, the complete wiring harness, as a customised, intricate product, amplifies its overall complexity and time [2]. As a result, the wiring harness is not only one of the most expensive components in the automobile, but it is also associated with a high degree of manual work delivered by several hundred thousand jobs worldwide.

Aside from the still heavily manual production, the wiring harness faces further challenges by the vulnerabilities of the supply chains posed by geopolitical conflicts, natural disasters, and pandemics. An additional important technical aspect is the absence of continuous digital data chains along the value chain. Software solutions are offered in proprietary silos and the digital flow of information is (consequently interrupted in many places.

From the point of view of the industry, the automation of information flows within the value chain would be a critical success factor in overcoming these challenges. So far, however, a lot of information has been geared towards human processing and has therefore only been minimally digitised. A possible approach could be the implementation of the (asset) administration shell in the wiring harness industry.

This publication first describes the value chain of the wiring harness and the associated challenges and developments. A special focus is on the aspect of the availability of wiring harness-related data over the entire life cycle. Subsequently, the concept of an Asset Administration Shell (AAS) is presented. The AAS will be used for the digital representations of all components of the wiring harness. Afterwards, the project "Asset Administration Shell for the Wiring Harness" (VWS4LS) is introduced. One major task of VWS4LS will be the creation of the model description as part of the AAS. In the second part, the current project status is presented and a brief outlook on further activities is given.

2 Changing the Value Chains for Wiring Harness

2.1 Wiring Harness in Current Value Chains

In the 1960s and 1970s, wiring harnesses tended to have the character of a series component due to the low level of electrification in automobiles. In the 1990s, the paradigm of the so-called KSK - the "customer-specific wiring harness" - has dominated since the 1990s. This term alone shows that this is no longer a mass-produced product, but a batch size of 1. Due to the drastic customisation of equipment in vehicles, German OEMs are installing KSK without exceptions. Up to now, this requirement has been mainly met by highly manual production processes in so-called "best cost countries" (low-wage countries). At the same time, the complexity of the wiring harness and the associated value chain continues to increase.

In the meantime, however, various trends in the automotive sector are forcing the automation of value creation processes. The resilience of value chains is ergo becoming increasingly important. Events or- vulnerability inducers, such as the COVID-19 pandemic, the grounding of the Ever Given in the Suez Canal, and the Russo-Ukrainian war have exposed the fragility of value chains. Without automation, greater regionalisation of production is inconceivable. However, a pure focus on the development of means of production with which manual activities can be replaced by automation will not be sufficient. Rather, in addition to the means of production, it is also necessary to consider their informational linkage in the sense of a continuously automated production value chain - each work step must be completely recorded or by means of information technology.

In addition, new challenges are arising around traceability of automotive components, partly caused by new regulations, such as the German Supply Chain Compliance

Act. Identifying human rights violations at all levels of their own supply chain will be the responsibility of each company. New regulations regarding sustainability will require one to record and aggregate the CO_2 footprint of each individual process step in the value chain with the purpose of initiating targeted improvements. Another important trend is the development of autonomous driving towards Level 4 and 5, which entails increasing data rates and higher requirements in functional safety [1].

2.2 Vision – Uniform Data Representation Along the Complete Life Cycle

One of the foundations for Industry 4.0-oriented automation of the wiring harness is the availability of complete digital information on each wiring harness, based on which development and optimisation can be realised along the entire life cycle from a wide variety of perspectives.

The data on the wiring harness can be used, enriched, and further developed by a wide variety of players. It is not only important that everyone has access to the same database during the engineering phase, but also during the subsequent (distributed) production and assembly process right through to service if necessary. All relevant components of both the "wiring harness" product and the respective production components (e.g. production equipment and processes with parameters) thus require a comprehensive digital description ("digital twin"). The decisive factor is not only the description itself, but also its automated interpretation, so that all subsequent processes can be automated as far as possible, especially in the case of changes.

This means that, based on a complete digital description, the systems of all the players in a value network can access complete and consistent data ("single point of truth"), enrich it, and control its use by means of authorisations. The digital twin is a fundamental prerequisite for end-to-end digital process chains across companies including the hierarchical description of components and their sub-components The AAS, in turn, is the technical concept of the interoperable digital twin of the wiring harness. The AAS, therefore, provides a comprehensive conceptual basis that can make a significant contribution to the solution.

2.3 The Asset Administration Shell

When the final report of the Industry 4.0 working group was published in April 2013, the question "What will the future look like under Industry 4.0?" was responded with the following statement: "Industry 4.0 will deliver greater flexibility and robustness together with the highest quality standards in engineering, planning, manufacturing, operational and logistics processes. It will lead to the emergence of dynamic, real time optimised, self-organising value chains that can be optimised based on a variety of criteria such as cost, availability, and resource consumption. This will require an appropriate regulatory framework as well as standardised interfaces and harmonised business processes" [3].

By developing the concept of Reference Architecture Model Industry 4.0 (RAMI4.0) and the related Industry 4.0 component, the basis for the implementation of the Asset Administration Shell (AAS) was set [4]. Primarily the AAS is introduced the digital representation of an asset. The concept supports the description of complex assets with all their components by using a hierarchical structure of AAS and further allows one to

link AAS of different assets. While current implementations of the AAS focus on the description of real physical assets, we expect soon, that information, functions, and as well as contracts will also be represented within an AAS. [5] the AAS, ergo, represents an approach to realise many of the ideas written down back in 2013.

2.4   Potentials by Using Digital Twins

By using the digital twin, numerous use cases can be realised over the life cycle of the wiring harness. For example, requirements can be transmitted digitally between OEMs and suppliers and components can be automatically selected in engineering. In addition, change management can be made much more efficient. With the help of the resultant database, new applications for simulation are emerging. As a result, optimisation potential in engineering, manufacturing, and recycling can be uncovered and realised.

The digital twin can not only be used in the development and production of the wiring harnesses. The data therein can also be used in repair shops to find faults more quickly, procure spare parts, and carry out repairs. In recycling, the data from the digital twin is valuable for quickly identifying which wiring harness is installed in the vehicle. This information can be used to quickly decide whether the wiring harness can be offered as a spare part or sent directly for recycling.

3   Asset Administration Shell for the Wiring Harness (VWS4LS)

The project "Asset Administration Shell for the Wiring Harness" (VWS4LS) has set the goal of developing a consistent representation of the Asset Administration Shell (AAS) for the wiring harness as a digital twin across all stages of the value chain from its initial specification to its disassembly.

The basis is formed by comprehensive information models of the product, production process, and means of production, which are uniformly designed throughout the industry. The norms and common standards in the industry are adhered to and extended as required. Uniform formats and protocols are required for the exchange of data so that the participants in the value chain can understand each other in terms of data technology. The starting point is the already known and used data types of the Cable Harness List (KBL) and the Vehicle Electric Container (VEC).

Another aspect is the infrastructure for data storage and data exchange for automated production control. The company-wide and cross-company exchange of data will increasingly take place in the cloud in the future. Group-specific, industrial, cloud-based platforms are currently being established by the OEMs (e.g., BMW Open Manufacturing Platform, Mercedes-Benz Cars Operations 360 (MO360), and Volkswagen Industrial Cloud). In the first step, the focus is initially on the internal group platforms for data exchange. In further steps, suppliers and service providers will also be integrated into these platforms.

While the concept supports hierarchical structures, today the AAS is mainly used for describing individual components and applying different types of sub-models and protocols, such as OPC UA and E-Class. In the future, however, the combination of AAS in hierarchical structures for assemblies will gain relevance. The AAS for the

wiring harness will generate a first example illustrating the effects on and the associated advantages of the concept of a hierarchical structure and the possibility describe relations of AAS via links.
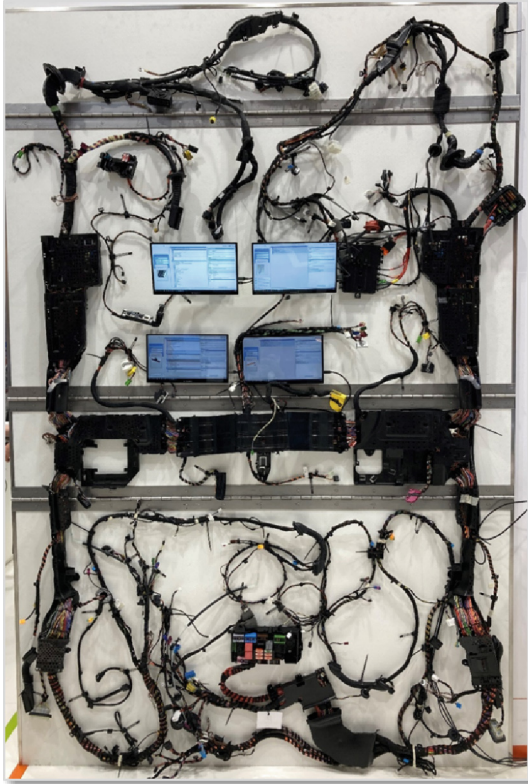
Fig. 1. Demonstrator panel for the Wiring Harness of a Mercedes Benz C-Class

Figure 1 shows a demonstrator panel of a real wiring harness of a Mercedes Benz C-Class with four screens, on which the digital representation as ASS of several components can be viewed (demonstrator is 3m high and 2m wide). The wiring harness therein gives an idea of how many individual components constitute the "wiring harness" product. Depending on the perspective of the value-adding company, the composite component or assembly can be defined differently. For the assembler, for example, the product is "the wiring harness" itself. The component manufacturer may already see a connector and connector housing as a composite component.

Furthermore, the goal of VWS4LS is to generate the AAS not only for the wiring harness product, but also for the machines used during production. The next step is the mapping of production processes in the AAS by linking the product and the production equipment. Doing so allows one to collect data, for example, on the quality of the production and the incurred costs. Those data can be supplemented by additional information

collected over the entire life cycle. The combined information can be comprehensively used to achieve a required level of quality within the given within a given set of financial and economic parameters.

4 Status of Development

The project is divided into ten subprojects and five use cases. In addition to the main areas of the value chain (development, production, and assembly), the subprojects include an information model and innovative topics, such as automated negotiation processes, data governance concepts, and the connection to Catena-X. The start-up phases of the sub-projects are sequentially timed. This is done in order to integrate the experience and content from the value-added areas into the more innovative topics.

As can be seen in Fig. 1, several thousand individual components and, depending on the angle of view, different composite components constitute the wire harness product.

The development of the information model serves as an important basis for the rest of the project. Since the digitisation of an entire value chain, with the goal of automation, is the principal focus, the product, process, and resource model (PPR model) were specifically used as the basis. The breakdown is intended to identify digitalisation gaps in the various areas of the value chain that are not yet covered by established data standards, such as KBL, VEC, and OPC-UA. Simultaneously, for the analysis of the current state of the value chain, the partners defined requirements for an end-to-end digitalised and automated value chain process. These requirements are considered as additional inputs for both the data model and the individual value creation areas.

The development process of the wire harness is fundamentally collaborative between different value creation partners. The resulting challenge is to develop a concept that supports the collaborative (shared digital twin) working method in the development of the control set, while maintaining data sovereignty and access. The basis for this was a reference process jointly developed in the project, which was supplemented with standardised inputs and outputs of the process steps. The known inputs and outputs are now to be converted into partial models, and thus, digitally, and automatically retrievable by means of a unique reference ID.

Like the development process, a reference process was also developed for the man-ufacturing and assembly of a wire harness. The aim was to obtain an overview of the information provided and required in the individual process steps. The reduction of man-ual activities and the information required for this essentially form the central element here. The current focus of work is to convert the reference process and its information into sub-models and then to make them available across companies through the AAS.

The development of the AAS and its completion with data, as shown in Fig. 2, is currently still manually carried out. The basis is formed by the sub-models standardised and published by the Industrial Digital Twin Organisation (IDTA). A standardised range of information for the most diverse application areas can be covered. For the digital continuity of the wiring harness value chain, this is not sufficient yet. For this purpose, our own sub-models will be developed at a later stage to meet the industry-specific data requirements. The implemented product data was provided by partners of the project VWS4LS and ARENA2036. The semantic required to be able to unambiguously interpret

data and information is ECLASS. ECLASS is a cross-industry classification standard that can uniquely describe information by a value [6].

Figure 2 shows the composite component in the AASX package explorer of a line set section. The left side of the figure represents the Industry 4.0 component, which consists of the asset (gray) and the AAS (blue). The lower left third depicts the repository, through which the AASs of the individual components of the composite component can be selected. The middle third shows a hierarchical structure from the AAS, through which various sub-models (SM) to the sub-model element (carries the respective information) are shown. The right third shows the information stored in the sub-model element and much more. In this case the hierarchical structure of a Bill-of-Material (BOM) can be seen there. The composite component shown here consists of five individual components (tape, cable, clip, connector, and flat connector housing), which all have an entity. The individual entities are then related to each other by the "Relationship Element" sub-model element.
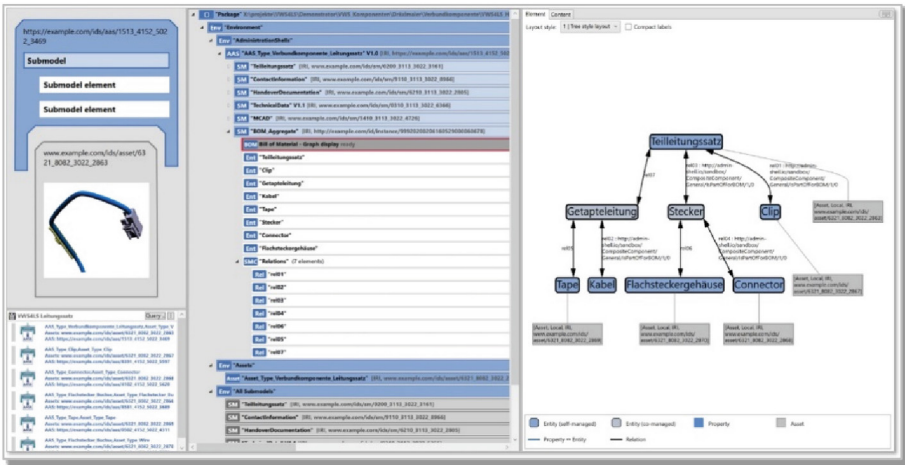
Fig. 2. AAS of a composite component

5   Conclusion and Further Development

The VWS4LS project has successfully completed the first steps, though there is more to come. New use cases have become possible. The creation of the complete wiring harness will enable a continuous access to all the data of the related supply chain. Based on the complete information, the system engineering will be significantly more efficient, especially given that they may even use automated processes. Integrated information, such as the carbon footprint of the complete wiring harness assets, including manufacturing can be determined. The data will be available over the complete life cycle allowing use cases, such as the components' re-use and recycling (Circular economy).

In the future, numerous further technical topics need to be addressed. One important aspect is the automated merging of AAS. The wiring harness consists of numerous individual components from different suppliers that are combined to form an assembly. Today, this is done with a great deal of manual effort. Within the project, the possibilities for automated production are to be investigated. An important aspect is change management; some questions that would arise herein include, but are not limited to the following: What happens to data when components change? And how can traceability still be guaranteed?

A particularly important topic in VWS4LS is the monetisation of data. Within the framework of the project, considerations on new business models and their implementations with the AAS must be developed. In this context, a data storage policy needs to be developed that defines who has access to what data and under what conditions. Another important aspect is automated negotiation processes. Here, negotiation scenarios and strategies need to be explored.

A very important aspect that is not covered in VWS4LS is how the data gets into the AAS. Since the new generation of AAS is currently still a manual process supported by tools, we see a great benefit by semi-automating the generation of assets by using semantic interoperability based on a neural language model [7].

References

1. Brabetz, L., Bordnetz, D.: In: Pischinger, S., Seiffert, U. (eds.) Vieweg Handbuch Kraftfahrzeugtechnik, vol. 9, pp. 877-911. Springer, Heidelberg (2021)
2. Hummel, H., Sedlmeier, R.: Das Bordnetz. ATZextra **15**(5), 97-99 (2010)
3. Kagermann, H., Wahlster, W., Helbig, J.: Recommendations for implementing the strategic initiative INDUSTRIE 4.0. Final report of the Industrie 4.0 Working Group. acatech, München (2013)
4. Hankel, M., Rexroth, B.: The reference architectural model industrie 4.0 (Rami 4.0). ZVEI **2**(2), 4–9 (2015)
5. Asset Administration Shell Reading Guide. https://www.plattform-i40.de/IP/Redaktion/EN/Downloads/Publikation/AAS-ReadingGuide202201.html. Accessed 20 June 2022
6. Both, M, Müller, J., Diedrich, C.: Automatisierte Abbildung semantisch heterogener I4.0-Verwaltungsschalen durch Methoden des Natural Language Processing. at – Automatisierungstechnik **69**(11), 940–951 (2021)
7. Xia, Y., Jazdi, N., Weyrich, M.: Automated generation of asset administration shell: a transfer learning approach with neural language model and semantic fingerprints. In: IEEE-ETFA Conference (2022)

Case Study on Localization for Robotic Wire Harness Installation

Markus Wnuk[1]([✉]), Manuel Zürn[1], Matthias Paukner[2], Sascha Ulbrich[2], Armin Lechler[1], and Alexander Verl[1]

[1] University of Stuttgart, 70174 Stuttgart, BW, Germany
markus.wnuk@isw.uni-stuttgart.de
[2] KUKA Systems GmbH, Augsburg, Germany

Abstract. Wire harness installation is one of the most challenging processing steps for automation in automotive production. Wire harnesses have an infinite number of degrees of freedom, such that they change their shape continuously during manipulation. As of today, the human's ability to perceive the wire harness and deal with its shape changes, is unmatched by any technical solution. Therefore, wire harnesses are still installed manually. This paper proposes a concept for wire harness localization and applies it for robotic wire harness installation. The concept uses two stereo cameras, one to perceive the shape of the wire harness and obtain rough position estimates of wire harness components, and a second for accurate 6D pose estimation of the individual components. The concept is evaluated with a case study on localization in a car chassis, where the accuracy and limitations of the concept are investigated by practical experiments.

Keywords: wire harness · localization · robot · installation

1 Introduction

Wire harness installation is a very challenging task. Especially in automotive production, where its size and complexity make it one of the heaviest and most expensive components [1]. Therefore, wire harnesses are mostly installed manually, which is a physically demanding task [2]. Compared to wire harness assembly, where several processes have been automated in the last decades, wire harness installation in the car remains to have a low level of automation [3]. However, automation offers the potential to increase process reliability, reduce physical load on workers, and decrease costs in the long term. Therefore, this paper aims in the direction of achieving higher automation in the process of wire harnesses installation in the car chassis, as shown in Fig. 1.

2 Problem Statement

In research, wire harnesses are referred to as semi-deformable linear objects (SDLOs) [4]. They consist of rigid components such as clips, plugs, wire channels,

Fig. 1. Installing a wire harness in a car chassis. The proposed localization concept is used to determine grasp poses for the robot.

and fixations interconnected by deformable parts [1]. Typical processing steps of wire harness installation include routing the wires, inserting the plugs into their mating parts, mounting clips, and connecting any wire ends to appropriate terminals [5]. Wire harness perception for the installation process can be split into:

(i) Finding the required component for an installation step: The challenge here lies in distinguishing the different components from each other. The components might have similar or identical geometry and may be used multiple times across a wire harness.
(ii) Determine the pose of a component: The challenge here is to determine the pose of the component accurately enough to be grasped by a robotic manipulator. The components along the wire harness are very small compared to the workspace in the car chassis. For vision systems, this poses the problem of covering a large field of view with a high resolution.

3 Related Works

Relevant works that address wire harness localization can be divided into two categories. Approaches from the first category, such as [7] and [8], aim at localizing the deformable parts of the wire harness. They use a series of segmentation and estimation methods to distinguish individual wires, even in entangled wire harness assemblies. However, these works consider aircraft wire harnesses and mainly address applications in aeronautics, such as wire defect detection. Therefore, the focus of these approaches lies on visual characteristics such as distances between wires, colors, or bending, while not considering the circumstances given in automotive wire harness installation. One of the most comprehensive works on the use case of automotive wire harness installation is presented in [9]. They address the problem of perception by attaching visual markers to the

wire harness and its components. This only yields a partial representation of the wire harness, at locations, where markers are attached and visible to the camera. Hence, there is no information about the mechanical coupling between the detected markers available. This is a limiting factor for certain installation steps.

Approaches from the second category neglect the deformable parts and only aim to localize individual components. In [4], the authors investigate the process of mating different electrical plugs of the wire harness. For localization, they use a global camera for detecting the component in the workspace using a deep learning neuronal network, trained on labeled images of the components. For accurate 6D pose estimation, they use a 2D in-hand camera system with template matching of 2D shape representatives obtained from 3D CAD data of the plugs. The approach is similar to the proposed concept in the sense that it uses two camera systems. However, it can only estimate components, but not the shape and configuration of the wire harness. Another approach from this category is presented by [10]. They use color heuristics to estimate the positions of plugs from 3D RGB-D data using traditional image processing methods. The focus of their method is position estimation and therefore, they do not obtain the 6D pose of the components. However, for certain automotive wire harness installation tasks, the orientation is crucial, especially when subsequent mating requires a specific orientation.

4   Contribution

Despite previous approaches have made great efforts to solve the challenge of localization, none of the investigated methods can fully fulfill the requirements for automotive wire harness installation. Therefore, this paper proposes a two-step localization concept. The concept estimates the current shape of the wire harness first. Hereby, we are capable of finding and distinguishing geometrical identical components and also estimate their mechanical coupling. In a second step, the pose of each component is obtained with high accuracy from a narrow field of view, by state-of-the-art template matching using the components' known geometric model. The presented concept is evaluated with a case study on clip installation of a reference wire harness with a seven-axis lightweight robot in car chassis.

5   Concept

The proposed concept uses two stereo cameras with different fields of view for the localization, see Fig. 2. The first camera is used to find the shape of the wire harness. Therefore, it is referred to as "**C**amera for **W**ire harness **L**ocalization" (CWL). The idea is to estimate a narrow region of interest for the second camera, which is referred to as "**C**amera for **C**omponent **L**ocalization" (CCL). The CCL is mounted on the robot end-effector to allow flexible repositioning. Eventually,
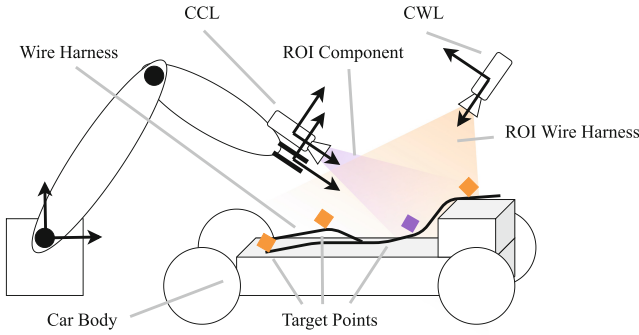
Fig. 2. Schematic of the test bench setup. Depicted is the concept of different fields of view of the CCL and CWL. The specific component of interest for accurate pose estimation is highlighted in purple.

the goal is to obtain an accurate estimation of the position and orientation of individual components, which are illustrated as target points in the figure.

The concept consists of two steps. First, the shape of the wire harness is determined. This allows for rough position estimations of all wire harness components. The estimated positions can then be used to distinguish individual components by comparing the estimated position of the component with their respective position on the wire harness. Note that this also allows estimating mechanical coupling between the components, which determines the available motions during a subsequent manipulation. In the second step, a specific component, chosen from a higher-level assembly logic, can be localized by moving the CCL over the estimated position of the component. This allows a high resolution pose estimation, as the camera operates in a narrow field of view of the component.

5.1   Wire Harness Localization

The CWL creates a point cloud of the entire workspace. The point cloud is preprocessed to segment the wire harness from the environment. Here we use 2D image processing, such as binary filters and background subtraction, together with 3D image processing, such as clustering and box filters for outlier removal.

After preprocessing, the localization is performed by registering a model of the wire harness to the obtained point cloud representation of the wire harness. For this step, we use non-rigid registration methods, such as [11] or [12], to estimate correspondences and match the model to the point cloud. This process is shown in Fig. 3a.

After registering the model to the point cloud, the rough positions of the components can be obtained, as each component is rigidly attached to the wire harness. Therefore, the local position of a component on the wire harness encodes its position uniquely in the workspace.
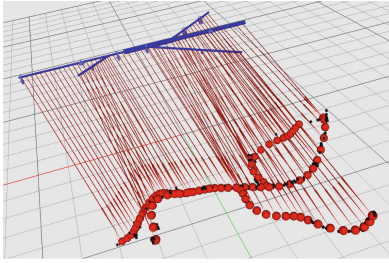
(a) Wire harness localization by non-rigid registration.
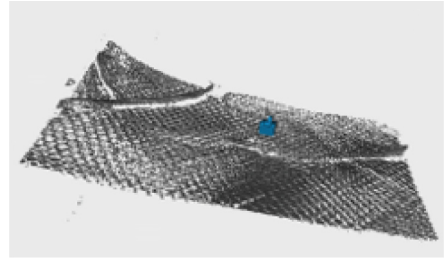
(b) Component localization for a clip by template matching.

**Fig. 3.** Estimation methods for the steps of wire harness and component localization.

5.2 Component Localization

Using the rough positions obtained from the estimated wire harness configuration, the robot can move the CCL, such that its field of view covers the targeted component. The CCL then acquires a high-resolution point cloud of the component. From this point cloud, we can obtain an accurate pose estimation using template matching [13]. Template-matching algorithms are often directly integrated into software libraries of modern stereo vision systems and can be readily used. They yield accurate estimations of the position and orientation given a CAD model of the component. For the system used in this paper, the matching result of a wire harness clip is shown in Fig. 3. From the estimated pose of the component, a suitable grasping pose for the robotic manipulator can be obtained by specifying the gripper geometry and desired grasping points.

6 Evaluation

6.1 Experimental Setup

For the experimental evaluation, a setup according to the concept of Fig. 2 is built. The scenario consists of a robot, two stereo vision camera systems, a car chassis, and an exemplary wire harness. The robot is a *KUKA iiwa* lightweight robot with 7 °C of freedom. This allows for high flexibility in the constrained workspace. The CCL is a *Roboception rc_visard 65 stereo camera*, which was hand-mounted on the robot end-effector. The CWL is a *Nerian Scarlet* stereo vision system, which is fixed in the workspace on the top of the car chassis. The camera setup is depicted in Fig. 4. A reference harness is used to evaluate the localization concept. Its design was created in consultation with automotive companies associated as partners within the research campus *ARENA2036*. The wire harness consists of one central trunk of 1 m in length and three additional branches of 0.44 m, 0.36 m, and 0.15 m, and a radius of 0.01 m. The wire harness is fixed in the car chassis with custom-made fastening clips, which are used as reference components in the evaluation. They are modified for easier grasping with a robotic gripper, as depicted in Fig. 4b.
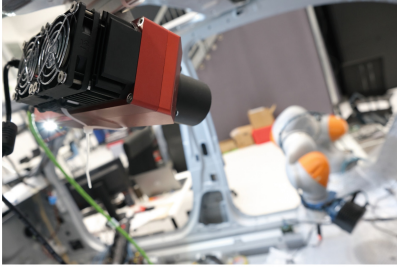
(a) Camera for wire harness localization.
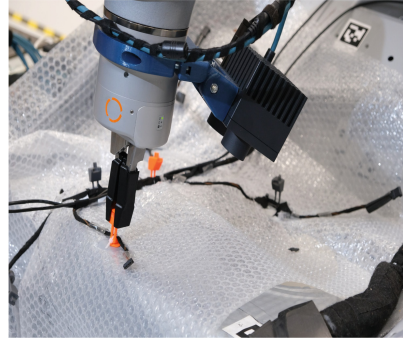
(b) Camera for component localization.

Fig. 4. Camera setup for the experimental evaluation of the localization concept.
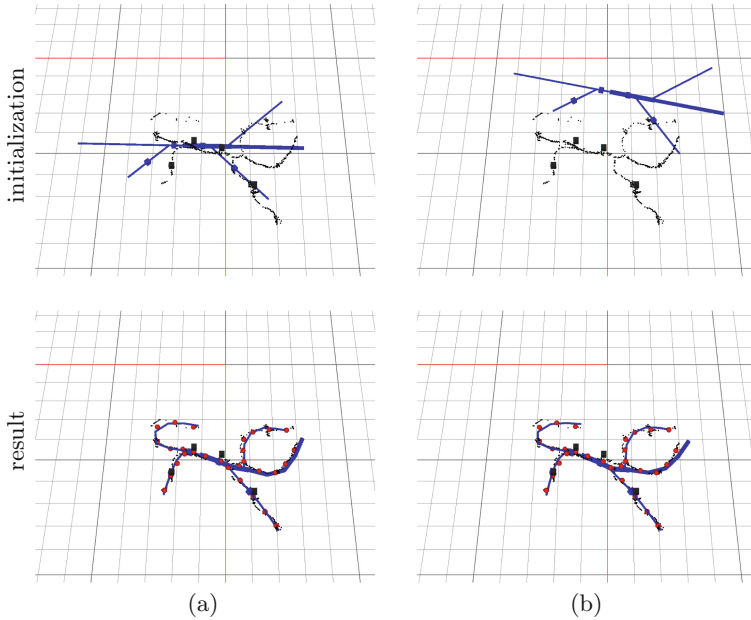
6.2 Evaluation of Wire Harness Localization

The evaluation of the wire harness localization aims at investigating if the configuration of the wire harness can be localized by the CWL with sufficient accuracy. Sufficient accuracy is measured by the subsequent localization with the CCL, as the rough clip position is encoded by the wire harness localization of the CWL.

For this purpose, a reference to measure and compare the performance of the localization is defined. This was achieved by fastening the wire harness with four clips in the car chassis. For each of the clips, successful grasping poses are determined. Successful grasping poses are used as a ground truth reference for further evaluation.

Assuming no prior knowledge of the ground truth clip positions, wire harness localization is performed. From the registered model, the positions of the four mounting clips are estimated. The estimated positions are then compared to the ground truth positions. Since it cannot be assumed that the initial configuration of the wire harness is accurately known, the model is initialized in different poses. The initial poses are randomly sampled with a positional uncertainty of $\pm 10\,\text{cm}$ in the mounting plane and a rotational uncertainty of $\pm 20°$ around the normal of the plane. The localization was performed for nine different configurations.

Exemplary results are shown in Fig. 5. The model is depicted in blue, the observed point cloud in black, the estimated configuration from non-rigid registration is shown in red, and the positions of the mounting clips are indicated by black boxes for the measured ground truth positions and blue for the model's estimates. To quantitatively assess the accuracy of wire harness localization, the error between the measured ground truth and the estimate is defined by the euclidean distance. To assess the accuracy of the wire harness localization, the error is averaged for each clip over the conducted experiments. Figure 6 shows the evaluation scenario with the enumeration of the clips depicted in Fig. 6a, and the measured mean position error for each clip given in Fig. 6b. The position error

Fig. 5. Exemplary tracking results for the evaluation of the wire harness localization with different initialization poses of the model.

remains below 5 cm for all experiments. This determines the maximum uncertainty which can be used to define a region of interest for the clip localization. CCL should be positioned such, that an area of ± 5 cm around the estimated clip position is covered, to ensure that the sought clip lies within its field of view.

For all randomly sampled initial configurations of the wire harness the clip error after registration lies below 5 cm, which is a sufficiently accurate estimate for targeting the field of view of the subsequent component localization towards the desired clip. The standard deviation over all experiments remains between ± 0.1 cm to ± 2 cm. This can be used to define a confidence measure for distinguishing different clips. Considering a radius of the standard deviation around every clip's estimated position, clips which lie outside the radius can be clearly distinguished and uniquely identified from their estimated positions alone. Within the radius, such a distinction cannot be made, and the clips need to be distinguished differently, for example by unique geometry, color, or texture information. However, in our experiment, all clips could be uniquely identified because they were mounted with a standard deviation greater than 2 cm apart.

6.3   Evaluation of Component Localization

The evaluation of the component localization aims to investigate if the individual components can be localized with sufficient accuracy for grasping. In the conducted case study the attached clips are considered the components and serve as
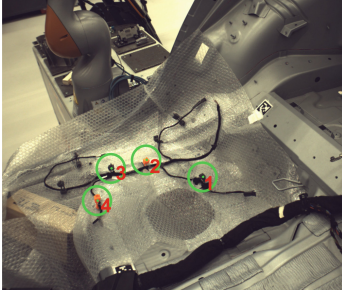
(a) Clip enumeration.
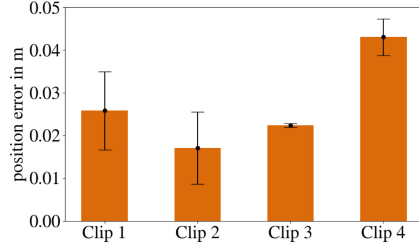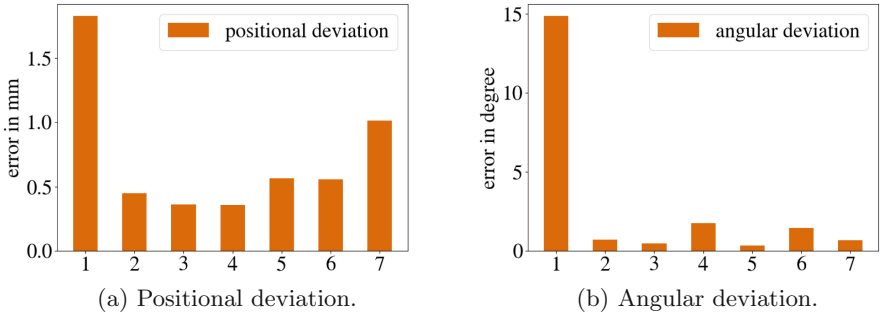
(b) Accuracy of wire harness localization.

Fig. 6. Evaluation of wire harness localization.

a reference for the evaluation. From the prior wire harness localization, a rough estimate of the position for each clip is available. Using this information, the robot is positioned such that the desired clip lies within the field of view of the CCL. However, many poses can be considered for positioning the CCL. Therefore, component localization is performed from different poses in the workspace.

The experiment is performed for a single wire harness clip, where its ground truth pose is determined, as in the previous experiment, by performing a successful grasp on the clip. The clip is localized by the CCL from seven different perspectives, each time rotating the robot end-effector for several degrees. For each pose, the positional and angular errors are determined from comparison with the ground truth. The positional error is measured by the euclidean distance between the estimated position and the ground truth. The angular error is measured as the norm of the Rodrigues angle between the measured orientation and the ground truth. Figure 7 shows the results.

From the results, it can be seen that the positional error remains below 1 mm and the angular error remains below 2 ° for six of the seven poses. The first pose yields a larger positional and angular error of up to, 1.82 mm and 15 °. From experience, we rate this measurement as an outlier. Therefore, the average positional accuracy is 0.46 mm, and the rotational accuracy is 0.94 °.

From the experiments, it can be observed that the measurement errors are in the range of 1 mm for the positional error and 2 ° for the angular error. This yields successful grasps, since errors within this range generally can be tolerated by the gripper design. However, outliers such as seen for the first pose, can occur and yield a probability to result in an unsuccessful grasping attempt. A practical approach to counteract such measurement errors might be self-centering clip and gripper geometries which provide positional and rotational tolerances during grasping.

(a) Positional deviation.

(b) Angular deviation.

Fig. 7. Evaluation of component localization for wire harness clips.

7 Conclusion

This paper presents a concept for wire harness localization. The process is split into two steps. This allows covering a large workspace while maintaining high precision for the pose estimation of individual components.

Within the experimental evaluation of the presented localization concept, clips could be estimated with an average error of 1 cm to 4.3 cm in the car chassis. This was sufficient for positioning the CCL to perform accurate localization with a positional accuracy of 0.46 mm, and a rotational accuracy of 0.94 ° in a narrow field of view. Grasping attempts with a positional accuracy of less than 1 mm and rotational accuracy of less than 2 ° allowed successful grasps in this case study. From the results, it can be expected that the concept can be generalized for the localization of other components such as plugs or wire channels.

Nonetheless, there remain limitations to the proposed method. First, the wire harness localization needs an entire view of the wire harness and a sufficiently good estimate of the initial configuration to perform successful registration. Second, overlapping of wires needs to be avoided because the non-rigid registration could converge in physically implausible local minima, which causes the localization to fail. Third, the components need to be constrained to specific orientations, as the robot cannot re-grasp and re-position the component arbitrarily.

Future research will be necessary to improve the concept. Methods that are capable to localize the wire harness from partial views and recover from false position estimates or failed grasps will make the process more robust and help to get closer to the goal of automated wire harness installation.

References

1. Trommnau, J., Kühnle, J., Siegert, J., Inderka, R., Bauernhansl, T.: Overview of the state of the art in the production process of automotive wire harnesses. Current research and future trends. In: Procedia CIRP (2019)
2. Enríquez Díaz, J., Weichel, J., Frieling, E.: Analyse körperlicher Belastung beim Einbau des Kabelbaums in das Fahrzeug und Empfehlung zur Belastungsreduktion - eine Felduntersuchung in einem Werk der deutschen Automobilindustrie. Zeitschrift Für Arbeitswissenschaft (2012)
3. Nguyen, H., Kuhn, M., Franke, J.: Manufacturing automation for automotive wiring harnesses. In: Procedia CIRP (2021)
4. Zhou, H., Li, S., Lu, Q., Qian, J.: A practical solution to deformable linear object manipulation: a case study on cable harness connection. In: 5th International Conference on Advanced Robotics and Mechatronics (ICARM) (2020)
5. Aguirre, E., Ferriere, L., Raucent, B.: robotic assembly of wire harnesses: economic and technical justification. J. Manuf. Syst. **6**(3), 220–231 (1997)
6. Abegg, F., Henrich, D., Wörn, H.: Manipulating deformable linear objects: Vision-based recognition of contact state transitions. Comput. Sci. (1999)
7. Abdallah, H., Orteu, J.-J., Jovancevic, I. & Dolives, B. Three-dimensional point cloud analysis for automatic inspection of complex aeronautical mechanical assemblies. J. Electron. Imaging (2020)
8. Guo, J., Zhang, J., Gai, Y., Wu, D., Chen, K.: Visual recognition method for deformable wires in aircrafts assembly based on sequential segmentation and probabilisitic estimation. In: IEEE 6th Information Technology and Mechatronics Engineering Conference (ITOEC) (2022)
9. Jiang, X., Koo, K., Kikuchi, K., Konno, A., Uchiyama, M.: Robotized assembly of a wire harness in car production line. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (2010)
10. Yumbla, F., Abeyabas, M., Luong, T., Yi, J.-S., Moon, H.: Preliminary connector recognition system based on image processing for wire harness assembly tasks. In: 20th International Conference on Control, Automation and Systems (ICCAS) (2020)
11. Tang, T., Tomizuka, M.: Track deformable objects from point clouds with structure preserved registration. Int. J. Robot. Res. (2019)
12. Wnuk, M., Hinze, C., Zürn, M., Pan, Q., Lechler, A., Verl, A.: Tracking branched deformable linear objects with structure preserved registration by branch-wise probability modification. In: 27th International Conference on Mechatronics and Machine Vision in Practice (M2VIP) (2021)
13. Rink, C., Marton, Z.-C., Seth, D., Bodenmüller, T., Suppa, M.: Feature based particle filter registration of 3d surface models and its application in robotics. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2013)

Traceability in Battery Production: Cell-Specific Marker-Free Identification of Electrode Segments

Günther Riexinger[1]([✉]) [iD], David J. Regina[2], Christoph Haar[1],
Tobias Schmid-Schirling[2], Inga Landwehr[1], Michael Seib[2], Jonas Lips[1],
Simon Fehrenbach[2], Julian Stübing[1], Daniel Carl[2], and Alexander Sauer[1,3]

[1] Fraunhofer Institute for Manufacturing Engineering and Automation IPA, Nobelstr. 12, 70569 Stuttgart, Germany
Guenther.Riexinger@ipa.fraunhofer.de
[2] Fraunhofer Institute for Physical Measurement Techniques IPM, Georges-Köhler-Allee 301, 79110 Freiburg im Breisgau, Germany
[3] Institute for Energy Efficiency in Production, University of Stuttgart, Nobelstr, 12, 70569 Stuttgart, Germany

Abstract. Digitalization in battery production, as well as the increase and stabilization of product quality of lithium-ion battery cells, require the elimination of information gaps between processes to enable the traceability of components and process steps to the finished product. In lithium-ion battery cell manufacturing, using a traceability system is considered a promising approach to reduce scrap rates and enable more efficient production. Today, traceability is possible from the assembled cell onwards. However, with a view to the new EU battery regulation, complete traceability down to the material needs to be ensured. One of the challenges in this context is to ensure both, traceability in continuous electrode production and cell-specific identification within the production chain. This paper presents an approach for identifying individual electrode segments without identification markers, using the individual microstructure of the electrode surface. The Fraunhofer Institute for Manufacturing Engineering and Automation IPA and the Fraunhofer Institute for Physical Measurement Techniques IPM are further developing and testing an identification technique known as Track & Trace Fingerprint. This technique is dedicated to serialization within battery production as part of the joint project DigiBattPro 4.0 and to be implemented at the Center for Digitalized Battery Cell Manufacturing (ZDB) of Fraunhofer IPA.
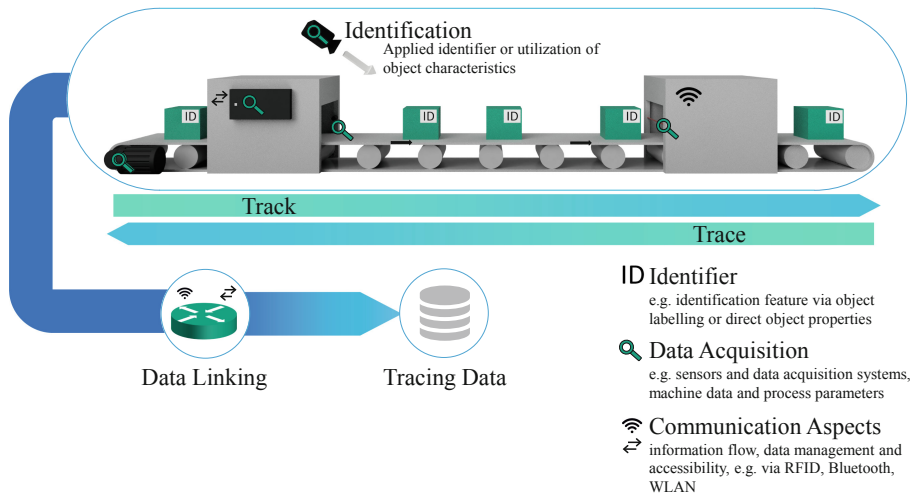
Keywords: Traceability · Identification · Battery Production · Digitalization · DigiBattPro

1 Introduction

1.1 Traceability in Lithium-ion Battery Production

Traceability not only plays an important role in production but also over the entire life cycle of a battery cell. Thus, the EU Commission anchors transparency along the entire supply and value chain in its proposal for a regulation concerning batteries and waste batteries. In addition to the CO_2 footprint during production, information about the origin and composition of the materials, repair or reuse, and end-of-life options such as recovery and recycling processes should also be provided [1].

In lithium-ion battery production, using a traceability system is considered a promising approach to reduce scrap rates, lower costs, and enable more efficient production [2]. It allows identifying possible problems or defects at an early stage of production. A traceability system collects information from trace objects, e.g. a single part or a segment of a continuous product during different phases of the product life cycle or at different production steps. It enables the assignment of process information based on tracing data for these objects. Generally, such systems consist of core elements, such as identification, data acquisition, data linking, and communication, as shown in Fig. 1 [3].

Fig. 1. Traceability system overview, adopted from [4].

In industrial battery production, an approach for cell-specific identification of electrode segments has not yet been established. Serialization becomes almost impossible when splitting the coil into subcoils or cutting out additional segments. Initial cell-specific serialization usually takes place after electrode production during cell assembly. Identifying cell-specific electrode segments within the electrode production process is currently one of the biggest challenges in cell production of lithium-ion batteries. The reasons can be summarized as follows:

- Change of batch structure (Raw materials → electrode paste → electrode coating → electrode sheets → cell assembly) [5].

- Manual removal of electrode segments due to product defects or to perform quality control.
- Prevention of material modifications or entry of any contamination due to the application of markings (e.g. ink).

1.2 Identification Techniques

Various techniques were investigated to select suitable identification solutions for lithium-ion battery production. Basically, trace objects are distinguished by their indirect and direct object features.

In the case of identification through indirect object features, the identification data is placed either on the workpiece itself or on workpiece carriers. The applied identifier is detected either contactless or tactile [6]. Optical markings belong to the group of contactless identification techniques. For this purpose, plain text, symbols, or an n-dimensional code such as a 1D bar code, a 2D Data Matrix code (DMC) or a color code is applied to the trace object and detected optically via digital image processing. Further options include electromagnetic or radio-based approaches. Trace objects could be equipped with active or passive radio modules (RFID, Bluetooth, WLAN modules), which transmit their identity signals. Tactile techniques require direct contact between identifier and sensor. For example, this includes devices that read the integrated circuits of smart cards.

Identification by direct object features eliminates the need for new identifiers to be applied to the trace object. Thus, a trace object is identified by its inherent characteristics, or the identity is calculated using logical models. For example, with some object materials it is possible to identify a trace object by its surface characteristics [7, 8] or to calculate e.g., the localization of a part or object segment based on process parameters via the use of physical-logical relations of the production sequence [9].
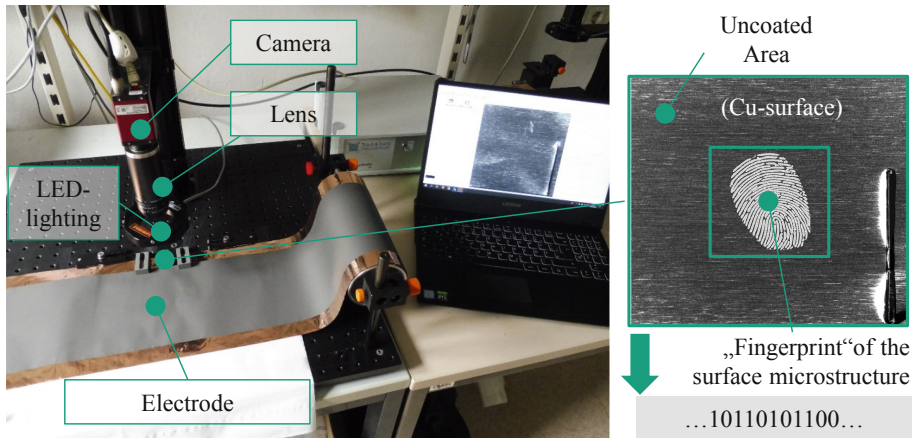
2 Cell-Specific Identification of Electrode Segments

Based on the described identification techniques and the requirements for electrode production, all labeling options have restrictions or limitations. Printed identifiers such as 1D barcodes or 2D Data Matrix code transmit printing ink [10], which may influence the battery's performance. Laser-marked identifiers may cause material lesions due to the laser process, and mathematical models have gaps once a part of the electrode is removed. Therefore, a marker-free technique is currently being researched to enable cell-specific traceability.

2.1 Marker-Free Identification Using Surface Microstructures

In the joint project DigiBattPro 4.0, scientists from Fraunhofer IPA and Fraunhofer IPM are working on marker-free identification via existing surface microstructures or direct object features in the area of individual electrode segments.

A camera-based system was designed to uniquely identify individual components without labeling, using only the individual microstructure of the surface. Laboratory tests

Fig. 2. Laboratory setup for the evaluation of electrode surface microstructures

showed that aluminum and copper electrodes are suitable for marker-free identification, as shown in Fig. 2.

The marker-free identification system is based on the Track & Trace Fingerprint Technology developed by Fraunhofer IPM. It is a tracking system for mass-produced parts, enabling individual serialization and authentication. The technology is driven by the fact that many semi-finished goods or components have an individually distinct microscopic surface or interwoven color structure. An industrial camera takes high-resolution images of selected areas on the component's surface. The specific structural pattern captured by the image and how they are positioned relative to each other is used to generate a numerical identification code, the so-called fingerprint. It is stored in a database and linked with other process and measurement data via a unique part ID. This process can be repeated to identify the component at a later stage in production.

Figure 3 shows an exemplary histogram on the fingerprint recognition of searched components or parts by comparing all fingerprints in a database. The x-axis shows the normalized Hamming distance, a measure of the relative similarity of two compared fingerprints. The smaller the value determined, the more similar the fingerprints are. The y-axis shows the number of fingerprints with the same degree of similarity. A typical successful comparison is shown in Fig. 3, where the sought-after part is sufficiently well separated from all other parts to make a statistically safe and unique identification. The distance between the center of the distribution of all other parts and the searched part in units of the standard deviation of the spread σ is given as a measure of the recognition reliability. When a comparison falls below a certain threshold of these measures, it is regarded as statistically unsafe and labeled as such. Thus, misidentifications are avoided. The value of these thresholds depends on the amount of fingerprints to be compared. A threshold value of 7 σ is sufficient to identify a single fingerprint out of a seven-digit number of fingerprints with statistical certainty.

Up to now, the technology has been implemented in industrial production lines to track single objects [11, 12]. However, the existing identification solution for discrete production objects or pieces does not meet the requirements for continuous electrode
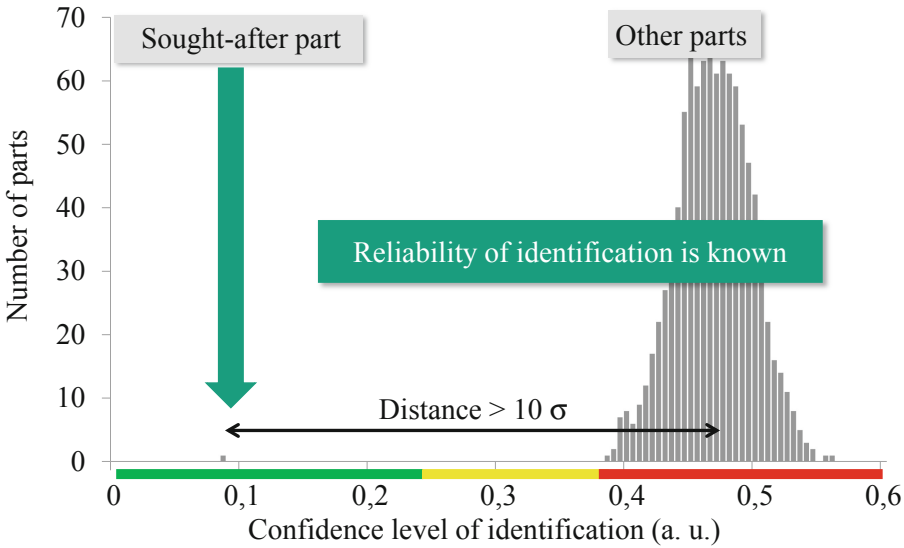
Fig. 3. Exemplary histogram with statistical reliability of fingerprint signature recognition. The colors on the x-axis indicate if the sought-after part is identified with statistical certainty (green), identified but with insufficient confidence (yellow) or not identified at all (red).

production. Therefore, further development of the fingerprint system is necessary to close the existing information gaps between the individual processes of electrode production and increase the granularity of serialization (from coil to cell-specific tracing of electrode segments), as well as to prevent material entry into battery cells by marker-free tracing. In continuous material production, a challenge for marker-free identification is the enormous amount of image data that must be dealt with when registering an object for the first time. Fingerprints have to be recorded and stored so that identification is possible, even when taking the image anywhere on the electrode. The developed technology can generate fingerprints within a few hundreds of milliseconds.

To identify cell-specific electrode segments, two cases have to be distinguished. In the first case, identification takes place prior to cutting the electrode. Here, identification can be enhanced by consecutive sequencing of individual cell-specific electrode segments. Furthermore, the system needs to identify if individual segments have been cut out, e.g., due to quality aspects. In the second case, when cutting is already done, it may occur that the order of the single electrodes is lost. As a result, the number of possible candidates is much larger than in the first case. The proposed solution is to compare and process the captured fingerprints on high-performance graphics processing units (GPUs), as this process can be largely parallelized. With recent improvements in GPU performance, processing such data volumes becomes feasible. Thus, the required detection time for identifying fingerprints depends on the hardware system and software configuration. The detection time will be validated and optimized within the planned operation of the fingerprint demonstrator system, enabling the traceability of materials used in continuous processes such as coils or foils.

2.2 Approach for Proving Concept Implementation

The following chapter describes current developments and the implementation of a traceability system for marker-free identification in continuous electrode production. To develop and evaluate the fingerprint demonstrator system, it is integrated into a roll-to-roll coating system operating at conveyor speeds between 0.5 and 20 m/min (see Fig. 4). Substrate foils and electrodes up to 200 mm in width can be handled and treated with infrared radiators and inline calender. The equipment enables the development and testing according to current specifications on lithium-ion production processes defined in advance within the project.
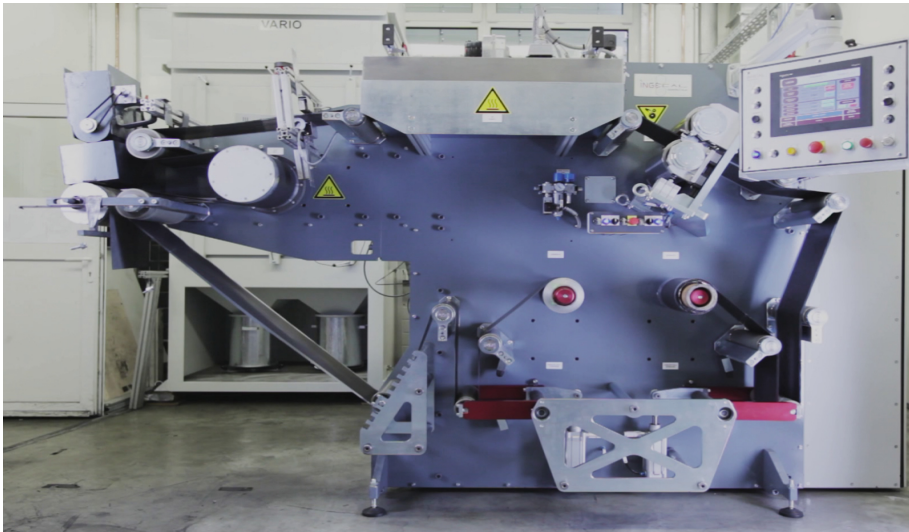
Fig. 4. Roll-to-roll coating system with inline calender as basis for the fingerprint demonstrator.

The implementation of the Track & Trace Fingerprint technology consists of at least two reader systems: One reader system scans and serializes the electrode and adds fingerprints to the database (to make them known to the system). The second reader system captures the fingerprints (tracing them in the database) and identifies the electrode segments. Consequently, a fingerprint reader system was developed (see Fig. 5) and two identical systems were set up from industrial computer vision components (see one of them in Fig. 6).

A reader system includes four high-resolution cameras with corresponding lenses, LED bars as illumination devices, and an IT interface to the central fingerprint management software. Stiff mechanical mounts were designed for capturing images without vibrations. Four cameras are needed to track four parallel tracks of electrodes. The sampling of the optical system is limited by the pixel size and amounts to 26,2 μm per pixel. The measuring field of a single image is 70 mm x 21 mm, while operating distance is about 230 mm. The actual area used for fingerprint generation is even smaller at 16 mm x 1 mm, resulting in a memory footprint of less than 1 KB per fingerprint. In the case

of adding and tracing discrete parts, a control system such as a programmable logic controller (PLC) sends add or trace commands to the fingerprint system.
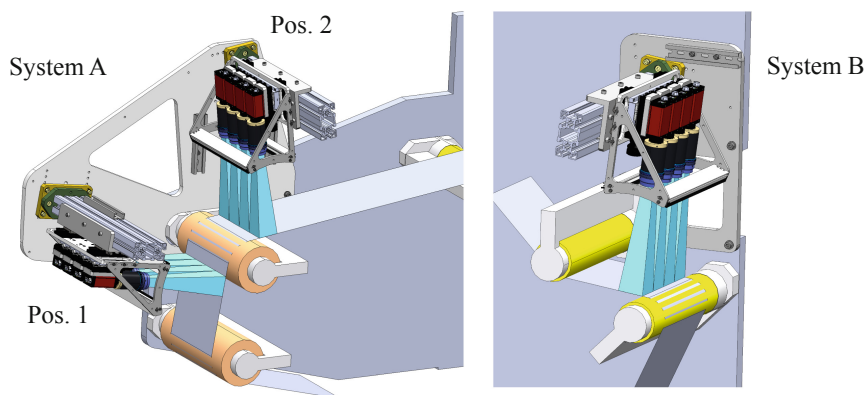
Fig. 5. Design of the two fingerprint reading systems in CAD, where System A has two possible mounting positions (Pos. 1 and Pos. 2).

When applying in combination with endless material, the reader system itself has to keep track of the coil position and add fingerprints at a fixed distance from each other automatically, independent of the electrode segment length. This is done by image processing of the detected surface microstructure and necessary to ensure the generation of at least one fingerprint in the captured tracing area. The distance between the individual fingerprints varies depending on the camera setup and field of view, fingerprint dimensions, as well as processing times.

A coil with 800 m length requires around 30 million fingerprint comparisons for a single request. As the comparisons can be computed in a highly parallel manner, a GPU acceleration is promising to speed up the process to the required identification rate. Early tests on a low-end mobile GPU already show a doubling in performance compared to the previous system. The objective is to identify cell-specific electrode segments in production cycle time.

A first test setup of the developed system including the lighting system was validated and will be finally implemented on a coating machine. In a first step, the uncoated area of the electrode segment was successfully serialized, identified, and recognized by the developed Track & Trace Fingerprint-based system. In general, all technical surfaces that exhibit an individual microstructure are suitable for the technology. Based on this, further tests will be carried out on different electrode materials together with different coatings (e.g., Cu and Al substrates and graphite coating). In the current state, the system can support up to a 10 m/min feed rate but is projected to reach 20 m/min or more with further optimization. For further use of the fingerprint serialization within subsequent production processes, it is necessary to transfer the fingerprint information from continuous electrode production to discrete cell manufacturing and housing. From the point of cell assembly, this is currently realized by conventional tracking codes such as DMC.

Fig. 6. Prototype setup of the demonstrator on a coating machine.

3   Conclusion and Outlook

Traceability of individual lithium-ion battery cells within electrode production is challenging, since there is currently no reliable way of implementing continuous traceability at single-cell level, thus linking electrode production and process data on a cell-specific basis. To meet this challenge and eliminate the existing identification and information gaps between process clusters, this paper established and presented a Track & Trace Fingerprint-based approach for continuous single-cell identification. Based on the newly designed Track & Trace Fingerprint system, a prototype setup was integrated and successfully implemented on a coating machine at the Center for Digitalized Battery Cell Manufacturing (ZDB) of Fraunhofer IPA. The hardware and software system will be further adapted and optimized in the next development steps to optimize the marker-free identification of continuous material. After testing precoated substrates or finished electrode rolls on the current coating and calendaring machine, a further expansion of the Track & Trace Fingerprint system is planned. The integration into additional coating lines allows evaluation of further aspects, such as modified material systems, residual moisture in the coating, and other influences accompanying the wet coating process.

The new marker-free identification approach enables the tracing of electrode segments without material changes and contamination due to the application of printed or lasered identifiers. Even if electrode segments are cut out, it will be possible to allocate electrode process parameters and data obtained during the coating process to the individual battery at cell level. This will provide a more in-depth understanding of the battery production process, a rapid localization of defects in the production chain, and

consequently will improve product quality. Quality improvements and determining influential process factors will become possible by a more granular analysis. Besides, new high-quality data sets allow for data-driven production analysis and optimization. With the developed Track & Trace Fingerprint system, cell-specific traceability of lithium-ion battery components and process steps to the finished product becomes possible. In the future, the technology can be used for other sectors and continuous manufacturing processes such as the production of continuous materials.

References

1. European Commission: Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL concerning batteries and waste batteries, repealing Directive 2006/66/EC and amending Regulation (EU) No 2019/1020 Document 52020PC0798. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0798. Accessed 30 May 2022

2. Michaelis, S., Rahimzei, E., et al.: Roadmap Batterieproduktionsmittel 2030 – Update 2020, pp. 126–130. VDMA Verlag GmbH (2020)

3. Wegner-Hambloch, S., van Betteray, K., Springob, K. (eds): Rückverfolgbarkeit in der Praxis: Artikel 18 und 19 der VO (EG) Nr. 178/2002 schnell und einfach umgesetzt, 1. überarb. Nachdr. Behr, Hamburg (2004)

4. Birke, K.P., Weeber, M., Oberle, M.: Handbook on Smart Battery Cell Manufacturing. World Scientific Publishing, Singapore (2022)

5. Heimes, H.H., Kampker, A., Lienemann, C. et al.: Produktionsprozess einer Lithium-Ionenbatteriezelle, 3rd edn., Aachen, Frankfurt am Main (2018)

6. Riexinger, G., Doppler, J.P., Haar, C. et al.: Integration of traceability systems in battery production. Procedia CIRP **93** 125–130 (2020)

7. Hofmann, A.: Track & Trace per "Fingerabdruck" (2017). https://www.ipm.fraunhofer.de/content/dam/ipm/de/PDFs/produktblaetter/PK/IVS/Track-trace-FINGERPRINT-de.pdf. Accessed 04 May 2022

8. Schmid-Schirling, T., Förste, A.: Track4Quality - Markierungsfreie Einzelteilverfolgung in anspruchsvoller Produktionsumgebung zur Prozessverbesserung und Qualitätssicherung: Schlussbericht zum Vorhaben, Fraunhofer-Institut für Physikalische Messtechnik IPM (2022)

9. Appelhanz, S., Schumann, M.: Improvement of traceability in continuous processes. In: Geldemann, Jutta, Schumann et al. (eds.) First International Conference on Resource Efficiency in Interorganizational Networks - ResEff 2013 -. Universitätsverlag Göttingen, Göttingen, pp. 192–204 (2013)

10. Sommer, A., Leeb, M., Haghi, S. et al.: Marking of electrode sheets in the production of lithium-ion cells as an enabler for tracking and tracing. Procedia CIRP **104**, 1011–1016 (2021)

11. Schmid-Schirling, T., Körber, T.: Bauteile per Fingerabdruck im Sekundentakt zurückverfolgen. In: Journal für Oberflächentechnik, pp. 40–42. Springer Fachmedien Wiesbaden GmbH (2022)

12. Saum, N., Schmid-Schirling, T.: Rückverfolgung von Bauteilen mit Track & Trace Fingerprint. In: Digital Manufacturing Magazin, pp. 42–43. WIN-Verlag GmbH & Co. KG, Vaterstetten (2021)

Multi-agent Interaction Structure for Enabling Subsidiary Planning and Control in Modular Production Systems

Simon Komesker[1,2(✉)], Jonathan Bartels[1,2], Achim Wagner[3],
and Martin Ruskowski[1,4]

[1] TU Kaiserslautern, 67663 Kaiserslautern, Germany
`simonkomesker@gmx.de`
[2] Volkswagen AG, 30884 Wolfsburg, Germany
[3] Deutsches Forschungszentrum Für Künstliche Intelligenz GmbH, 67663 Kaiserslautern, Germany
[4] Technologie-Initiative SmartFactory KL E.V, 67663 Kaiserslautern, Germany

Abstract. Modular production systems enable resilient production processes through decoupled production processes. On the way to implementing flexible and adaptable production systems, information support plays a decisive role. Only the use of intelligent and structured information processing across previous system boundaries and areas enables the coordination of requirements and capacities in dynamic production environments. The rigid communication structures in information systems of current production systems therefore need to be replaced by dynamic interaction, both horizontally between entities and vertically between different hierarchical levels. Multi-agent systems (MAS) are one way to meet the requirements for centralized and decentralized decision making in complex (cyber physical production) systems (CPPS). To prepare the instantiation of a MAS, it is necessary to structure and describe the information flows of a production system.

In this paper, the results of a simulation experiment for the implementation of collaborative, subsidiary decision making based on a model-based system structure are presented. Productivity potentials of more than 10% can be shown by using collaborative manufacturing strategies.

Keywords: modular production · multi-agent systems · interaction model · systems architecture

1 Introduction

Resilient production systems guarantee robust production processes in the event of unforeseen deviations in the operating sequence. Kern's Modular Production proves to be more resilient to demand and capacity fluctuations compared to linear production systems [1, 2]. With the use of Cyber Physical Systems (CPS), mechanical and mechatronic elements in (socio-technical) production systems acquire a higher decentralized decision-making capability and possess complex interaction [3]. Systems engineering

principles enable a better understanding and designing of complex phenomena [4]. Thus, a system structure is needed that enables structured information processing in modular production systems [2].

The requirements for information processing in modular production systems with alternative process design according to Kern have been identified in former research [2, 5]. This paper follows up on these and validates the previous research in the form of a simulation experiment. For this purpose, the related work of preparing structured information processing in multi-agent systems (MAS) for matrix production like Kern's Modular Production is given in Sect. 2, as well as the basics of systems engineering. In Sect. 3, the basic functions for operating a MAS in modular production systems are identified and put together in an interaction model. This model is supported with a system structure in Sect. 4. Section 5 evaluates the implementation before the paper is concluded in Sect. 6.

2   State of the Art and Related Work

In the following section, the related work in the area of systems structures for cross-system information processing is presented.

2.1   Structured Information Processing in Modular Production

Modular Production Systems require cross-system information processing [4]. The control paradigm of strictly hierarchical control, e.g. the ISA95 is not suited for dynamic and flexible cross-system information processing [6]. Alternative control paradigms that allow vertical and horizontal communication are hybrid control paradigms [7]. A promising approach to enable hierarchically but flexible information processing is the application of multi-agent systems, that allow for complex communication between encapsulated agents as subsidiary [8, 9]. As previous research stated, a special focus lays upon the systems architecture and structure to support the requirements for information processing between different domains [2, 5]. There are existing architectures that serve as orientation for MAS development [10, 11]. An approach preferably suited for reconfigurable MAS are AOSE methods like ADMARMS which is based on high level design principles and a rather functional oriented development [10].

The coordination of systems is based on interaction of product, processes and resources in CP(P)S [12]. Different possibilities exist to apply coordination by planning and scheduling [13]. These could be used to coordinate systems and its inner systems and elements on different hierarchy levels.

2.2   Structuring Systems for Developing a Systems Design

A System can be seen as a combination of interacting elements organized to achieve one or more stated purposes [4]. Systems Engineering is focused on the system as a whole, it looks at the system from the outside as well as from the inside using different principles and concepts [14]. The functional concept describes the functions of a system and what a system does. The structural concept describes the interior of the system, as

well as the relationships of elements within this system. Elements of a system can also represent systems on a lower level (system of systems) [15]. This nested relation can be described by the hierarchical concept [4, 16]. A subdomain is the systems design, that incorporates the architectural, logical and physical setup. By using a model-based architecting approach a suitable support of informational processes can be created [17, 18].

In addition technical cybernetics analyze the information flow through the system and how it is processed. It considers how this can be used to manage and control itself as a control loop [19].

3   Modelling Interaction for Planning and Control in Modular Production Systems

An Industrie 4.0 compliant systems structure for modular production systems needs to support complex decision making [5].

The procedure for decomposing a system top-down follows a basic principle "From the General to the Detail" [16]. A decomposition is carried out to managerial and operational independence of functions to create a functional architecture [18]. The coordination of functions follows the bottom-up principle and couples different functional blocks to process (steps). As part of the systems analysis, the system's elements are identified as well as the tasks, roles and interdependencies for a system design view [5]. The application of ADMARMS design methodology supports a MAS architecture with strong focus on maintaining the independence of functional requirements [10].

3.1   Top-Down Decomposition and Bottom-Up Aggregation of a Modular Production System

The system is basically designed as a fractal structure with similar subsystems on different hierarchy layers. The subsystems are designed as encapsulated entities, which can be supported by the concept of AAS and ASD [5] (see Fig. 1).
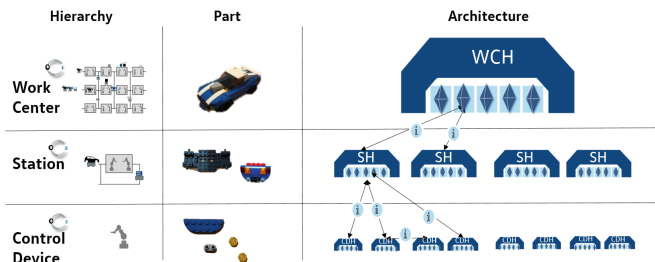

Fig. 1. Decomposition of the production system for structured information processing for a holonic architecture

The structuring was done by defining hierarchy layers for the Modular Production by Kern according to the RAMI4.0 levels followed by a functional requirements engineering matching the manufacturing system design. Each system contains similar functional blocks that are encapsulated as holons on different hierarchy levels according to the RAMI4.0 and preparing a subsidiary decision making.

The challenge of modular production systems was identified in the information processing for a resilient production flow in skill-based modular production systems. The functions and basic elements of a system from an information point of view were extracted based on requirements on previous research [2, 5]. They are separated as representation and self-description functions on the one hand and coordination functions on the other:

I Representation and Self-description Functions

The **Resource Agent (RA)** knows capabilities and availability status of an element or system. It includes skillset, process times and setup matrix. The capabilities or skills are aggregated bottom up from each level to fulfil orders and suborders. The **Product Agent (ProdA)** gives an overview of all manufacturing steps for a specific product in a subsystem. It defines configuration, quality requirements and start and end dates for each item in the production program. It contains information about the order and the manufactured product, with update information from the resources for completed process steps. This interacts with the shell that contains order specific blank options for the product being built. The material is available in sufficient quantity in the supermarket and the transportation is coordinated with the production flow. The production flow manager need the information of material availability for the planned production program. The **Quality Agent (QA)** evaluates quality data from products and parts/material. The **Process Agent (ProcA)** describes skills of the production processes needed to manufacture products with suitable resources.

II Coordination Tasks

The **Production Flow Manager (PFM)** ensures, that the right processes are performed on the right resources in the right time. Therefore the PFM needs information about the product process steps, the precedence graph and the offers from the resources for processing orders and suborders. The PFM schedules and reschedules the orders on the different resources. The **Production System Manager (PSM)** ensures that the right resources are available to produce the production program. A potential measure could be the reconfiguration of a system e.g. a resource adaption or integration of unplanned orders, because the (sub-)system matches the required skills for that order. KPIs to consider in this context are transportation time, variant flexibility, value add time, setup time, makespan, output. The **Material Agent (MA)** ensures, that the right material is at the right place at the right time and the right volume The **Data Manager (DA)** collects all the actual and requested information in the system and provides a consistent data base. The **Deviation Agent (DevA)** identifies and assesses deviations initiates activities in the system. The mechanism for planning and scheduling is presented in [5] and defines a subsidiary decision making process for multi-level modular production systems. The **Orchestration Agent (OA)** is responsible for the execution of decisions and closing the control loop of an integrated planning and control system.

3.2   Interaction Model for Cross-System and Cross-Level Information Processing

Based on the identified agent functionalities, an interaction model was developed as basis for a functional architecture in the sense of a model-based architecture [15]. The formalized interaction model derived from the functional description in Sect. 3.1 is depicted in Fig. 2. The interaction model is used for every fractal system of the production system to support its subsidiary decision making. A fractal is defined per hierarchy level as a work center, a station or a resource, each represented as a holon (see Fig. 1). The interaction is described on an abstract level as control/command and inform. Control/command interactions need a confirmation of receive, an inform a typical message into an agents inbox. Asynchronous horizontal and vertical communication is realized by the DA that serves as a message broker between agents, structuring data in different topics to which other agents are subscribed and therefore enabling cascaded feedback control loops.
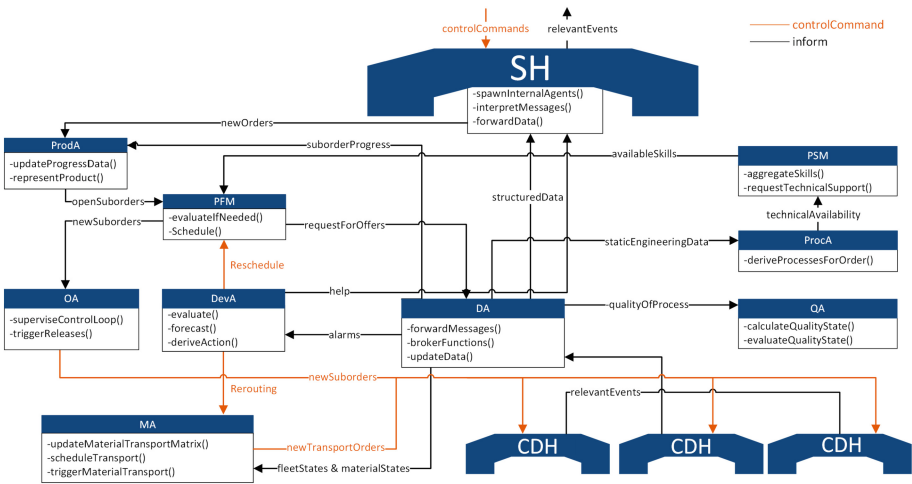
Fig. 2. Interaction model for multi-level and cross-system holistic planning and control

4   Implementation of a Subsidiary Planning and Control for Enabling a Resilient Production Flow in Modular Production

In this section a system structure for a use case is implemented based on the interaction model for subsidiary decision making. This is done by using the functional decomposed blocks and aggregate it bottom-up in a multi-level production system.

4.1   Introduction of the Use Case

The production system contains 84 resources to produce 3 different models of a car with different but similar product structure. The system is structured hierarchically in 3 levels according to RAMI 4.0 consisting of fractals designed as holons with the same agent

architecture. The **work center level**, that schedules and orchestrates the production of the product. The **station level**, that schedules and orchestrated the production of modules of the product in stations. And **the control device level**, where the processes to manufacture parts for modules are being executed the scheduled sequence. The control device level marks the lowest level of granularity of functions for operational and managerial independence. The production program consists of 300 products, each of which consists of 16 modules and the associated production orders, 3 of which can be manufactured in parallel. The module orders contain 2 to 8 production orders, so that the total number of process steps for a product amounts to up to 90 steps. A data and information model is supporting the function of a skill-based coordination and allocation of the different product needs and resource capabilities with the respective process skills.

4.2   Implementation of a Model-Based System Structure

To create the system structure, the interaction model is prototyped as a holonic MAS. The intelligent medium- and long-term decision-making is realized with the help of agent-based modeling and simulation (ABMS). The coupling with a production system is realized using the industrial-grade Discrete Event Simulation (DES) tool Plant Simulation, which is coupled via a middleware with the ABMS and a scheduling module.
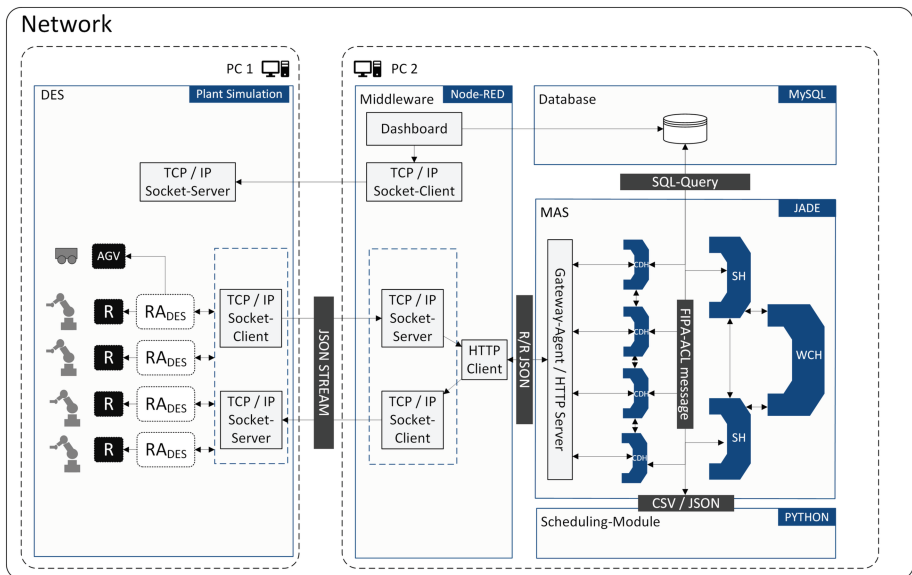
Fig. 3. Implemented system architecture from of the interaction model and planning mechanism

The **DES** provides the data basis for the decisions as a simulated production system. Based on the skill-based approach, the resources have the decomposed skills to execute process steps to manufacture a part, module or product. The formerly passive entities of the DES were agentified in order to be able to communicate horizontally and decide

decentrally for ad-hoc initiation of transport orders and execution of short-term alternative strategies. The material transport is modeled by freely moving AGV within the station and by lane-bound AGV between the stations.

By using a low-code **Middleware** and TCP/IP based interfaces a synchronous communication between the DES and the MAS can be established. The goal is to connect each passive resource of the DES with an active equivalent in the MAS. The middleware forwards the event-based JSON-formatted updates from the DES to an http-enabled gateway agent of the MAS via the TCP/IP client/server socket interface and returns derived actions, as shown in Fig. 3. Via a web-based user interface, the middleware enables configuration of the MAS, scheduling module and DES.

The **MAS** is the digital representation of the production system and structures the agents within it based on the interaction from Fig. 2. The MAS is a multi-agent system. In implementing the interaction model, the focus was on cross-level processing and functions were aggregated. Based on the FIPA-ACL compliant agent platform JADE, a MAS was developed that implements the sequencing and allocation planning [4]. For this purpose, each control-device level resource was implemented as a holonic CDH in a 1:1 relationship in the MAS. In the prototypical case, each Holon encapsulates several of the agents presented in Sect. 3, within the respective system as a black box. These include the CDH representation function and the DM's function, which interacts with an SQL database. In addition, the function of checking the available skills of resources of the PSM is carried out during initialization and failure events. In the case of level-specific sequence and allocation scheduling, a scheduling module is triggered in each case, corresponding to the function of the PFM. Through the interaction of individual encapsulated holons as well as the cross-level decomposition of orders, the holons solve the problem of sequence and allocation scheduling subsidiary and forward the resulting plans as executable orders for implementation to the executing control device level of the DES.

The **Scheduling Module** serves as an implementation of the PFM for the subsidiary coordination of resources, processes and products. The scheduling problem of the use case describes an extended flexible job shop problem. The scheduling module for solving the problem was implemented in Python and solved using a heuristic based on the Tabu-Search method. The holons use the module to perform an allocation of the hierarchically distributed operations of a task with the resources available in the respective holon. The agents have the possibility to use the heuristics of the module when requesting an offer as well as to optimize the final schedules by the heuristics and to forward concrete orders based on them. The scheduling module optimizes the schedule according to the lead time.

The developed system structure represents a production system as **a system of systems,** which is able to solve subsidiary decisions on a short-term decentralized shop-floor level as well as medium- and long-term planning problems through the coupled MAS. After an initial decomposition of the planning problem in the MAS, a sequence plan is generated by the holons communicating with each other. The final production orders are derived and forwarded to the DES resources using the middleware. The ad-hoc transport planning as well as the short-term decentralized reaction of the resources to disruptions is realized by the horizontal communication in the DES.

5   Simulation Setup and Results

In this section, the configuration and the corresponding execution of the simulation runs are explained in more detail. In a first step, the different variables are presented as well as the evasion strategies of the resources on control-level-device and station-level within the simulation runs.

The overall layout of the DES, consisting of 84 control-devices containing skills grouped into 16 stations in 1 work-center is designed as a matrix. The production program of 300 products and three types is initially planned for 110 products and then released in groups of 30 products when the number of products in the system is lower than 80 products. The buffer capacity of resources on control-device-level is limited to 2 with an availability of 95% with an MTTR of 2:30. The transport within each station is realized by 5 AGV, for inter-station transport with 40 AGV. The station buffer capacity is set to 20.
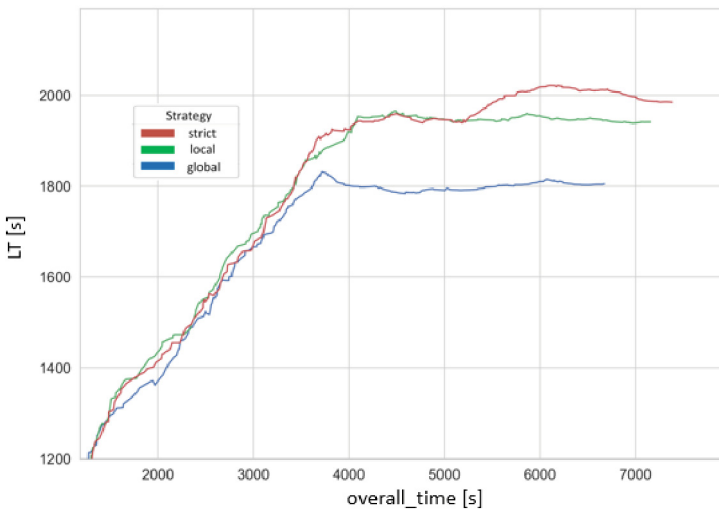
Fig. 4. Exemplary graph of three different simulation runs following different manufacturing strategies

The simulation runs follow different manufacturing strategies to show the effects of different collaborative decentral and central control mechanisms. Three strategies *strict (strict plan execution), local (local negotiation for alternatives) and global (global negotiation for alternatives)* were followed to simulate different situations of deviation events at control device or station level. The *strict* strategy follows the initial plan from the scheduling tool and does not allow alternative process sequences. The *local* strategy allows to reassign a product in case a resource failure occurs. In this case, the alternatives that the technical precedence graph allows are evaluated. Either the process step can be completed in the same station at a different resource with the required skill or a different order with a different skill requirement can be assigned. Compared to the *local* strategy, the *global* strategy additionally allows the selection of another station within

the overall system to process the open order by using the agent interaction for negotiation of alternative resource allocation and product sequences.

While the lead time (LT) for 300 products of the *strict* simulation runs is 7670 s on average, the LT decreases to 7313 s in case of the *local* strategy. The *global* strategy reduces to an averaged 6891 s. The transportation time percentage (AVG_TP) with the *global* strategy of all products are on average 2.8% higher than with the *strict* strategy. At the same time, the storage time percentage (AVG_SP) of the products are reduced by 5.8%, which results in an overall increase of the percentage of value-added production time (PP) by 3%. Comparing the lead times for the mentioned production program, this results in an average reduction of the LT of 11.2% for the same product mix and the same incident volume. Figure 4 shows an exemplary graph of lead time of products LT leaving the system in an exemplary graph. In the start-up phase of the system, the strategies do not differ due to low competition for resources. After the ramp up phase, the saving corresponds to about 11.2% (Table 1).

Table 1. Results of average makespan [overall_time], transportation time percentage [AVG_TP], average storage time percentage [AVG_SP], average value-added production time percentage [AVG_PP] and average leadtime [AVG_LT] of the different manufacturing strategies in 27 simulation runs

Strategy	overall_time [s]	AVG_TP [%]	AVG_SP [%]	AVG_PP [%]	AVG_LT [s]
global	6891.108	13.9	57.3	28.9	1890.744
local	7313.016	12.6	59.9	27.6	1986.008
strict	7670.065	11.1	63.2	25.9	2122.766

6 Conclusion

This paper shows a system structure for holistic information processing with centralized and decentralized intelligence. This is based on an interaction structure for cross-system (horizontal) and cross-level (hierarchically) interaction for an autonomously organized production flow. A hybrid control paradigm was implemented using MAS and shows its potential for the use of centralized and decentralized planning and control in a use case of a modular production systems. This allowed the implementation of an agent-based planning mechanism that realizes sequence planning over multiple levels. The planning and control mechanism was designed using an interaction model for a holistic support. For the execution of the planning and control in dynamic environments a hybrid decision support is needed, which was prototypically implemented in a simulation experiment. The interaction model was implemented focusing on cross-level interaction and collaborative decision making between holons of different systems in a MAS. The collaborative decision making based on cross-level and cross-system interaction in a holonic MAS enables an optimization of the lead time by more than 10% for dynamic production situations. Agent based scheduling and control is used in different scenarios

in the experiment. Firstly, a global optimal schedule for a multi-level production flow was created by negotiating between different local optima in the manufacturing system. Secondly, during manufacturing execution the agents optimize the global production flow autonomously by cross-system and cross-level interaction in system. By using the interaction structure, an alternative production flow is created and orchestrated, which solves unforeseen events e.g. deviations caused by machine failures. In future experiments, the interaction structure will be thoroughly validated by adding additional agents to the holonic architecture. For further research, additional scheduling methods can be applied as well as further enhancements of the cascaded control loops to allow a more dynamic and predictive assessment of deviations and possible risks.

References

1. Kern, W.: Modulare Produktion – Methodik zur Gestaltung eines modularen Montagesystems für die variantenreiche Serienproduktion. Springer Vieweg. Stuttgart, Univ., Diss, Wiesbaden (2021)
2. Komesker S. et al. Structured information processing as enabler of versatile, flexible manufacturing concepts. In: Weißgraeber P., Heieck F., Ackermann C. (eds.) Advances in Automotive Production Technology – Theory and Application. ARENA2036, pp. 108–116. Springer Vieweg, Berlin (2021). https://doi.org/10.1007/978-3-662-62962-8_13
3. Kagermann H, Wahlster W, Helbig E.: Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0: Abschlussbericht des Arbeitskreises Industrie 4.0, Bundesministerium für Bildung und Forschung. Frankfurt am Main, Germany: Forschungsunion Wirtschaft und Wissenschaft (2013)
4. Bajzek, M., Fritz, J., Hick, H.: Systems engineering principles. In: Hick, H., Küpper, K., Sorger, H. (eds.) Systems Engineering for Automotive Powertrain Development. P, pp. 149–194. Springer, Cham (2021). https://doi.org/10.1007/978-3-319-99629-5_7
5. Komesker, S., et al.: Enabling a multi-agent system for resilient production flow in modular production systems. Procedia CIRP **107**, 991–998 (2021)
6. Trunzer E. et al.: System architectures for Industrie 4.0 applications: derivation of a generic architecture proposal. Prod. Eng. **13**, 247–257 (2019)
7. Cruz Salazar, L.A., Ryashentseva, D., Lüder, A., Vogel-Heuser, B.: Cyber-physical production systems architecture based on multi-agent's design pattern—comparison of selected approaches mapping four agent patterns. Int. J. Adv. Manuf. Technol. **105**(9), 4005–4034 (2019). https://doi.org/10.1007/s00170-019-03800-4
8. Ruskowski M. et al. Production Bots für Production Level 4: Skill-basierte Systeme für die Produktion der Zukunft. atp magazin **62**(9), 62–71 (2020)
9. Smart Factory e.V., Production Level 4 – Der Weg zur zukunftssicheren und verlässlichen Produktion, Whitepaper, SF_Whitepaper-Production-Level-4_WEB.pdf (smartfactory.de), Accessed 08 June 2022
10. Farid, M., Ribeiro, L.: An axiomatic design of a multiagent reconfigurable mechatronic system architecture. IEEE Trans. Industr. Inf. **11**(5), 1142–1155 (2015)

11. Valckenaers, P., et al.: Designing Holonic manufacturing systems. Rob. Comput.-Integr. Manuf. **14**, 455–464 (1998)
12. Lueder, A.: Flexibility in production systems by exploiting cyberphysical systems. Computer **53**(1), 81–85 (2020)
13. Leusin, M., et al.: Solving the job-shop scheduling problem in the industry 4.0 era. Technologies **6**(4), 107 (2018)
14. Kossiakoff, A., et al.: Systems Engineering Principles and Practice, 2nd edn. John Wiley, New York (2011)
15. Haberfellner, R., et al.: Systems Engineering Fundamentals and Applications. Springer, Switzerland (2019)
16. Ropohl, G.: Allgemeine Technologie: eine Systemtheorie der Technik. KIT Scientific Publishing, Karlsruhe (2009)
17. Weilkiens, T., et al.: Model-Based System Architecture, 2nd edn. John Wiley, New York (2022)
18. Buede, D.: The Engineering Design of Systems: Models and Methods, 2nd edn. John Wiley, New York (2008)
19. SEBoK Guide to the systems engineering body of knowledge (SEBoK). http://www.seb okwiki.org/w/index.php?title¼Guide_to_the_Systems_Engineering_Body_of_Knowledge_ (SEBoK)&oldid¼53122. Accessed 08 June 2022

Resilience to Pandemics through Flexibility in Sourcing, in Order Fulfillment, and Production Capacity of the Automotive Supply Chain

Marc Gebauer[1]([✉]) [ID] and Cyrine Tangour[2] [ID]

[1] Brandenburg University of Technology, Platz der Deutschen Einheit 1, 03046 Cottbus, Germany
marc.gebauer@b-tu.de

[2] Fraunhofer-Zentrum für Internationales Management und Wissensökonomie, Neumarkt 9, 04109 Leipzig, Germany

Abstract. Since the COVID-19 pandemic, the automotive industry which is regarded as a best practice considering its supply chain has experienced new threats which render its supply chain vulnerable. For instance, the many lockdowns, associated with collapsing of global distribution channels sunk vehicle sales dramatically. The purpose of this study is to identify capabilities to strengthen the resilience of automotive supply chains to pandemics. By using the supply chain resilience framework developed by Sytch et al. [4] we analyze the resilience of the automotive supply chain with evidence from the literature to a pandemic crisis with vulnerability factors similar to those of COVID-19.

We do find evidence that seven out of ten capabilities we looked for are present in the automotive supply chain. Capabilities to improve are i.e., multiple sources for tier 1 suppliers, improving risk pooling/sharing, and defining means of production postponement.

With the evidence on resilience factors for pandemics, we provide managers with a set of factors to focus on in pandemics. Thus, our study helps managers to better prepare their supply chain to resist global crises such as the COVID-19 pandemic. We used a methodology that can be applied with more secondary and also primary sources and therefore is interesting for researchers.

Keywords: resilience factors · vulnerability factors · supply chain · automotive industry

1 Introduction

The COVID-19 pandemic, which started in China in December 2019 uncovered many vulnerabilities of global supply chains [1, 2]. Many countries closed their borders to travelers and goods transport – by air, land, or sea- to limit the spread of the virus on their territories [2]. Direct consequences are the collapse of global distribution routes, economic self-isolation, "consumers reduced income" and "production shutdowns" [3], which impacted both the demand and supply sides of supply chains worldwide [2].

Because of its heavy reliance on a fragmented and global supply chain, the automotive industry is a typical sector that got hit heavily by the COVID-19 outbreak [2–4]. Major disruptions, such as the COVID-19 pandemic, result in deep cuts in sales for production companies like the automotive industry (e.g. [5, 6]). For instance, vehicle sales sunk by almost 20% in February 2020 compared to the previous year Free and Hecimovic [2] relate that in the USA, 93% of all automotive production was stopped by late March 2020. In Europe, it is estimated that 1.1 million workers in the automotive industry lost their jobs because of the pandemic [7].

The automotive supply chain is highly vulnerable to a turbulent external environment, consequently to its complex and globalized supply chain which is associated with higher uncertainty than other sectors' supply chains [8, 9]. Managers estimate that supply chain risks constitute the most severe threat to their company [10]. Consequently, managing risks and uncertainty of supply chains is an essential ability in the automotive industry [9], which can be understood as the resilience of supply chains.

Starting in the middle of the 1990s, researchers began to develop approaches to increase the resilience of supply chains to disruptions [11]. First, they identified resilience approaches to improve supply chain efficiency, such as just-in-time production [12]. Since the COVID outbreak, the focus of research on the resilience of the supply chain to various vulnerability factors grows strongly. A search conducted on 27th May 2022 on the Web of Science database with the terms "resilience" and "supply chain" results in eight publications for 2010 and 573 publications for 2021. Resilience in a general meaning is the ability to resist disruption. The resilience of supply chain research is segmented into at least five fields, namely: (1) types of disruptions, (2) resilience phases, (3) resilience strategies, (4) resilience methods and tools, and (5) capabilities for resilience. While research in each distinct field is abundant, research that crosses fields is less frequent. In the study of Sytch et al. [4], a framework is conceptually developed that bridges the research fields of type of disruption or vulnerabilities and the capabilities for supply chain resilience. We use a qualitative content analysis approach in combination with the framework of Sytch et al. focusing on capabilities for resilience to future pandemics based on information from scientific literature. To what extent are the capabilities of the automotive industry resilient to vulnerability factors related to a pandemic situation similar to COVID-19? The focus is on flexibility in sourcing, in order fulfillment, and production capacity of the automotive supply chain as main capabilities to be resilient to a pandemic.

The paper is structured as follows: After the introduction, we provide an explanation of the resilience of supply chains followed by remarks about the resilience of the automotive supply chain, recommendations for the automotive supply chain based on our analysis, and the conclusion.

2   Resilience of Supply Chains

A generally agreed-on definition of supply chain resilience has not yet emerged in the literature [11]. It can be constructed from its elements "resilience" and "supply chain". Resiliency in general is "…the capacity for an enterprise to survive, adapt, and grow in the face of turbulent change…" [13] and "…the ability of a system to return to its

original (or desired) state after being disturbed" [14]. This system can be a company's supply chain. It is defined as "…the network of companies involved in the upstream and downstream flows of products, services, finances, and information from the initial supplier to the ultimate customer" [15]. Thus, we do define supply chain resilience as the ability of the supply chain to return to its original state (measured in dimensions of performance) or a higher one after being disturbed. The literature on supply chain resilience is divided by a) phases, i.e., pre disruption, during disruption, and post disruption, b) strategies, which can be proactive, concurrent, or reactive, and c) the capabilities to anticipate, adapt, respond, recover, and learn. [16] This study focuses on capabilities. Companies can be resilient by developing capabilities (C_i) to overcome the vulnerability factors (V_j). In this case, resilience factors (V_j, C_i) are present as shown in Table 1 [4]. Vulnerability factors are turbulences, deliberate threats, external pressures, resource limits, sensitivity, connectivity, and supplier/customer disruptions. And capabilities factors can be flexibility in sourcing, flexibility in order fulfillment, capacity, efficiency, visibility, adaptability, anticipation, recovery, dispersion, collaboration, organization, market position, security, and financial strength [15]. This study's focus lies on flexibility in sourcing which is "the ability to quickly change inputs or the mode of receiving inputs", flexibility in order fulfillment meaning the "ability to quickly change outputs or the mode of delivering outputs" and capacity which is the "availability of assets to enable sustained production levels" [15] since flexibility and capacities have been found to be effective strategies for reducing the effects of disturbances [17].

Table 1. Framework for resilience factors (Source: own table based on [6])

	Vulnerability factors (V_j)
Capabilities (C_i)	Resilience factor (V_j, C_i)

3   Supply Chain Resilience of the Automotive Industry

3.1   Procedure

For the collection of data, the method of a systematic literature review is used for which the Search-Appraisal-Synthesis-Analysis (SALSA) framework of Booth et al. [18] is followed. To capture more relevant studies, two search levels are defined. First, one of the largest scientific databases, Web of Science, is systematically searched by using the following predefined keywords: [TS = Resilience AND automotive AND "supply chain"]. The search scope is limited to only peer-reviewed academic publications in English or German with no timeframe limitations. 38 papers are first found. Their appraisal is conducted through a streamlined deductive qualitative content analysis of the full-text versions. Thus, the main selection criteria is the availability of information to be extracted to fill the theoretical framework table. Since not all information is found in the selected papers, thus, a second search was conducted on the more inclusive scientific database Google Scholar. The aim is to find other secondary sources, such as industry

reports or conference proceedings, to complete missing information of the theoretical framework. More papers are selected, which rises up the number of secondary sources selected for this study. The data from the secondary sources are summarized in a table. This table is the base for the analysis in this study. It shows the specific capabilities that are present or absent in the current automotive industry, with respect to flexibility in sourcing, order fulfillment, and production capacity.

Data analysis and recommendation generation are conducted by two researchers independently, based on the filled theoretical framework table. Both outcomes of the analysis and recommendations are then compared and discussed. Differences are argued and the consensus is reached. The results are presented in Sects. 3.3 and 4.

3.2 State of Art of the Automotive Supply Chain

The automotive supply chain produced 80 million cars including commercial vehicles in 2021 and results in a yearly turnover of 800 million € for only the top 100 suppliers [19]. It spreads over 67 countries and is therefore a global supply chain [4]. A global supply chain is to be understood as

"[r]aw materials and intermediate goods are now frequently shipped across the globe several times before final products are exported to final consumers around the world, coordinated by accounting technologies predicated on global labor arbitrage, cost minimization, lean inventory management, and tax avoidance" ([2, S. 2] based on a literature review).

On the one hand, the automotive industry is seen as "the leader in supply chain management" [20, S. 704]. This results in automotive supply chains being well researched (for instance [6]). We find research about the automotive supply chain and its resilience in Japan, Thailand [21], China [22], Iran [23], Brazil [24], Turkey [25], Portugal [17], and Germany [26]. On the other hand, the automotive supply chain has been hit heavily by the Covid-19 pandemic. "…automakers' supply networks are not nearly as robust […] to global disruptions [like the] pandemic" [4, S. 129]. Reasons for the vulnerability of the automotive supply chain lay in the focus on cost reduction and efficiency [17] as a goal reached by lean management, just-in-time delivery, and outsourcing [27, 28]. The analysis of automotive supply chains generally results in the structure of an OEM, 1st-tier suppliers, 2nd-tier suppliers, and more entities. General Motors counts 193 unique tier-1 suppliers, 899 unique tier-2 suppliers, and 2,875 unique tier-3 suppliers. For Volkswagen, the numbers are 213, 1,026, and 2,901 ([4] with data from Bloomberg). More specifically companies in the automotive supply chain act as supplier of parts and materials, are assemblers, logistic service providers or specialize in research or testing. Additionally, regulatory bodies are included [26]. Thus, it can be considered complex [17] providing also complex products compared to other global supply chains [29] (see Fig. 1).
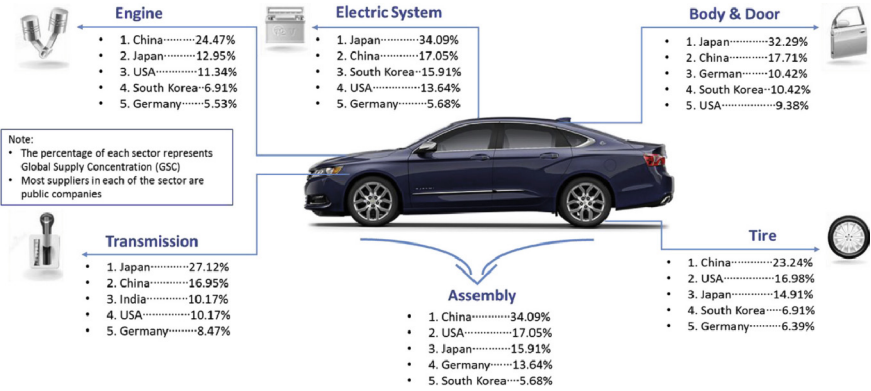
Fig. 1. Supply chain mapping for the global automotive industry [29]

3.3   Sourcing and Fulfillment Capabilities of the Automotive Supply Chain

In summary, only seven of the ten capabilities we focus on in this study are found in the current automotive supply chain (see Table 2).

Table 2. Capabilities of the automotive supply chain from the literature (Source: own table)

Capability		Status	Ref.
Flexibility in sourcing	Part/input commonality	Mostly present	[6, 30]
	Modular product design	Mostly present	[31]
	Supplier contract flexibility	Mostly present	[32, 33]
	Multiple sources for Tier 1 suppliers	Mostly absent	[6]
	Multiple sources for Tiers 2 and 3 suppliers	Mostly present	[26]
Flexibility in fulfillment	Alternate distribution channels	Mostly present	[34, 35]
	Risk pooling/sharing	Mostly absent	
	Production postponement	Mostly absent	
Capacity	Reserve capacity	Mostly present	[24, 26]
	Redundancy	Mostly present	[4]

We did find evidence for most of the resilience capabilities belonging to flexibility in sourcing, flexibility in fulfillment, and capacity in the literature on automotive supply chains. Starting from the beginning of automotive production until now parts, e.g. axles or wheels are used partly as standards for different products to cut costs [6, 30]. Thus, we do find evidence for the capability part and input commonality. Considering the modular product design which is meant to keep control of the growing variety of products [31]. The supplier contract flexibility can be found in the automotive supply chains

for the quantity of supply, e.g. of engines [32, 33]. The use of multiple sources is not really found for the tier-1 suppliers. Toyota, as an example, "interacts directly and closely with the first-layer suppliers […] [s]ince such key components are customized, only a few pre-selected vendors are in its first layer…" [6, S. 219]. The other layers with less specialized components tell a different story. Here we do find cases in which "[i]f one supplier fails […] still at least one [will be available] ensuring the delivery of parts" [26, S. 245]. Going further with the attributes for flexibility in fulfillment cases of alternative distribution channels can be found. One exemplary company "…has a nationwide network to facilitate the customers. This network includes 349 dealers and their workshops, 1541 authorized service centers covering 897 cities and towns" [34, Abs. 3.2]. Additionally, the means of transportation have been changed in one case after the attacks on the twin towers [35]. For risk pooling or sharing as another resilience capability and delayed commitment and production postponement, we did not find evidence in the literature. Lastly, the elements of capacity, reserve capacity, and redundancy are generally found present according to examples from literature, such as safety stocks [24, 26] and "…manufacturing presence in multiple countries…" [4, S. 125].

4   Status Quo and Recommendations

Based on the COVID pandemic, four subfactors are defined to describe the vulnerability that can result from a pandemic crisis, namely: barriers to import and export, lockdowns, the unpredictability of demand, and fluctuations in currencies and prices. The effect of the current capabilities of the automotive supply chain (see Table 2) on the mitigation of vulnerabilities subsequent to a pandemic is evaluated. The results are summarized in Fig. 2 and detailed in the text hereafter. Based on these findings, a recommendation for a higher resilience of the automotive supply chain is drawn.

Parts/components commonality (C1) and modular product design (C2) have similar effects on the resilience of the automotive supply chain. The capability of the automotive industry of using parts and components from more than one product or product families decreases the resilience of its supply chain to a barrier of import and export in which the delivery of parts by global suppliers is interrupted. Furthermore, lockdowns, in which workers are not allowed to go to factories, lead to the interruption of the production of parts. When parts are common to many products, such is the case with the automotive industry, by ripple effect the supply chain is disrupted. Similarly, Modular product designs decrease automotive supply chain resilience to barrier of import and export and lockdowns. Especially in European and USA, the automotive industry tends toward a modular product design, in which they outsource car modules from 1st-tier suppliers [6]. The resilience of their supply chain is highly reliant on suppliers of modules (i.e., 1st-tier suppliers). Modular product design and parts commonality increase the resilience of the automotive supply chain to the unpredictability of demand, as they increase their flexibility in adapting their production lines to fluctuating demand. No discernable effects on the resilience of the automotive supply chain to fluctuations in currencies and prices could be identified.

Supplier contract flexibility (C3) and reserve capacity (C6) have similar effects on the resilience of the automotive supply chain. In a flexible supplier contract, it is agreed

Capabilities (C_i)		Vulnerability from Pandemic (V_i)			
		V1 Barrier of import and export	V2 Lockdown	V3 Unpredictability of demand	V4 Fluctuations in currencies and prices
C1	Parts/components commonality	**DR 1,1**	**DR 1,2**	IR 1,3	NE 1,4
C2	Modular product design	**DR 2,1**	**DR 2,2**	IR 2,3	NE 2,4
C3	Supplier contract flexibility	IR 3,1	IR 3,2	IR 3,3	IR 3,3
C4	Multiple sources for tiers 2 and 3 suppliers	IR 4,1	IR 4,2	NE 4,3	IR 4,4
C5	Alternate distribution channels	IR5,1	IR 5,2	NE 5,2	IR 5,4
C6	Reserve capacity	IR 6,1	IR 6,2	IR 6,3	IR 6,4
C7	Redundancy	IR 7,1	IR 7,2	NE 7,3	IR 7,4
DR: Decreases automotive supply chain resilience					
IR: Increased automotive supply chain resilience					
NE: No effect on automotive supply chain resilience					

Fig. 2. Diagnostic of the resilience of the current automotive supply chain to COVID-19-like pandemic crisis (Source: own figure)

that a predefined amount of an order can be revised up or down. Currently, flexible supplier contracts are used by defining a corridor in which the volume to be supplied can vary [32]. Such flexibility increases undeniably the automotive supply chain's resistance to all vulnerability sub-factors associated with a pandemic situation. According to the literature, automotive firms use different tools to define safety stock levels to mitigate crises [24, 36]. This capability increases the automotive supply chain's resilience to a pandemic situation in all four aspects.

Multiple sources for tiers 2 and 3 suppliers (C4), alternate distribution channels (C5) and redundancy (C7) have similar effects on the resilience of the automotive supply chain. In general, the capabilities of flexible sourcing through alternative suppliers of parts mitigates disruption from the barrier of import and export, lockdowns, and fluctuations in currencies and prices. Similarly, the capability to use flexible ways of transportation and to switch transportation means, when necessary, allows the automotive supply chain to mitigate barrier of import and export, lockdowns, and fluctuations in currencies and prices. Finally, redundancy, such as by establishing manufacturing facilities in many countries [4], increased the resilience of the automotive supply chain to the barrier of import and export, lockdown, and fluctuations in currencies and prices. The effect of multiple sources, alternate distribution channels, and redundancy is not discernable on the unpredictability of demand.

Based on the results of this paper, we recommend the following. First, to increase the resilience of the automotive supply chain, the three capabilities that do not exist in the current automotive supply chain should be developed, namely: (1) multiple sources for tier 1 suppliers, (2) risk pooling/sharing, and (3) production postponement. As tier 1 suppliers produce and supply more critical parts of a car than tier 2 and 3 suppliers, the effect of relying on multiple sources would be more pronounced on supply chain resilience. It is recommended that automotive assemblers should have more than 1 tier supplier to increase supply chain resilience to the barrier of import and export, lockdowns, and fluctuations in currencies and prices. Furthermore, risk pooling and sharing is a key capability to mitigate vulnerability due to fluctuating demand and prices, supply and demand uncertainty [37], such as the barrier to import, export, and lockdowns. Moreover, the capability of production postponement allows to shift manufacturing activities across different products [37], thus allowing companies to be resilient to all four vulnerability aspects associated with a pandemic situation.

When taken apart, two capabilities render the automotive supply chain more vulnerable to pandemics which are parts commonality and their modular product design. At the same time, these capabilities are a cornerstone of the high efficiency of the automotive supply chain. The recommendation is to find a balance between high efficiency and high resilience of the automotive supply chain, by considering key capabilities in connection with each other. For instance, parts that should be common to many product lines should have multiple suppliers and redundancy in their manufacturing and storage. In that way, the capability of parts commonality (C1) is moderated by multiple sources (C4), reserve capacity (C6), and redundancy (C7).

5 Conclusion

The pandemic shows vulnerabilities in the automotive supply chain which can be regarded as one of the best. Parallelly, the complementary interest in research in supply chain resilience has been growing in the last years. The literature presents us with concepts, simulations, and with some empirical studies about supply chains being disturbed by vulnerabilities. Our study provides evidence from the literature that the automotive supply chain could improve its resilience by finding multiple sources for tier 1 suppliers, improving risk pooling/sharing, and finding possibilities for production postponement. These recommendations are based on a single literature study. Thus, they offer orientation but need to be weighed carefully by managers. Further research can be done with different secondary or new primary data.

References

1. Golan, M.S., Jernegan, L.H., Linkov, I.: Trends and applications of resilience analytics in supply chain modeling: systematic literature review in the context of the COVID-19 pandemic. Environ. Syst. Decis. **40**(2), 222–243 (2020). https://doi.org/10.1007/s10669-020-09777-w
2. Free, C., Hecimovic, A.: Global supply chains after COVID-19: the end of the road for neoliberal globalisation? Account. Auditing Accountability J. **34**(1), 58–84 (2021). https://doi.org/10.1108/AAAJ-06-2020-4634

3. Marcucci, G., Mazzuto, G., Bevilacqua, M., Ciarapica, F.E., Urciuoli, L.: Conceptual model for breaking ripple effect and cycles within supply chain resilience. In: Supply Chain Forum, Milton Park, Abingdon, Oxon, England. Taylor & Francis Ltd. (2022). https://doi.org/10.1080/16258312.2022.2031275

4. Sytch, M., Kim, Y., Page, S.: Supplier-selection practices for robust global supply chain networks: a simulation of the global auto industry. Calif. Manag. Rev. **64**(2), 119–142 (2022). https://doi.org/10.1177/00081256211070335

5. Balakrishnan, A.S., Ramanathan, U.: The role of digital technologies in supply chain resilience for emerging markets' automotive sector. Supply Chain Manag. Int. J. **26**(6), 654–671 (2021). https://doi.org/10.1108/SCM-07-2020-0342

6. Matsuo, H.: Implications of the Tohoku earthquake for Toyota's coordination mechanism: supply chain disruption of automotive semiconductors. Int. J. Prod. Econ. **161**, 217–227 (2015). https://doi.org/10.1016/j.ijpe.2014.07.010

7. ACEA: COVID-19: Jobs of over 1.1 million EU automobile workers affected so far, data shows (2020). https://www.automotiveworld.com/news-releases/acea-covid-19-jobs-of-over-1-1-million-eu-automobile-workers-affected-so-far-data-shows/

8. Govindan, K., Azevedo, S.G., Carvalho, H., Cruz-Machado, V.: Impact of supply chain management practices on sustainability. J. Clean. Prod. **85**, 212–225 (2014). https://doi.org/10.1016/j.jclepro.2014.05.068

9. Junaid, M., Xue, Y., Syed, M.W., Li, J.Z., Ziaullah, M.: A neutrosophic AHP and TOPSIS framework for supply chain risk assessment in automotive industry of Pakistan. Sustainability **12**(1), 154 (2020). https://doi.org/10.3390/su12010154

10. FM Global: Managing Business Risk through 2009 and Beyond (2007). http://www.protectingvalue.com

11. Pettit, T.J., Croxton, K.L., Fiksel, J.: The evolution of resilience in supply chain management: a retrospective on ensuring supply chain resilience. J. Bus. Logist. **40**(1), 56–65 (2019). https://doi.org/10.1111/jbl.12202

12. Revilla, E., Saenz, M.J.: The impact of risk management on the frequency of supply chain disruptions: a configurational approach. Int. J. Oper. Prod. Manag. **37**(5), 557–576 (2017). https://doi.org/10.1108/IJOPM-03-2016-0129

13. Fiksel, J.: Sustainability and resilience: toward a systems approach. IEEE Eng. Manag. Rev. **35**(3), 5 (2007). https://doi.org/10.1109/EMR.2007.4296420

14. Christopher, M., Rutherford, C.: Creating supply chain resilience through agile six sigma. Crit. Eye **7**(1), 24–28 (2004)

15. Pettit, T.J., Fiksel, J., Croxton, K.L.: Ensuring supply chain resilience: development of a conceptual framework. J. Bus. Logist. **31**(1), 1–21 (2010). https://doi.org/10.1002/j.2158-1592.2010.tb00125.x

16. Ali, A., Mahfouz, A., Arisha, A.: Analysing supply chain resilience: integrating the constructs in a concept mapping framework via a systematic literature review. Supply Chain Manag. Int. J. **22**(1), 16–39 (2017). https://doi.org/10.1108/SCM-06-2016-0197

17. Carvalho, H., Barroso, A.P., Machado, V.H., Azevedo, S., Cruz-Machado, V.: Supply chain redesign for resilience using simulation. Comput. Ind. Eng. **62**(1), 329–341 (2012). https://doi.org/10.1016/j.cie.2011.10.003

18. Booth, A., Papaioannou, D., Sutton, A.: Systematic approaches to a successful literature review (2012)

19. Kords, M.: Statistiken der weltweiten Automobilindustrie (2022). https://de.statista.com/themen/2670/weltweite-automobilindustrie/#topicHeader__wrapper

20. Wicher, P., Stas, D., Karkula, M., Lenort, R., Besta, P.: A computer simulation-based analysis of supply chains resilience in industrial environment. Metalurgija **54**(4), 703–706 (2015)
21. Haraguchi, M., Lall, U.: Flood risks and impacts: a case study of Thailand's floods in 2011 and research questions for supply chain decision making. Int. J. Disaster Risk Reduction **14**(3), 256–272 (2015). https://doi.org/10.1016/j.ijdrr.2014.09.005
22. Millar, M.: Supply Chain Risk - the harsh lessons of Tianjin. Intralogistics Mag. (2015). chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/. https://static1.squarespace.com/static/54f601e1e4b02adee20cb495/t/567a4b60c21b86984b67ad3e/1450855264419/Supply+Chain+Risk+Lessons+of+Tianjin%2C+IntraLogistix+%28UK%29+Mark+Millar+1115.pdf
23. Kaviani, M.A., Tavana, M., Kowsari, F., Rezapour, R.: Supply chain resilience: a benchmarking model for vulnerability and capability assessment in the automotive industry. Benchmarking Int. J. **27**(6), 1929–1949 (2020). https://doi.org/10.1108/BIJ-01-2020-0049
24. Vanalle, R.M., Lucato, W.C., Ganga, G.M.D., Alves Filho, A.G.: Risk management in the automotive supply chain: an exploratory study in Brazil. Int. J. Prod. Res. **58**(3), 783–799 (2020). https://doi.org/10.1080/00207543.2019.1600762
25. Evren, Y., Sakarya, A.: Resilience, the rivet effect, and the changing geography of production chains in the Turkish automotive industry. Planlama-Plann. **28**(1), 12–21 (2018). https://doi.org/10.14744/planlama.2017.36035
26. Thun, J.-H., Hoenig, D.: An empirical analysis of supply chain risk management in the German automotive industry. Int. J. Prod. Econ. **131**(1), 242–249 (2011). https://doi.org/10.1016/j.ijpe.2009.10.010
27. Boysen, N., Emde, S., Hoeck, M., Kauderer, M.: Part logistics in the automotive industry: decision problems, literature review and research agenda. Eur. J. Oper. Res. **242**(1), 107–120 (2015). https://doi.org/10.1016/j.ejor.2014.09.065
28. Sturgeon, T., Van Biesebroeck, J., Gereffi, G.: Value chains, networks and clusters: reframing the global automotive industry. J. Econ. Geogr. **8**, 297–321 (2008)
29. Xu, M.: Supply chain sustainability risk and assessment. J. Clean. Prod. **225**, 857–867 (2019). https://doi.org/10.1016/j.jclepro.2019.03.307
30. Heikal, E.A., El-Kharbotly, A.K., El-Beheiry, M.M.: A proposed index to assess commonality among aircraft product families. IOP Conf. Ser. Mater. Sci. Eng. **610**(1), 012082 (2019). https://doi.org/10.1088/1757-899X/610/1/012082
31. Parry, G., Graves, A. (eds.): Build to Order. Springer, London (2008). https://doi.org/10.1007/978-1-84800-225-8
32. Niemann, J., Seisenberger, S., Schlegel, A., Putz, M.: Development of a method to increase flexibility and changeability of supply contracts in the automotive industry. Procedia CIRP **81**, 258–263 (2019). https://doi.org/10.1016/j.procir.2019.03.045
33. Thomé, A.M.T., Scavarda, L.F., Pires, S.R.I., Ceryno, P., Klingebiel, K.: A multi-tier study on supply chain flexibility in the automotive industry. Int. J. Prod. Econ. **158**, 91–105 (2014). https://doi.org/10.1016/j.ijpe.2014.07.024
34. Agarwal, N., Seth, N.: Analysis of supply chain resilience barriers in Indian automotive company using total interpretive structural modelling. J. Adv. Manag. Res. **18**(5), 758–781 (2021). https://doi.org/10.1108/JAMR-08-2020-0190
35. Iskandar, I., Darestani, S.A., Ghavami, S., Azizi, A.: Supplier selection problem: a multi-objective model considering both all-unit and incremental discounts with multiple-item. Appl. Mech. Mater. **606**, 271–275 (2014). https://doi.org/10.4028/www.scientific.net/AMM.606.271
36. Spieske, A., Birkel, H.: Improving supply chain resilience through industry 4.0: a systematic literature review under the impressions of the COVID-19 pandemic. Comput. Ind. Eng. **158**, 107452 (2021). https://doi.org/10.1016/j.cie.2021.107452

37. Dehdar, E., Azizi, A., Aghabeigi, S.: Supply chain risk mitigation strategies in automotive industry: a review. In: 2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Bangkok, pp. 84–88, December 2018. https://doi.org/10.1109/IEEM.2018.8607626

Requirements for a Process to Remanufacture EV Battery Packs Down to Cell Level and Necessary Design Modifications

Melina Graner[1(✉)], Frieder Heieck[1], Alexander Fill[2], Peter Birke[2], Woidy Hammami[3], and Katharina Litty[3]

[1] Hochschule Für Angewandte Wissenschaften Kempten, Bahnhofstraße 61, 87435 Kempten, Germany
melina.graner@hs-kempten.de
[2] Universität Stuttgart, Pfaffenwaldring 47, 70569 Stuttgart, Germany
[3] Daimler Truck AG, Mercedesstraße. 128, 70327 Stuttgart, Germany

Abstract. In case of electric vehicles (EV) powered by lithium ion traction batteries (LIB), remanufacturing processes nbecome increasingly important due to their rising market share and valuable raw materials. LIB can account for up to 40% of the total EV cost. Often, only a small portion of the cells are significantly degraded when the usable battery capacity falls below 80%, which is currently considered the standard end-of-life criterion. However, in order to enable efficient remanufacturing, novel battery design principles are required. This paper discusses the requirements, opportunities and challenges of future remanufacturing processes of LIBs down to the cell level using a battery system of a commercial vehicle as an example. It gives an overview of the current state-of-the-art manufacturing processes of battery systems and shows the developed overall remanufacturing process including condition assessment, disassembly and reassembly. Subsequently, requirements on future designs are discussed. The state-of-the-art of EV batteries is evaluated based on these requirements to determine where incompatible connections such as welded contacts or adhesive joints conflict with remanufacturing design principles.

Keywords: Remanufacturing · EV battery · Condition assessment · Modular design · Circular economy

1 Introduction

Battery technology enables the transformation of the mobility sector towards sustainable drive systems. Electrical driving technology appears to be the future of transportation and offers climate and eco-friendly transportation of passengers and goods. EVs are mostly equipped with LIBs. With increasing market share of EVs, the question about how to process a battery after its usage time arises and processes for the recovery of materials and components such as recycling and remanufacturing gain importance. The industrial research and development (R&D) efforts in battery design mainly focus on

the increase of energy and power density, cost reduction, fast charging and improved safety [1]. Tremendous progress has been made in the optimization of battery design on the material level (material for cathode, anode etc.), electrode level (e.g. electrode thickness), cell level (e.g. shape) and system level (mechanical design, battery management system (BMS) etc.) [2]. Concepts for a recycling-oriented design and manufacturing received less attention. In current LIB-systems, single battery cells are connected to form a battery module. Several modules together with additional electrical periphery (e-parts like battery management etc.) form a complete traction battery.

The research gap addressed is the concept of a remanufacturing process for LIBs down to cell level and the associated changes regarding design and assembly of the components. This paper first provides a state-of-the-art review on LIB systems with respect to their life cycle and upcoming recycling directives. Second, an exemplary LIB system with prismatic cells is analysed with respect to the manufacturing process and applied joining techniques, which are decisive for an efficient remanufacturing process. Third, the challenges of the current system regarding an efficient and potentially automated remanufacturing process and necessary design modifications are discussed.

2   State-of-the-Art/Literature Review

This chapter provides a review of the relevant contributions from the existing body of scientific publications, legal regulations and articles in the fields of LIB design, aging behaviour of components and their treatments after end of life back to the year 2006. The European Directive 2006/66/EC takes producers of batteries or producers of components incorporating a battery into responsibility for the waste and recycling management of batteries that they place on the market [3]. A recent update of the directive requires recycling efficiencies and recovery of materials for batteries of 65% by 2025 [4]. Even though there are legal regulations for the material efficiency, there is currently no standardised procedure for the processing of returned batteries. When an EV battery reaches its end of first life, manufacturers have three options: disposal, recycling, or reuse. In most regions, regulation prevents mass disposal. Established recycling processes currently only focus on the highly valued materials such as cobalt and nickel [5].
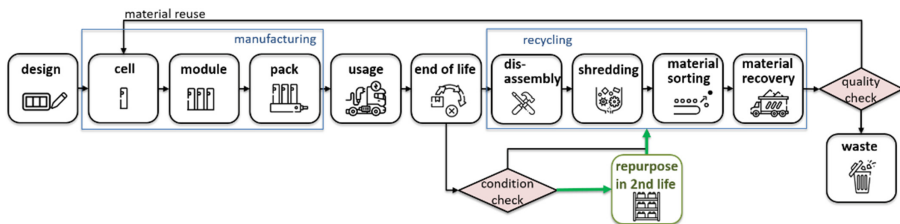
Fig. 1. The life cycle of a LIB includes the battery design, manufacturing, usage life and either a second life or direct recycling of material after the end of first life.

Pyrometallurgy (smelting) offers only a low recycling rate with high-energy input, as cheap base metals such as lithium, graphite or carbon get lost in the process. Hydrometallurgy (leaching, precipitation and resynthesis of elements) uses huge amounts of chemicals, which require professional disposal. New processes that recover more material are not yet fully mature. The third option is the reuse of batteries in stationary energy-storage applications with lower current and energy density demands. However, second-life applications can only extend the lifespans of LIBs. Critical materials still need to be recycled after the end of second life, as it is shown in Fig. 1 [6, 7].

A LIB drops out of first life when it no longer meets EV performance standards, which typically means maintaining 80% overall capacity [5]. It is possible, that the total capacity is determined by a few degraded cells. Kampker et al. carried out an experimental confirmation of cells reliability in 2021. A LIB was disassembled after roughly 288 deep cycles and all 196 cells were tested, resulting that 89% of the cells were still reusable whereas 9% had an insufficient capacity for reuse and 2% showed other types of failures. Of the reusable cells, 68% were probably still useable for automotive applications, 16% only for stationary applications, and 5% were probably not worth being reused, even if they did not fail [8]. These results confirm a simulation by Mathew et al. [9], which estimates that most cells of a used battery pack are worth recovering. The virtual replacement of 5%–30% of the worst aged cells resulted in a restored state of health (abbreviated SOH) of almost 100%. These results, viewed in the context of increasing scarcity of resources and the legal regulations aiming for circular economy for LIBs raise the question how single cells can be restored and reused.

Remanufacturing transforms a degraded product into a quasi-new or an improved functional state by reprocessing and exchanging single components, thus increasing its resource and economic efficiency and enabling circular economy principles. Remanufacturing on LIBs means the exchange or preparation of degraded components or cells. According to the case study scenarios of Alfaro-Algaba et al. [10] and the results of Kampker et al. [11], remanufacturing of LIBs offers great potential in economical and ecological savings. However, there is currently no industrial application of reprocessing LIBs down to cell level. This is due to designs unsuitable for an automated disassembly and varying designs from different manufacturers. Upcoming industrial trends like the cell-to-pack design (with cells directly integrated in the housing without stacking to modules) intensify these problems or simply do not allow remanufacturing [12].

3   Analysis of a State-of-the-Art Battery System

This paper focuses exemplarily on the LIB system of a commercial vehicle. In the e-axle of a commercial vehicle multiple battery packs are combined to reach higher battery capacity for higher power and range. Nevertheless, the composition and components of the LIB system itself are comparable to LIB systems in passenger cars.

Methodology
Based on an overview of the current manufacturing process of LIBs derived from state-of-the-art commercial vehicle LIB systems as well as from a literature review, this paper shows a process step sequence for a remanufacturing process. Various requirements for an effective implementation of this process model are defined. Subsequently

this publication discusses requirements on future designs, which enable the processes derived earlier. Furthermore, it evaluates the exemplary battery design based on these requirements. Figure 2 represents the procedure graphically.
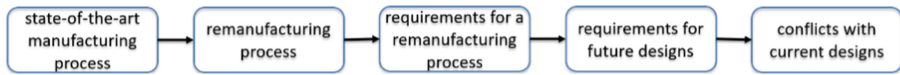
Fig. 2. Methodology of the research.

The Manufacturing Process of a State-of-the-Art Battery System

Figure 3 describes the manufacturing process of the exemplary LIB system (Fig. 4), both derived from a commercial vehicle's LIB system as well as a literature review.
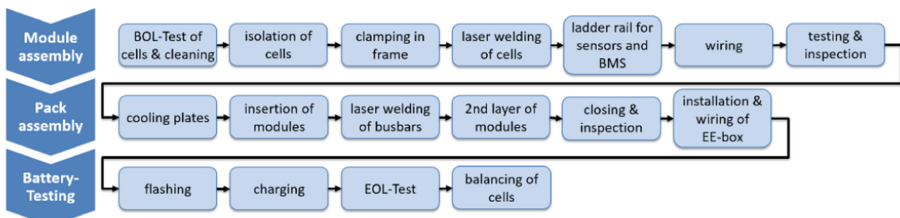
Fig. 3. The manufacturing of a state-of-the-art LIB transferred from [13] to the system in Fig. 4.

An initial test sorts out defective cells using impedance spectroscopy, voltage measurement or capacity analysis. The single prismatic cells are enclosed with adhesive foils for isolation and are combined to cellblocks and clamped in a module frame. A welded ladder rail contacting system interconnects the cells and connects all cells of a module with the cell monitoring board (CMB for voltage and temperature, mounted to the module block). The modules are interconnected with busbars. The most common electric contacting technologies are laser welding, laser bonding or ultrasonic welding [15]. The following is a test of conductivity at the joints. For later connection to the master system, the modules are equipped with cable harnesses. Before positioning the modules in the housing, a cooling plate is placed on the bottom of the pack. Thermal glue or foam between the modules and the cooling plate and additional screw connections to the cross bars of the housing tub secure the module blocks [16].

In the exemplary LIB system in Fig. 4 two layers of modules are arranged in separate housings to increase the battery capacity. They are placed on top of each other, separated by a solid sealant. The topcover closes the housing with screws and a solid sealing. A leak test guarantees its tightness. On top of the two module layers there is the EE-Box (containing electric/ electronic peripheral devices such as contactors, fuse, pyrofuse and busbars to which the wiring harness is attached). It is mounted to the housing and also connects the high-voltage (HV) connector, low-voltage (LV) connector and master BMS. A liquid cooling system for all packs is mounted on one of the packs (not displayed in
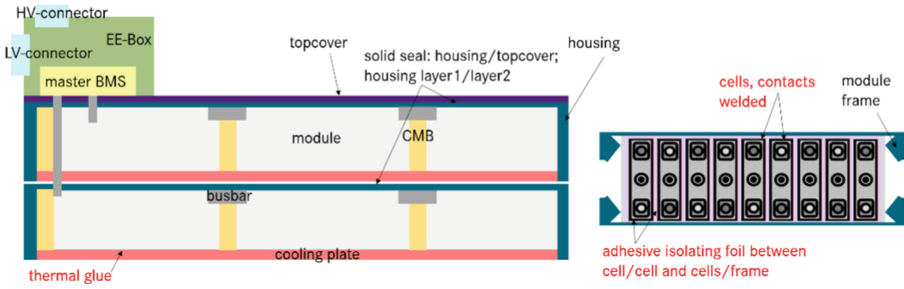
Fig. 4. Schematic design of an exemplary LIB and its components (left) and a module (right).

Fig. 4). The mounting of those components is detachable as usually screws, clips or plugs are used [16].

Remanufacturing Scenarios for LIB Packs

The remanufacturing process is not simply the reversal of the manufacturing process, due to various inextricably designed connections as well as degraded components. Based on the generic design presented in Fig. 4, it can be described by the following steps: analysis and condition assessment, removal process, check and preparation of single components and cells and reassembly/ replacement of components.
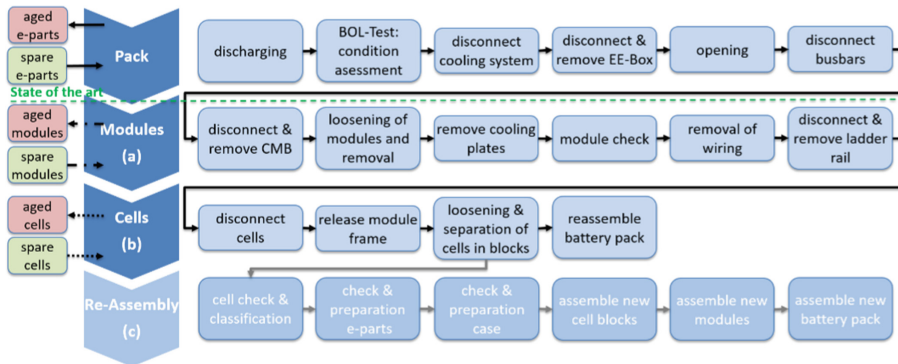
Fig. 5. The developed remanufacturing process for LIBs with different strategies adapted and supplemented from [16].

The first step is to discharge the packs. After that, a Begin-Of-Line Test (BOL) estimates the overall condition of the pack and identifies defect components or cells. The SOH and aging condition of single cells decide about their further usage. The same applies to the functionality of the e-parts, some of which could use an overhaul.

The process sequence depends on the chosen business model or remanufacturing strategy. Figure 5 indicates the strategies (a) with an exchange and replacement of aged modules or (b) aged cells. Strategy (c) means disassembly of the whole system, where components are used as spare parts or in newly reassembled packs. Also a combination

of these strategies is possible: if a module contains a single or few degraded cells they are replaced, otherwise the whole module is exchanged. Some of the remaining cells from the removed module then might serve as spare parts. Similar to strategy (c), Kampker et al. [13] introduce a remanufacturing architecture to build new LIB systems from disassembled components. Strategy (a) is similar to the business model of the Nissan Leaf, where aged modules of LIBs can be replaced- enabled by a special design with very small modules and bolted connections [14].

4   Remanufacturing Challenges and Design Modifications

Challenges and Requirements of Remanufacturing

Due to the wide variety of cell-, module- and pack-types, today the dismantling of LIB systems is time-consuming manual work. Considering the rising amount of returning LIBs (one million are expected to return by 2028), the remanufacturing process needs to be partially or fully automated to implement it industrially and economically [15]. Figure 6 indicates that the expected high number of returning LIBs will require automated processes. It must be further investigated whether a flexible line in terms of variants and quantities is better than an automated sub system separated by cell type (cylindrical, pouch or prismatic) or process steps (e.g. glued or screwed connections).
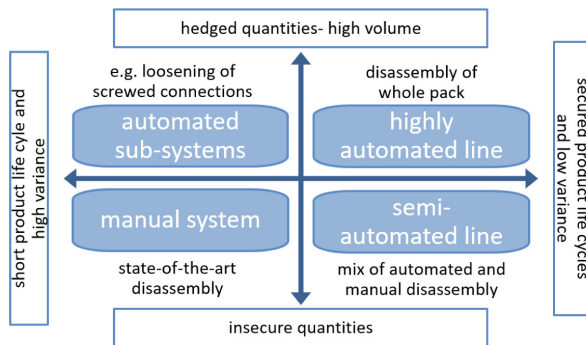
Fig. 6. Different levels of automation in production systems based on the illustration in [15].

To generate the quantities, a well-targeted and systematic collection has to feed the returned LIBs to the remanufacturing process. This also means that batteries from damaged vehicles are handled in different processes due to potential dangers.

A standardised labelling of LIBs supports the idea of automation, so the used materials and joining technologies can be identified ahead for the flexible systems to react or set up accordingly. This is similar to the idea of a battery passport that carries helpful information about the manufacturer, design, data sheets and disassembly manual [17].

A reliable BOL-test needs to identify damaged packs to avoid safety issues or problems in disassembly and to access the quality. LIBs not worth for remanufacturing will be treated differently. In addition, the BOL should evaluate the state of charge.

A safety concept should define the treatment of packs in case of a thermal runaway. Michaelis et al. suggest a standardised BMS-interface to extract life cycle data and information about the usage profile, possible failures in the pack and the SOH [16]. The condition assessment needs reliable, standardised, affordable and non-destructive analytical methods [17]. In addition, reliable inline measurement principles for cell checks and their classification as well as for periphery e-parts need to be developed. Lately a cooperation between Audi and Volkswagen released a diagnosis tool, which assesses the overall condition of the LIB pack and single cells within a few minutes [18].

For operational safety regarding high voltage and states of stress in the components, specially trained staff uses insulated and securing tools. A crane helps handling heavy components, as a pack can weigh over 700 kg in case of heavy-duty vehicles.

Short-circuit hazards also have to be avoided and the process forces and positioning of grippers and tools need to be very precise. The highly sensitive cells and sensors need to be separated but not damaged by mechanical stress or thermal load. Tolerance chains have to be considered in the resulting stress conditions and expansion behaviour.

Looking closer at the LIB design in Fig. 4, the previously described manufacturing process is not reversible due to the applied production methods and materials. The irreversible steps of the manufacturing process require either a non-destructive dismantling technology or a change in the manufacturing process going hand in hand with the redesign of the battery system itself. As the results of the disassembly experiments of LIBs with prismatic cells of Kampker et al. [8] and Schäfer et al. [19] show, there are no such non-destructive dismantling technologies down to cell level with potential for automation. Obviously, adaptions in the design of LIB systems are necessary.

The Main Difficulties of Disassembly of State-of-the-Art LIBs to Cell Level Are:

- welded electrical connections between cells and welded busbars are hard to separate without removal of material (such as drilling or milling of the contacts)
- adhesive isolation between the cells and between cell blocks and the cooling plate makes it very difficult to loosen and separate the cells without damaging them
- the cells must not be harmed in any way (e.g. by penetration or destruction of the outer case, short circuits and deformation due to excessive pressure)
- in order to exchange an individual cell in the module frame, the compression within in the block should be maintained as the cells expanded during usage
- the overall SOH within a block should be balanced, which is why the inserted cells' SOH should fit the SOH of remaining cells
- corrosion, distortion due to thermal and mechanical loads, altered tolerance chains and thereby induced residual stresses complicate the dismantling

Requirements on Future Designs for Remanufacturing:
LIBs with new designs must be comparable to the state-of-the-art LIBs regarding performance, weight and dimensions. Alternative electrical contacting methods should not have higher electrical resistance than on state-of-the-art batteries. Based on the literature review and analysis of the exemplary LIB system, Table 1 suggests R&D activities regarding innovative materials or integration of functions to improve the potential

of future LIB systems for economical remanufacturing scenarios. The overall design should be modular with interfaces for automated disassembly. As few components as possible should be destructed during the disassembly. Inseparable bonds such as welding or gluing should be avoided or at least made in such way that they can be separated.

Table 1. Design modifications for a remanufacturing friendly LIB design

State-of-the-art	Design modification approaches
Thermal glue for heat dissipation	Function integration of cooling channels and/or isolation in the housing of cells, modules or packs
Adhesive gap fillers	Alternative materials such as thermal elastomers with lower adhesion characteristics [20]
Double-sided adhesive film to connect cells	Single-sided adhesive film easier to separate
	insert cells in a plug system/frame for grouping [21]
CMB with welded connection to ladder rail	Function integration of sensors in the housing of cells, modules or packs
Poor accessibility of components which need to be disassembled	Provide space, gaps or edges to level up a component or to ensure the accessibility of components
	Exchange cells without removing the modules if accessibility is given
Welded electric contacts are not detachable	Cable plugs for wiring sensors and cable harnesses
	Laser bonding or micro-clinching are detachable [22]

Based on the above-mentioned requirements, the exemplary design of a state-of-the-art LIB system in Fig. 4 is evaluated regarding suitability for remanufacturing based on the used materials and joining technologies.

The EE-Box including HV-Connector and LV-Connector can be easily removed from the housing even though corrosion over the usage time can complicate the loosening of screws, plugs and clips. Wiring and cable harnesses of the BMS Master and the HV- and LV-module can be disconnected. The design of the topcover and the housing enables an opening of the battery pack.

The busbars for connecting the modules are permanently welded. A separation is only possible with the removal of material (such as drilling or milling of the contacts). To lift out the module (frames) first the screwed connections to the pack housing have to be loosened. The thermal glue or foam between the cell blocks and the cooling plate complicates the removing of the modules and is the cause of possible damages.

Similar to the module level, the welded connections for contacting the cells and CMB cannot be separated non-destructively. Laser cutting of busbars exposes the cells to possible damages from high heat input (risk of fire and reduced efficiency due to chemical reactions in the cell), molten material and debris. Additionally, the adhesive foils between the cells inhibit the separation of cells without mechanical damaging.

5 Discussion of Results

The proposals for remanufacturing friendly design changes made in this paper are partly adaptable to the design introduced in Fig. 4. Function integration of cooling channels in the components involves the risk of undetectable leaking in the system. Better accessibility requires larger installation space. The same applies to the framed plug-in system for cell grouping. Even though the gaps between cells favour ventilation cooling and provide space for the expansion of cells during usage, the greater need for space is a disadvantage compared to cell connection by means of adhesive foils. Alternative materials for thermal isolation with lower adhesion characteristics are subject of current research in the automotive sector. Without adhesive connections and sufficient accessibility, the exchange of cells directly in the pack could be possible without the removal of modules. By developing alternative contacting methods, measurement technology can be integrated for an inline assessment of the quality of the connection and would replace quality control after contacting. [15] In summary, the challenge in implementing remanufacturing friendly design changes is to avoid interference with classical battery system development goals. New LIB designs must be comparable regarding weight and dimensions in order to offer similar power density as state-of-the-art LIBs.

6 Summary and Conclusion

The analysis of the manufacturing process and used materials and joining technologies results, that the current state-of-the art designs of LIB packs like in Fig. 4 do not allow a disassembly down to cell level. The displayed state-of-the-art LIB-design currently allows the exchange of e-parts on pack level. Mainly the adhesive bonds and non-detachable welding connections inhibit dismantling. This paper suggests design modifications regarding better accessibility of components for disassembly, avoidance of adhesive gap fillers or glue and the use of alternative electrical contacting. High potential lies in the contacting methods laser bonding or micro-clinching. Cooling function integration in components or alternative materials for isolation help to reduce the use of glue and require further research activity.

References

1. Deng, J., Bae, C., Denlinger, A., Miller, T.: Electric vehicles batteries: requirements and challenges. Joule. **4**, 511–515 (2020)
2. Han, X., Lu, L., Zheng, Y., Feng, X., et al.: A review on the key issues of the lithium ion battery degradation among the whole life cycle. eTransportation **1** (2019)

3. European Commission: Directive 2005/36/EC of the European Parliament and of the Council of 7 September 2005. Off. J. Eur. Union, pp. 216–261 (2006)
4. European Commission: Proposal for a Regulation of the European Parliament and of the Council concerning batteries and waste batteries. Off. J. Eur. Union (2020)
5. Engel, H., Hertzke, P., Seccardo, G.: Electric vehicles, second life batteries, and their effect on the power sector. https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/second-life-ev-batteries-the-newest-value-pool-in-energy-storage. Accessed 24 May 2022
6. Tan, D.H.S., Xu, P., Chen, Z.: Enabling sustainable critical materials for battery storage through efficient recycling and improved design: a perspective. MRS Energy Sustain 7(1), 1–13 (2020). https://doi.org/10.1557/mre.2020.31
7. Szombathy, M.: Das Recycling von Altbatterien ist ein entscheidender Zukunftsmarkt, https://recyclingportal.eu/Archive/45716. Accessed 13 Dec 2021
8. Kampker, A., Wessel, S., Fiedler, F., Maltoni, F.: Battery pack remanufacturing process up to cell level with sorting and repurposing of battery cells. J. Remanuf. 11(1), 1–23 (2020). https://doi.org/10.1007/s13243-020-00088-6
9. Mathew, M., Kong, Q.H., McGrory, J., Fowler, M.: Simulation of lithium ion battery replacement in a battery pack for application in electric vehicles. J. Power Sources 349, 94–104 (2017)
10. Alfaro-Algaba, M., Ramirez, F.J.: Techno-economic and environmental disassembly planning of lithium-ion electric vehicle battery packs for remanufacturing. Resour. Conserv. Recycl. 154 (2020)
11. Kampker, A., Hollah, A., Kreisköther, K., Lienemann, C.: CoFAT 2016-Electromobile Remanufacturing-Nutzenpotenziale für batterieelektrische Fahrzeuge. In: Presented at the Conference on Future Automotive Technology (2016)
12. Why Tesla Is So Interested In "Cell-To-Pack" Batteries. https://jalopnik.com/why-tesla-is-so-interested-in-cell-to-pack-batteries-1843463349. Accessed 21 Dec 2021
13. Kampker, A., Heimes, H., Lienemann, C., Grauel, D., Jones, M.: Development of a novel remanufacturing architecture for lithium-ion battery packs. In: Electric Vehicles International Conference (EV), pp. 1–6 (2017)
14. Hanley, S.: Nissan begins offering remanufactured batteries for LEAF. https://cleantechnica.com/2018/05/15/nissan-begins-offering-remanufactured-batteries-for-leaf/. Accessed 6 Jan 2022
15. Michaelis, D., Rahimzei, E., Kampker, P., Heimes, H. et al.: Roadmap Batterie-Produktionsmittel 2030 - Update 2020. VDMA Batterieproduktion (2021)
16. Heimes, H., Kampker, A., Offermanns, C., Kreisköther, K., et al.: Recycling von Lithium-Ionen-Batterien. PEM RWTH Aachen University, BLB TU Braunschweig und VDMA, Frankfurt am Main (2021)
17. Arno, K., Hagelüken, C., Kohl, H., Buchert, M. et al.: Ressourcenschonende Batteriekreisläufe – mit Circular Economy die Elektromobilität antreiben. acatech - Deutsche Akademie der Technikwissenschaften (2020)
18. Second Life oder Recycling? BattMAN rettet Batterien vor ungerecht frühem Lebensende!. https://www.audi-mediacenter.com:443/de/pressemitteilungen/second-life-oder-recycling-battman-rettet-batterien-vor-ungerecht-fruehem-lebensende-14304. Accessed 7 June 2022
19. Schäfer, J., Singer, R., Hofmann, J., Fleischer, J.: Challenges and solutions of automated disassembly and condition-based remanufacturing of lithium-ion battery modules for a circular economy. Procedia Manuf. 43, 614–619 (2020)
20. Reich, U.: Freie Fahrt für den Wärmetransport, https://www.fst.com/de/corporate/newsroom/pressemitteilungen/2020/freudenberg-thermoconelastomer-presse/. Accessed 7 June 2022

21. Kampker, A., Heimes, H.H., Ordung, M., Lienemann, C., Hollah, A., Sarovic, N.: Evaluation of a remanufacturing for lithium ion batteries from electric cars. Int. J. Mech. Mechatron. Eng. **10**, 1929–1935 (2016)
22. Das, A., Li, D., Williams, D., Greenwood, D.: Joining technologies for automotive battery systems manufacturing. World Electr. Veh. J. **9** (2018)

Author Index