UNIVERSITEIT
iYUNIVESITHI
STELLENBOSCH
UNIVERSITY

100
1918 · 2018

INTERNATIONAL CONFERENCE ON COMPETITIVE MANUFACTURING

COMA'19

Proceedings

Knowledge Valorisation in the Age of Digitalization




30 January 2019 – 1 February 2019
Stellenbosch, South Africa

Organised by
Department of Industrial Engineering
Stellenbosch University

International Conference on Competitive Manufacturing (COMA 19) Proceedings

COMA '19

30 January 2019 – 1 February 2019
Stellenbosch, South Africa

Organised by the
Department of Industrial Engineering
Stellenbosch University

Editors:
Prof Dimiter Dimitrov[1]
Devon Hagedorn-Hansen[2]
Konrad von Leipzig[3]

[1, 2, 3]Department of Industrial Engineering
Stellenbosch University

ISBN No: 978-0-7972-1779-9

Department of Industrial Engineering
Stellenbosch University
Private Bag X1
Matieland 7600
Stellenbosch, South Africa
Tel: +27 21 808 4234
Fax: +27 21 808 4245
E-mail: devonh@sun.ac.za; kvl@sun.ac.za

About CIRP

CIRP was founded in 1951 with the aim to address scientifically, through international co-operation, issues related to modern production science and technology. The **International Academy of Production Engineering** takes its abbreviated name from the French acronym of **C**ollege **I**nternational pour la **R**echerche en **P**roductique (**CIRP**) and includes some 600 members from 50 countries. The number of members is intentionally kept limited, so as to facilitate informal scientific information exchange and personal contacts.

CIRP aims, in general, at:

- Promoting scientific research, related to
  - manufacturing processes,
  - production equipment and automation,
  - manufacturing systems and
  - product design and manufacturing
- Promoting cooperative research among the members of the Academy and creating opportunities for informal contacts among CIRP members at large
- Promoting the industrial application of the fundamental research work and simultaneously receiving feedback from industry, related to industrial needs and their evolution.

CIRP has its headquarters in Paris, staffed by permanent personnel and welcomes potential corporate members and interested parties in CIRP publication and activities in general.

CIRP Office, 9 rue Mayran, 75009 Paris, France. Web : http://www.cirp.net

In Loving Memory of
Professor Dimiter Marinov Dimitrov
The driving force behind COMA

Foreword

Welcome to this seventh International Conference on Competitive Manufacturing hosted by the University of Stellenbosch and organised by the Department of Industrial Engineering.

In a small world where global trade is the new driving force conquering countries and continents alike, international competitiveness is becoming the ultimate challenge of the new millennium. It requires high quality products manufactured with state-of-the-art technologies at low cost under the assumption of highly efficient operations management as well as clear corporate goals and strategy. This in turn is facilitated by and dependent on improved engineering training, education, and relevant applied research, fuelled by active interaction between academia and industry.

The main objective of COMA ’19, the International Conference on **Co**mpetitive **Ma**nufacturing is to present recent developments, research results and industrial experience accelerating improvement of competitiveness in the field of manufacturing. The 90 papers and presentations invited or selected to be delivered at the Conference, deal with wide aspects related to product design and realisation, production technologies and systems, operations management as well as enterprise design and integration. The worldwide participation and range of topics covered indicate that the Conference is truly a significant meeting of people striving similar aims. The event is an additional opportunity for communication between paper authors and attendees, which undoubtedly will serve as a further step towards exciting developments in the future. It also provides ample opportunities to further exploit international collaboration.

The Chairmen and the Organising Committee express heartfelt thanks and gratitude to the Members of the International Programme Committee, who have given their help and expertise in refereeing the papers and will chair the plenary and technical sessions during the Conference, as well as to the authors for participating and ensuring that the high standards required on an International Conference were maintained. These thanks and gratitude are extended to our highly regarded plenary speakers.

The Chairmen convey sincere thanks to the conference sponsors for their generous support, which made this event possible.

The International Academy of Production Engineering (CIRP) is gratefully acknowledged for the scientific sponsorship given to the Conference.

Finally, the tremendous effort of the Organising Committee is appreciated. Grateful thanks are due particularly to the Conference Secretariat for ensuring the success of COMA ’19.

We hope that you will find the Conference interesting and stimulating!

Mr KH von Leipzig
Conference Chair

Prof DM Dimitrov
Conference Co-Chair

Conference Chair
Von Leipzig, K - Stellenbosch University, South Africa

Co-Chair
Dimitrov, D - Stellenbosch University, South Africa
Schutte, C - Stellenbosch University, South Africa

International Programme Committee
Bártolo PJ, The University of Manchester, United Kingdom
Basson AH, Stellenbosch University, South Africa
Blaine D, Stellenbosch University, South Africa
Byrne G, University College Dublin, Ireland
Butala P, University of Ljubljana, Slovenia
Damm O, LHA Management Consultants, South Africa
du Preez ND, Indutech, Stellenbosch, South Africa
Eversheim W, RWTH Aachen, Germany
Fourie CJ, Stellenbosch University, South Africa
Gwangwava N, BIUST, Botswana
Hummel V, Reutlingen University, Germany
Kals HJJ, University of Twente, The Netherlands
Klocke F, RWTH Aachen, Germany
Kochan D, ZAFT-HTW Dresden, Germany
Kuppuswamy R, University of Cape Town, South Africa
Laubscher R, University of Johannesburg, South Africa
Levy GN, TTA, Switzerland
Lutters E, University of Twente, The Netherlands
Matope S, Stellenbosch University, South Africa
Mbohwa C, University of Johannesburg, South Africa
Mhlanga S, NUST, Zimbabwe
Monostori L, MTA Sztaki, Hungary
Niemann J, Hochschule Düsseldorf, Germany
Olakanmi E, BIUST, Botswana
Oosthuizen GA, Stellenbosch University, South Africa
Putz M, Fraunhofer IWU, Chemnitz, Germany
Schreve K, Stellenbosch University, South Africa
Stich V, RWTH, Aachen, Germany
Teti R, University of Napoli, Italy
van der Merwe A, Stellenbosch University, South Africa
van Dyk W, Stellenbosch University, South Africa
van Houten FJAM, University of Twente, The Netherlands
Wertheim R, Fraunhofer IWU, Chemnitz, Germany
Wegener K, ETH Zurich, Switzerland
Yadroitsev I, Central University of Technology, South Africa

Organizing Committee
Damm O, Stellenbosch University, South Africa
Hagedorn-Hansen D, Stellenbosch University, South Africa
Matope S, Stellenbosch University, South Africa
Oosthuizen GA, Stellenbosch University, South Africa
van der Merwe A, Stellenbosch University, South Africa

SUBMISSION REVIEW PROCESS

A formal "Call for papers" for the 7th International Conference on Competitive Manufacturing (COMA '19) was issued in May 2018 to submit an 'Abstract' within the identified tracks/themes. Abstract submissions were subjected to an internal reviewing process, whereby successful submissions were notified and invited for presentation to the conference. Authors were subsequently invited to submit the 'Full Paper', which was published as a conference proceeding. Both the Abstracts and Full Papers were submitted online through the Sun Conferences submission page http://conferences.sun.ac.za/index.php/COMA19/coma19 where acknowledgement of receipt was sent to authors. Authors were informed that a double-blind review process is applied to Full Paper submissions.

The following dates were set by the organising committee:

- Call for papers ....…………..1st May 2018
- Submission of abstracts: …………………………… 30th June 2018
- Notification of acceptance of abstracts: ………….. 15th July 2018
- Submission of full papers: ....17th September 2018
- Feedback on paper reviews: .... 30th November 2018
- Revised paper submissions: .... 3rd January 2019

Abstracts were required to be a maximum length of 400 words. Full Papers were required to be a maximum length of 6 pages, but leniency was given for the Author biography's and references. Full Paper submissions were required to adhere to a specific template and format which was placed on the conference site here:
http://conferences.sun.ac.za/index.php/COMA19/coma19/schedConf/cfp

A double-blind reviewing process was used for the Full Paper submissions. As such, both the reviewer and author identities are concealed from the reviewers, and vice versa, throughout the review process. Each Full Paper submission was sent to a minimum of two reviewers, with a third reviewer being requested in case of non-consensus between the first two reviewers. The reviews were completed by national and international academics, and experts in the respective field, listed on the International Programme Committee page.

A total of 35 reviewers participated in the review process, each reviewing between two and five papers. Reviewers were asked to review submissions according to the following criteria and were encouraged to provide recommendations and suggestions.

- Does the title reflect the contents of the paper?
- Does the paper relate to what has already been written in the field?
- Do you deem the paper to be proof of thorough research and knowledge of the most recent literature in the field of study?
- Is the paper clearly structured, easy to read and with a logical flow of thought?
- Are the arguments employed valid and supported by the evidence presented?
- Are the conclusions clear and valid?
- Does the paper conform to accepted standards of language and style?
- Any other recommendation(s)?
- Select reviewer recommendation: 'Accept Submission', 'Revision Required', or 'Decline Submission'.

Reviewer feedback was saved on the submission system, where acceptance emails together with review comments were sent to the authors, allowing them to revise the submission. The authors were given between 2 and 4 weeks to incorporate changes, after which the final document was submitted for approval and publication as a conference proceeding.

Papers were invited in the following areas relevant to the conference theme:

Product Design and Realisation: Design for Manufacturing and Assembly, Reverse Engineering, CAD/CAE, Concurrent Engineering, Design for Additive Manufacturing, Biologically Inspired Design Approaches, Virtual Prototyping, Networks in Product Development, Open Design.

Production Technologies: Expert Systems in Manufacturing, CAD/CAM Systems, HSC, EDM, Forming, Casting, Metrology, Mechatronics, Precision Manufacturing, Bio-manufacturing, Robotics, Reliability, Sensing, Assembly, Automation, Intelligent Manufacturing, Software for Manufacturing, Biologically inspired manufacturing processes, Non-conventional machining, Environmental aspects, Machining of non-metallic materials and composites, Abrasive processes, Hybrid Machining

Production Systems and Organisations: Factory Planning, Production Planning and Control, Logistics, Modelling and Simulation, Scheduling, Quality Assurance, ERP-Applications, Supply Chain Management, Communication Networks, Social Manufacturing, Digital Factory, Biological Transformation in Production Systems and Supply Chains, Cyber-physical approaches

Enterprise Design and Integration: Knowledge Management, Product Life Cycle, Human Interface, Web-based Design and Manufacturing, Technology and Innovation Management, Total Quality Management, Distributed Control Systems, Socio-economic and Environmental Issues, Enterprise Engineering, Industry 4.0, Machine Learning, Big Data.

ACKNOWLEDGEMENTS

Sincere thanks to our distinguished supporters and sponsors, whose generosity made the success of this Conference possible.

TABLE OF CONTENTS

Session B5: Processing of Advanced Materials I

Session B6: Novel Manufacturing Concepts

Session B7: Learning Factories and Education

Session C3: Enterprise Engineering Tools I

Session C4: Modelling and Simulation II

Session C5: Enterprise Engineering Tools II

Session C6: Product Design and Graph-Based Languages

Plenary Session

Manufacturing Technology: Concepts, Adoption and Commercialisation

New Machining Concepts - Compensated, Intelligent, Bioinspired

K. Wegener[1],T. Gittler[2], L. Weiss[2]
[1]Institute of Machine Tools and Manufacturing (IWF), ETH Zürich, Switzerland
[2]Inspire AG, Zurich, Switzerland

Abstract
The impact of Industrie 4.0 onto machine tools is significant, despite the fact, that quite some of the novelties discussed within this new paradigm have their roots decades earlier. But especially the concerted action, which strives the development of sensors, controls, data processing together with connectivity, unprecedented data integration and the notion of cyber physical production systems open up new development lines towards manufacturing systems as enablers for the progress in manufacturing. Highly developed compensation concepts are developing into state depending AI-supported strategies. Maintenance becomes predictive, as learning of machines becomes global and model based. Further inspirations taken from biological systems are adopted for machining centres and drive a biological transformation of manufacturing machines. Machine intelligence becomes the basis for executing manufacturing processes, which requires a close integration of process intelligence (CAM-systems) and machine controls.

Keywords
Biologicalization, Industrie 4.0, Cognitive Technical Systems, Machine learning Strategies, Bioinspired, Biointelligent Manufacturing, Adaptive Thermal Compensation, Self-Organization, Self-Healing, Sensors in Abundance

1. INTRODUCTION

In the past 10 years it became obvious, that manufacturing provides the basis for a stable society. 20% of GDP as industrial production or 25% as production of tangible goods seems to be the basis of wealth even in a service oriented society. In the end, all thoughts of engineers have one unique dedication, namely to provide products to fulfill requirements of mankind. Therefore, it is not only money, that makes the world go round, but production machinery. The modern world is populated by about a new value of 350 billion $ of CNC driven machine tools, which is approximately 2.5 million CNC machines. As shown in Figure 1 each year a value of 50 to 60 million CNC machines is added, still increasing the population of CNC machines, because some of them replace non-counted non-CNC machines. Approximately 2/3 of them are cutting machines.

In 1951, the CNC control has already been invented, being one of the most groundbreaking developments in production machines, but the transition is still on its way. According to [47] from 2018 to 2025 the population of CNC machines in Europe will increase from 700'000 to 800'000, while non CNC-machines will strongly decrease in number. Since then the overall increase in productivity, which according to [58] is approximately 2 to 3% per year is basically supported by an incredible amount of ameliorations and better exploitation of the technology. Just as an example: a tool segment for the tire industry produced on one of the old punchtape based machines would require 1000 kg of punchtape, which moves this production clearly in the vicinity of technical overkill at that times. Figure 2 shows for a gear grinding machine the overall increase in productivity over the years, driven by enhanced process technology, enhanced tools, machine tool development being capable to withstand higher forces and vibration excitation, faster controls, more powerful drives, sensory, and better automation. Naturally also set point generation has largely enabled higher cutting speeds and it can be estimated that in case of more reliable cutting materials and better tool benign tool paths a reduction of production time in the two-digit percent range can be achieved. Within CIM in the 80s and 90s an attempt was made to automatize the strategy planning, an intelligence or experience requiring act within the process chain of cutting. This was not successful due to limited computational power and economic interests. In the meantime, additive manufacturing still in the stage of rapid prototyping showed the possibility to start manufacturing only a few minutes after the workpiece geometry has been provided, which takes hours in case of cutting. To face the truth, produced parts in rapid prototyping lack several industrial requirements indispensable in the mature cutting technology. But the discussion on the notion of a "five-minute-machine", requiring machine intelligence has been implanted.

Compensation, as kinematic, thermal gravitational or dynamic compensation is one of the achievements in recent times towards accuracy, reliability and robustness. The physical modelling behind are highly complicated, if they take into account all the essential influential effects within a machine tool. Models in use and scientific discussion spread from physical to phenomenological approaches, and in this range from complex and nearly unmasterable to simple while vice versa the parameter identification effort ranges from simple to awfully huge. Self-learning approaches based on some model seem to be the suitable work around for dilemmas like this.

Obviously, the human ability to take up pieces of information, store and combine them and take decisions on the basis of information from different times and environments is valuable in those cases of complicated and even complex behavior. Also other capabilities of biological systems besides cognition, learning capabilities and intelligence are highly desirable for machine tools and deserve research, such as machine autonomy, abundance of sensors, sensor fusion and redundancy, swarm intelligence and fleet learning, strong reliance optical image recognition, self-healing or wear and failure compensation and also teaching competence. Bio inspired structures are topics already introduced at least on the scientific level into machine tools as demonstrated in [41], which is insofar a fairly weak approach, as the condition of manufacturability needs to be taken into consideration, which today is only done by intuition of some designer. But for sufficient development of additive manufacturing for large structures a really topology optimized lightweight structure can be realized. Bio inspired functional surfaces for machine tools suffer from the fact, that these structures are extremely vulnerable, and their lifetime is strongly limited. Biological systems always come up with self-healing capabilities, restoring the fine surface structures. Therefore, the capability for self-healing opens new potentials in very different aspects from tool sharpening to play filling but is still in its infancy.

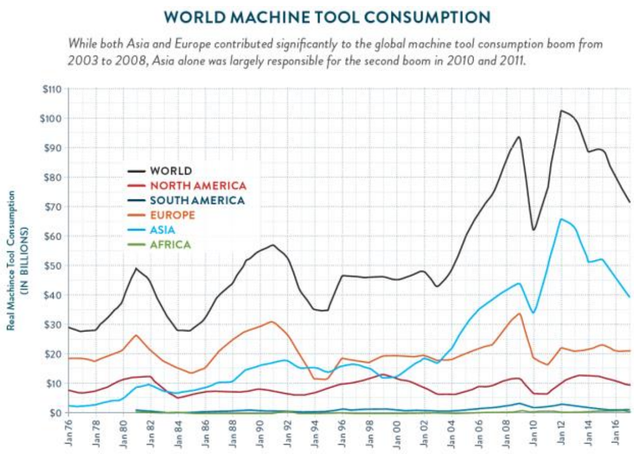


Figure 1 - Increase of the machine tool population over time [60]

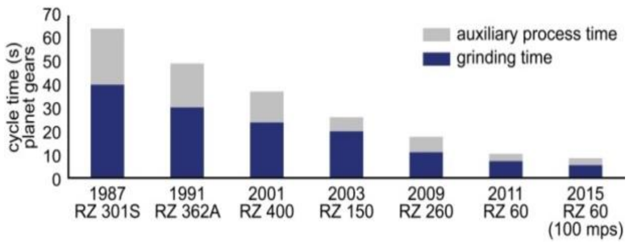


Figure 2 - Reduction of manufacturing times (Courtesy Reishauer) [56]

2. COMPENSATION

2.1. General remarks

Driven by accuracy requirements of machine tool applications the expenses for the mechanical improvements of machines by mechanical means are becoming more and more expensive. This requires a paradigm change which was attempted in several waves starting decades ago, namely to charge the control system with compensating for erroneous movements of the machine. These are always due to the fact, that positioning, and position measurement of machine tool axes takes place far away from the TCP which shall be controlled and between those two the machine mechanics with its physically given insufficiency reigns. Several aspects today justify independent of earlier failures a reconsideration of this topic:

1.) A consequent design for compensation declaring repeatability as the highest principle has not yet been attempted [57].
2.) IT-technology of today offers drastically increased computational performance, especially if intensive number crunching required by the modelling approach is performed on the GPU (Graphical Processor Unit). Thus real time delivery of new compensation values becomes achievable.
3.) Huge investments of manpower have enriched the world with model order reduction algorithms of high and predictable precision as for instance demonstrated by [49].
4.) Meta-modelling allows fast provision of new compensation values. Meta models also can be adapted on the fly to cope with changing machine and machining conditions. They are a preferred entry point of self-learning algorithms.

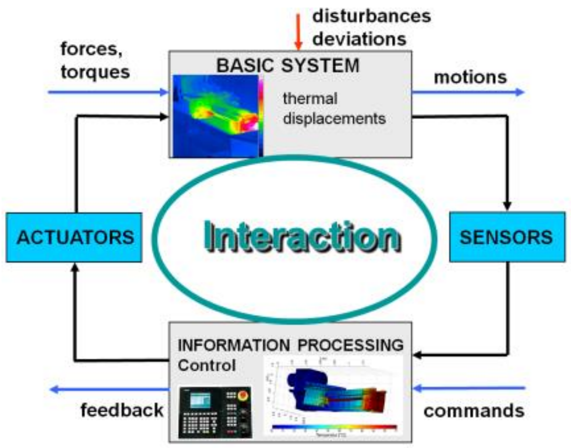
Compensation is done according to [57] to correct erroneous behaviour of the machine tool out of different reasons:

- Inaccurate kinematics
- Gravity of moving parts
- Dynamical deflection due to inertial effects as intalk, crosstalk, coupling forces

- Thermal displacements
- Wear status

While dynamical compensation needs extremely fast reaction or feed forward with the help of models, the other compensations are not requiring that high dynamics or are even static. Compensation requires models of the machine tool and / or the process, in order to take effects into account, that happen during the long-life time of a machine, those models become extremely complicated and complex, self-adaptability and machine learning become of crucial importance as will be shown in the following sections.

2.2. Thermal compensation



Sensor technology	Modelling	Actuator Technology
NC-program	Physical field equations	Cooling, heating
Power intake of components	Reduced physical field equations	Axes of machine
Temperature measurements	Thermobalance models	Additional axes
Elongation measurements	Phenomenological models	
Position measurements	KI approach, neural networks	

Figure 3 - Mechatronic concept of thermal compensation and morphological box [57].

As stated by Mayr *et al.* [35], thermal errors are by far the most prominent source of manufacturing errors, and because thermal compensation has less time restrictions and thus really becomes industrially viable, it is elaborated to more detail. The standard means to cope with thermal elongations in high precision machining is to create a thermally stable environment by air conditioning of the shop floor and by run in to get a stable distortion or tempering the machine to mitigate thermal gradients. The energy consumption of machine tools is under discussion for over one decade, and becomes more and more suspiciously observed. With the aspect of knowledge based machining the modern approach is to go for a thermal compensation, where the machine tool axes execute the error corrections without any utilization of further power as in the cooling and air conditioning approach. Thus, not only from the accuracy perspectives, but also from the energy efficiency perspective, there is no way around thermal compensation. Figure 3 shows different ways how to setup a thermal compensation action chain. Knowledge basis is the model which takes up data from the basic system by sensors and changes with the basic system with the help of actuators. Everything that supplies information to the compensation intelligence is considered as sensor. This can be displacement measurements, thermal measurements and also the CNC program is considered as a sensor. Actors can be a chiller, the machine axes as well as auxiliary positioning devices. From top to bottom the models become less physical, which on practical basis means that its modelling complexity as well as the computation time is reduced to become more and more real time capable. This needs to be then counterbalanced by measured parameters displaying the complexity of the system and thus the calibration effort needed to setup the model in the beginning is increasing from top to bottom.

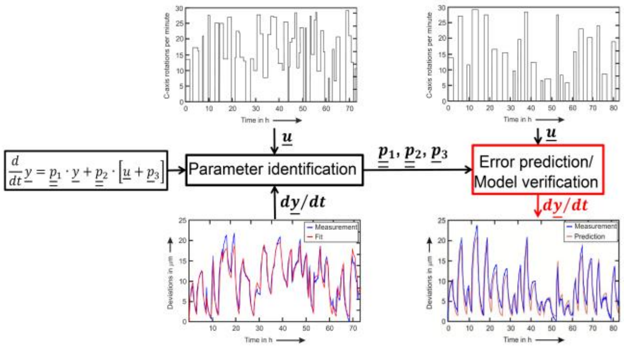


Figure 4 - Procedure of phenomenological modelling, calibration and compensation [57].

Figure 4 displays the procedure for phenomenological modelling with sensor input from the CNC program. As basic system the rotary axes block of a 5-axis milling machine is used. As phenomenological model a fairly crude representation of the physics of thermal behaviour is used, keeping only the history dependency of errors on the movements of the machine as physically correct representation. Similar modelling has been presented e.g. by Mayr *et al.* [34], Yang and Ni [64] and Brecher *et al.* [7]. This gives a system of first order differential equations as shown in Figure 4 for the location error vector $\underline{y}$ of the rotation axes depending on the vector of thermal inputs $\underline{u}$ which contains the power inputs of the axes, environmental temperatures etc.. All the elements of the matrices $\underline{p_1}$, $\underline{p_2}$ and vector $\underline{p_3}$ need to be then identified during a calibration cycle.

This calibration cycle consists of 70 hours of randomly chosen movements of B- and C- axis. The displacement ZOT of the table surface in vertical direction as an example for one of the components of the vector $\underline{y}$ is shown in Figure 4 as simulation result together with the measuring results, which show expectedly good accordance, because for this motion the parameters have been identified.

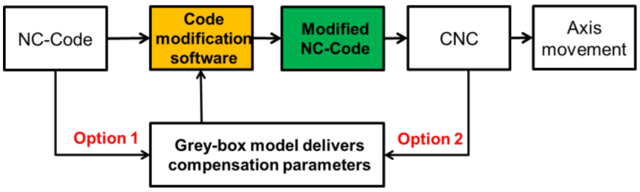


Figure 5 - Block diagram for thermal compensation [19]

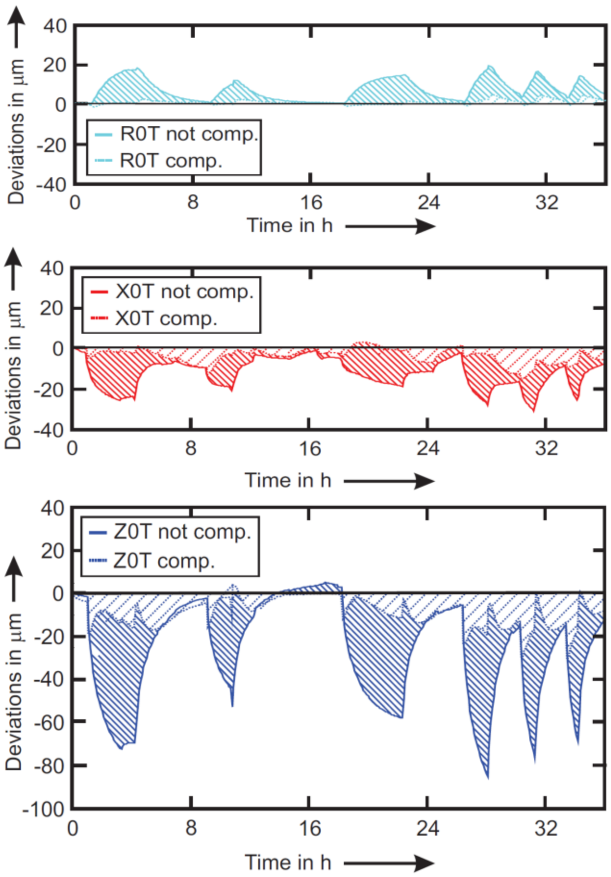


Figure 6 - Achieved compensation results with the phenomenological model [57]

To verify the model to more than only suitability a different randomly chosen motion profile for B and C is evaluated with the model and measured as well. Again, the model represents very well the measured behaviour of the machine and the simulated error values can then be used for compensation according to Figure 5. Figure 6 then shows the achieved compensations, which demonstrates that this kind of modelling is clearly capable to serve as a model for online compensation of machine tools.

The scientific questions still open are, how short a calibration cycle might be to achieve a given quality of compensation and what happens for very long run times or changes in the mode of utilization of the machine. For the latter, a solution is provided in chapter **Error! Reference source not found.**.

3. AUTONOMOUS MACHINE LEARNING

3.1. General remarks

Learning is a fundamental element of Intelligent Manufacturing Systems (IMS), and closely linked to the notion of Biointelligence. Learning processes in general can be classified in three categories, depending on the available feedback: Supervised learning, reinforcement learning and unsupervised learning.

Supervised learning postulates that the desired outcome (solution) is available together with the respective data. Reinforcement learning provides an evaluation of the solution given by the learning algorithm, usually of qualitative or binary nature. The unsupervised learning process does not provide any (external) solution [37]. The solution lies within the data, usually referred to as a structure. It is usually used to cluster and trim datasets, such as automated outlier detection for instance. The notions of other learning types, such as fleet learning, deep learning and model based learning are specific applications of supervised, unsupervised or reinforcement learning. Hybrid approaches of the previously named approaches are possible, and practically applied in many of the presented cases (e.g. [6,31]). Fleet learning is a term derived from condition monitoring applications in aircrafts, and has also found applications in automotive autopilot training [51]. However, it has not yet been published in conjunction with manufacturing related approaches. Model-based learning defines any learning approach relying on pre-installed knowledge (pre-taught patterns, inclusion of expert knowledge), showing a larger and faster learning success. It can be considered as a Funnel of Nuremburg for Machines. Internal research at IWF (ETH Zurich) hints that model-based learning is also a useful measure to compensate for low-quality low-quantity historical data for the training of appropriate models. As a complete overview over all machine-learning techniques goes beyond the scope of this report, the reader's attention should be drawn to [62] for an in-detail overview of machine learning approaches in manufacturing and their particularities.

3.1.1. Learning with structured and reliable *data*

Learning approaches are mostly characterized by the quality of the relevant and available data. A plethora of high quality datasets registered over appropriate timespans is a rather rare case, wherefore few successful applications of unsupervised learning approaches in manufacturing have been published. One of the few exceptions that provided an application for unsupervised learning algorithms in manufacturing was published by Lieber *et al.* [30] in 2013, using unsupervised learning for preprocessing and feature extraction purposes of datasets generated in a rolling mill. Luckow *et al.* [31] described in 2016 the possible application of unsupervised learning for visual inspection processes in manufacturing, as well as for recognizing features in camera images of robots enabling self-learning techniques. For the visual inspection process, a large database of already-existing imagery was used in order to train the algorithm, therefore circumventing the issue of collecting the data manually. However, an important aspect here is the notion of transferring learning from one case to another, which is also applied by [40] and [45]. In [45], the fusion of knowledge of manufacturing operations data is carried out by representation learning. Representation learning describes a learning method in which the feature set is auto-discovered by the learning algorithm. The feature engineering or extraction is carried out automatically.

Even if data becomes more and readily available, and processing tools and units decrease in cost while they increase in performance, the collection of the right data within a sufficiently defined context at the right time in a sufficiently large quantity remains an issue [31]. As for the design of learning approaches, this challenge is not always easy to overcome. Reinhart and Steil [44] go as far as suggesting hybrid approaches for inverse kinematic controls of robots. According to their findings, parameter identification remains less cumbersome as the data collection for a pure machine-learning approach. The machine learning approach is used to model only complex characteristics outside of the identified parametric model. The combination of physical and statistical modelling however is not an isolated case, as underlying physical models can be used to decrease the need of generated data for proper results. Together with the transferal of already acquired knowledge in different applications, these hybrid approaches appear to be a valid approach for a learning approach in manufacturing, given a lack of available or accessible data. As some sort of physical knowledge is always used at the foundation, most approaches can be considered as hybrid. However, the degree of hybridity differs severely between most known approaches of learning strategies in manufacturing.

Transfer learning in a direct manufacturing context seems to have not been published yet. However, it is widely used in image recognition in conjunction with Convolutional Neural Networks (CNN) [43]. As CNNs train different layers for subsets of image recognition (e.g. identification of vertices and edges), layers can be reused for different purposes, considerably lowering the necessary training data set volume. The underlying principle of data clustering and analysis remaining the same, only the mapping to predefined objects or states (e.g. conform or non-conform pieces) needs to be trained when applied to manufacturing systems.

The supervised learning approach is by far more frequently used than the non-supervised learning approach. Denkena *et al.* [11] implemented a support vector machine approach in 2016 in order to create a self-optimizing cutting process. The obtained process data is continuously modeled, in order to optimize the cutting parameters for the satisfaction of predefined boundary conditions. Wuest *et al.* [61] used a supervised machine learning approach in 2014 for product quality monitoring. Product and process states were introduced in order to partition the manufacturing program into sequences, allowing for a standardized data collection. This ultimately enables a “collection of data across a wide spectrum of product and process information”. In 2016, Haas *et al.* [23] proposed an iterative learning control approach to reduce systematic tracking errors in machine tools. In 2017, Arinez *et al.* [2] demonstrated the use of reinforcement learning in order to train a gantry scheduling policy, allocating material to a buffer serving two machines at once. The trained manufacturing cell showed a significantly lower production loss than the previously installed first-come-first-serve approach. Escobar and Morales-Menendez [13] published a machine learning and pattern recognition technique in 2017, allowing to detect defective welds from an ultrasonic metal welding process. Another interesting application of supervised machine learning in manufacturing was introduced by Shin *et al.* [50] in 2014: By the combination of STEP-NC features with NC data extracted via MT Connect, a model was trained which could predict the power consumption of machine tool. However, the performance was rather low, with a prediction error in between 11% and 21%, depending on the material used. The here within presented example of the autonomous and adaptive thermal error compensation by Blaser *et al.* [6] accordingly falls under the definition of supervised learning with structured data. Summing up the use of the so-far discussed machine learning methods, the limits of the application seem to lie in the challenge of an automatized collection of structured and reliable data.

3.1.2. Learning with unstructured, imprecise, fuzzy or probabilistic data

However, cognitive systems should be able to learn based on uncertain or probabilistic data. Additionally, associative learning (transferring knowledge from one case to a different setting) as a method of creation of new knowledge, can be considered a trait of advanced cognitive systems. Moreover, learning and adaptation based on experts' advice, as well as learning on rather gray and fuzzy instructions by a worker for instance, are also learning strategies that should be considered.

An example for the learning with unstructured data is the creation of a clustering or standardization framework: In 2014, Wuest *et al.* [61] designed specific conventions, in order to allow for a standardized learning cycle, yielding favorable results. They defined a set of characteristic product states, allowing to measure and cluster data, as well as detecting and classifying changes in product conditions.

In the context of learning with unstructured or imprecise instructions, the notion of fuzzy control, or fuzzy learning comes to mind. This describes the ability of the machine to properly interpret a qualitative input, usually by a human interaction. [13,37,61,62] made use of, referred to or mentioned fuzzy control in their respective publications.

Telling from the variety of different learning approaches used in the publications cited in this chapter, the absence of robust, efficient general learning approaches becomes obvious. There currently is not a one-size-fits-all measure, but it seems as if learning approaches are surprisingly diverse. As the handling of system changes over time, a current and certainly future focus is the (self-)adaptation of algorithms, as presented for instance by Blaser *et al.* [6] in chapter 3.2.

An open question waiting for answers is the optimality between modeling and learning effort. It might differ considerably with the application cases (and with available data quality and quantity), but a general framework is yet to be developed. This is also interesting in the context of using similar models in different situations, where the results need to be scaled for comparison (i.e. application of similar learning approaches to different machine types). As internal research from IWF (ETH Zurich) hints, the modeling effort rises considerably with the application of similar learning strategies to even slightly differing machine types.

With the increase of high volume high quality data, the modeling effort decreases. However, a main problem remains to acquire sufficient and reliable data, especially for unsupervised learning approaches. Furthermore, the utilization of data from other machine types in transfer learning cases lacks model applications and guidance. This issue hints at a further development of fleet learning as used by Tesla Motors in order to answer the question of comparability between data from different machine sizes and types. Summing up the previous argument, the notions related to associative learning (transfer learning, fleet learning) should be elaborated more in-detail in manufacturing contexts.

Alternatively, successful and universal approaches of introducing unreliable and unstructured data are of interest. Beneficially might be to pursue the question of registration and evaluation of each and every input of a skilled and virtuous user, so that implicit knowledge can be made visible. It becomes evident that the combination of rule and model based expert systems with learning capability enhances manufacturing processes.

3.2. Adaptive learning control for thermal compensation

As pointed out in section 2.2 models for machine tool covering their full behaviour become infinitely complicated. Reduced models combined with machine learning is therefore here considered as superior. Especially the integration of history dependent power input data, the modelling of time delays introduces errors on the long run of a machine. Mou and Liu [38] pointed out, that working conditions might change such, that the identified parameters of the simplified phenomenological model are no more suitable. Starting with the model explained in section 2.2 a self-learning algorithm, developed by Blaser *et al.* [6], as adaptive learning control (ALC) can be established as sketched in Figure 7. This method requires an active thermal displacement modelling, which predicts the thermal displacements, transfers the correction data to the machine control via some provided interface as FOCAS2 of Fanuc, which is then used for correcting each individual axis. From time to time measurements of real data from the machine take place to compare the predicted displacements with the actual displacements on the machine. For this compensation scheme only providing the relationship between displacements and heat inputs, no other than position measurements can be used for this, which means that a defined geometrical object needs to be placed on the work piece side and a probing system on the spindle system is used to determine the position of this geometrical object. As most machine tools today provide an exchangeable probing system, this is used to probe a precision sphere positioned 160 mm outside of the center of the C-axis in 4 different positions of the C-axis as also shown in Figure 7. These measurement values are then used to recalibrate the parameters within the matrices and vectors of the model equation.

The principle of mitigating thermal errors is then presented in Figure 8.

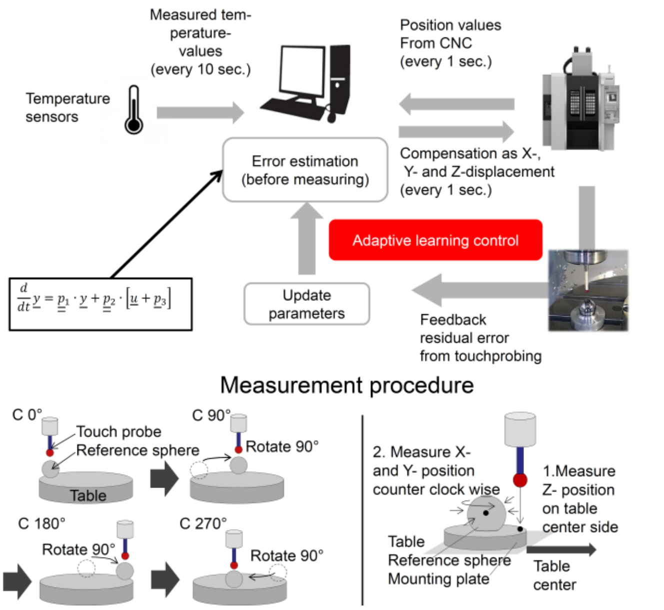


Figure 7 - Principle of adaptive learning control for thermal error mitigation in machine tools in the bottom is shown the measurement procedure [6].

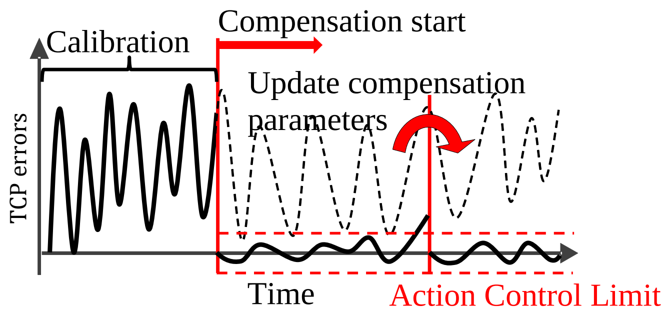


Figure 8 - Illustration of the concept of ALC for thermal error compensation, adapted from Blaser *et al.* [6]

The effect of ALC is shown in Figure 10 for the machine tool with V [w C2' B' b [Y1 Y2] X [Z1 Z2] (C1) t] as kinematic chain according to ISO 10791-2 shown in Figure 9. The first calibration is set to a reduced time interval of 12 h, which is typical for run in times at the machine tool maker. The correlation between achieved location errors and run in times is shown in Figure 11 for a situation with thermally stabilizing cutting fluid and without. It can be recognized, that depending on the cutting fluid supply the calibration times exceed these 12 h strongly. Despite 12 h seems to be too short for a good calibration, it is applied here to demonstrate the effect of ALC. Here the thresholds for recalibration are set to 15 µm for linear and 25 µm/m for rotational axes. The system starts with 30 min between each recalibration. When these thresholds are surpassed, the recalibration times are automatically reduced to 15 min and a time delay of two hours is set for the application of new data, to have sufficient measuring points for the recalibration. Figure 10 shows, that the ALC approach is suitable for enabling compensation with limited accuracy models.

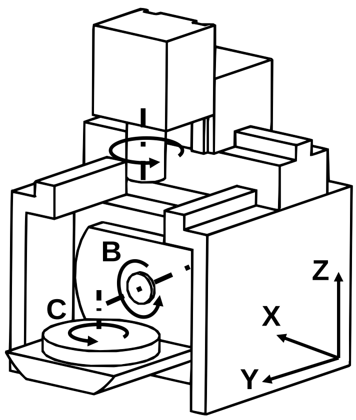


Figure 9 - Structure of the investigated machine tool

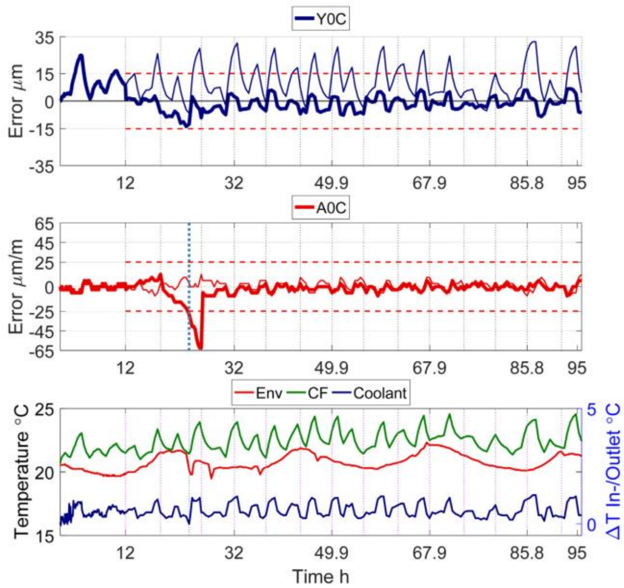


Figure 10 - Results for the thermal compensation with ALC approach

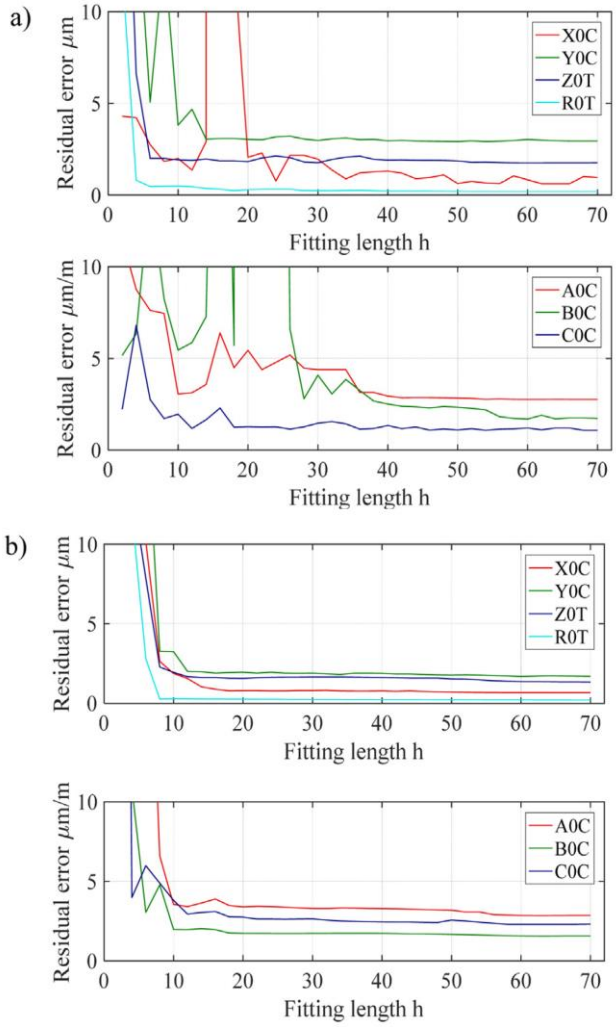


Figure 11 - Correlation between the calibration time and the achieved location errors. Without cutting fluid in a and with stabilizing cutting fluid in b [6]

Drawback of the method is the necessity to perform position measurements, which are time consuming as they cannot take place parallel to machining. Reconstruction of the model such that temperature measurements at specific points within the machine can be used instead is at the moment under preparation. More elaborate learning algorithms are developed to change the structure of the model if required. Deeper reconstruction of the model, even change of thermal sensor placement is possible with a machine learning approach, but anyway requires rules and algorithms for those modifications to be predefined.

3.3. Iterative learning control for path planning

The concept of iterative learning control can also be applied for the increase in accuracy of tool paths in machining, which was reported in [23]. Difficult for this task is the limited bandwidth of typical motion controllers, which therefore requires feedforward control with knowledge out of previous trials. Fast exploitable models per axis with limited physical exactness are used as shown in Figure 12.

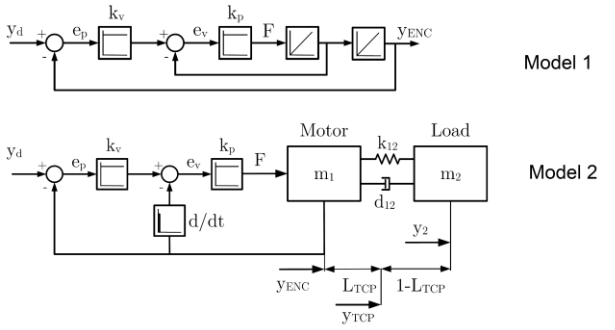


Figure 12 - Single axis models of limited physical accuracy used for estimation of the TCP position [22]

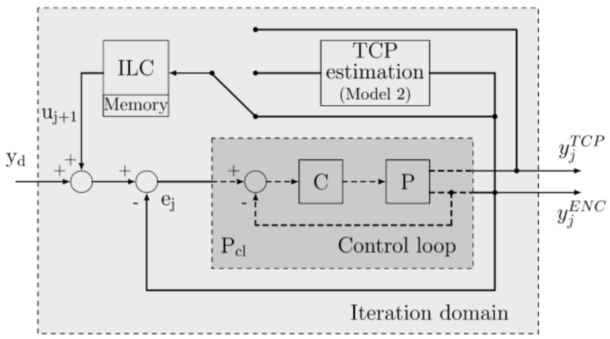


Figure 13 - Block diagram for the ILC for path planning containing three different feedback possibilities: (1) feedback from the encoder signal directly, (2) feedback from an estimation of the TCP out of model 2 and the encoder signal (3) feedback from direct measurement of the TCP with an external cross grid [22]

For the feedback either the encoder signals are exploited or an external measurement system directly tracing the TCP is used for the learning feedback. The latter possibility is not suitable for practical utilization in real manufacturing, but shows which accuracy increase is possible, if exact knowledge of the TCP movement were accessible. With the encoder signals an estimation of the TCP movements with the help of the two-mass model from Zirn [68] is used instead of a measurement at the TCP. Thus, the scheme of the ILC for path planning containing three different possibilities of setup of ILC is presented in Figure 13.

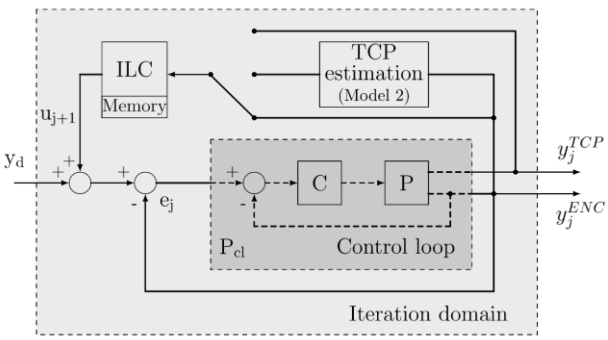


Figure 13Figure 14 - test bed for the validation of the ILC algorithms for path planning

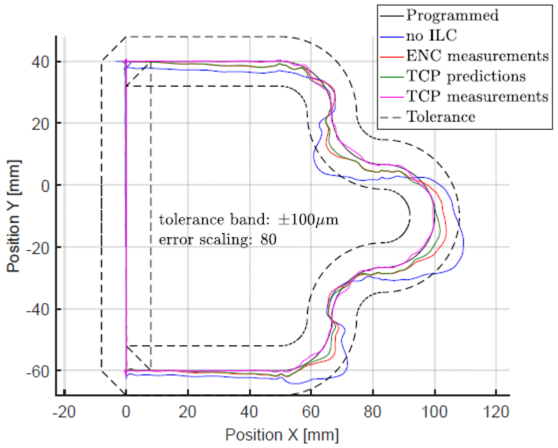


Figure 15 - contour accuracy of three different ILC schemes [22].

For the test bench from Figure 14 the results of the ILC for a chosen path and the three different learning approaches is presented in Figure 15.

Naturally the feedback of exact TCP data enhances this ILC best, but the simple models for the prediction of the TCP behaviour depending on the commanded paths show respectable increase of path accuracy and in this case, bring the TCP over the whole prescribed path within the tolerances.

Depending on the iteration scheme, only very little numbers of iterations are required until the final accuracy of the path, which makes this feasible for repetitive manufacturing tasks. A generalization to unknown manufacturing paths, where training for a feedforward is not possible, the learning algorithm is capable to learn frequently repeated sections or features of paths, which can then be stitched together to give an enhanced TCP path. This approach specifies a self-enhancing expert system for path planning on a specific machine, which can be setup without intense modelling of machine structures.

4. EXPERT SYSTEMS AND TEACHING MACHINE

Expert systems on machine tools comprise the available knowledge of how to run the process. The only way how “keep it simple” works in those cases is, to hide the full physical complexity of the process the user and let the expert system on the machine do the work. But nevertheless, for machine tools experienced users outperform any expert system as they have for special applications specialized knowledge and therefore need to have the access also to all details and parameter selection. Especially for strongly experience based processes as it is the case for grinding and EDM expert systems are offered within the machine, which is the basis for creating intelligent machines. The TechnologyIntegrated solution by Fritz Studer AG (Thun, Switzerland) [17] is an example. Here, an expert system was designed, in which previously digitalized knowledge is used to automatically assess a near-optimal parameter set for a grinding process. The worker is then instructed by the system on how to setup the process and the grinding parameters are automatically selected from the expert system. As shown in see Figure 16 the expert system ranked among the three fastest workers which showed a fulfillment of a 100% of all requirements. On the other hand, this demonstrates the headroom for improvements if the system is capable to learn from experienced users as shown in Figure 17. The machine becomes the main information storage and for this concept needs to have close contact to technological data bases, if these are not directly stored within the machine control. In this learning and teaching approach the machine directly behaves like a biological system. An interesting question in this respect is, how the machine control distinguishes between the apprentice and the master.

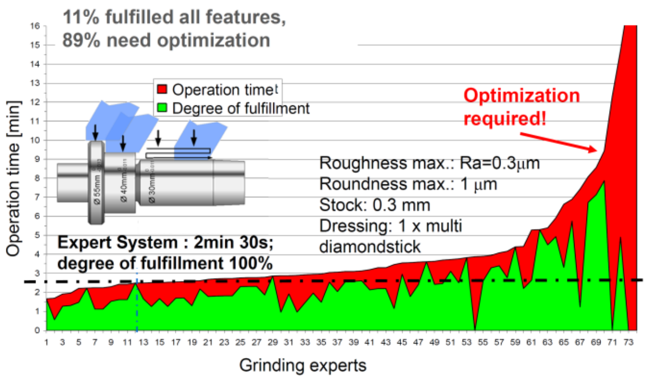


Figure 16 - Performance of the Studer Technology Integrated solution in comparison with the performance of human experts for a dedicated grinding process [18].

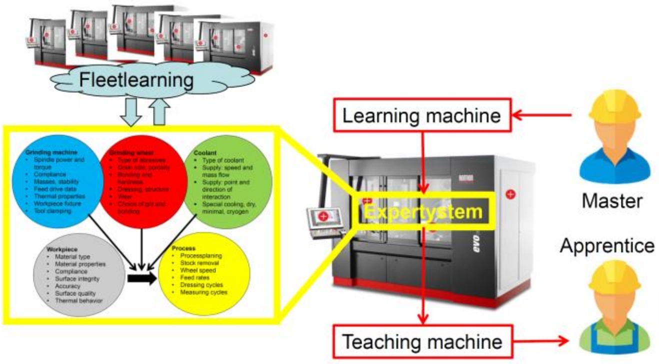


Figure 17 - Learning and Teaching Machine.

5. SELF-HEALING AND SELF-ORGANIZATION

An important trait in the biological transformation is the ability to handle exceptions, in which the system needs to adapt to disturbances, partial or complete breakdowns. By comparing the way how systems treat possible failures, a distinct line can be drawn in between self-organizing and self-healing systems:

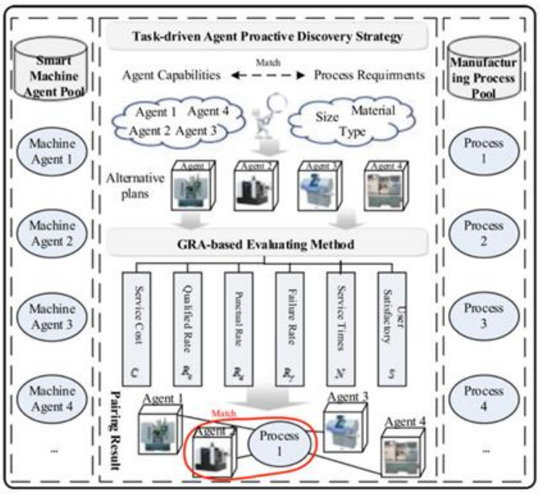


Figure 18 - Task-driven Self-Organizing Model [67]

Self-Organizing Systems: Systems that have a larger space of possible solutions (including a defined set of disturbances), which can swiftly switch to a different solution in case of a detected misbehavior or failure. An example would be a manufacturing system fulfilling an optimization problem, in which multiple solutions are optimal, and can be selected arbitrarily, such as in [67] and presented in Figure 18. The systems are defined as self-organizing systems, which are able to adapt to changing conditions by the selection of different available elements or different process sequences to be used. This does not include repair of the defective or misbehaving unit or fraction, but a mere mitigation of defects of member entities. An example referring to machine tools would be the autonomous detection of a precision loss incurring within a section of a ball screw, and the subsequent shift of the process window outside of the section suffering from precision loss. Self-Organizing systems can be characterized by

- Predefined disturbance and solution sets,
- problem circumvention (redundancy, diversity) rather than root cause elimination,
- the lack of ability to reconstitute an original, desirable state,
- and cognitive capabilities limited to detect symptoms belonging to the predefined solution space.

Self-Healing Systems: Systems that are able to reconfigure, identify and repair autonomously following a misbehavior, failure or error. Systems fulfilling the following characteristic traits can be defined as an actual self-healing system:

- A priori unbound disturbance and solution space,
- retention of a functional state not entirely dependent on redundancies,
- ability to reconstitute an original desirable state,
- distinct cognitive capabilities allowing to intrinsically or heuristically develop solutions to a previously undefined disturbance.

From software development, an analogous definition of necessary traits can be drawn. They include consistency-maintenance mechanisms, failure-detection techniques and recovery techniques [10,16]. Self-Healing Systems require a large array of sensors in order to allow for the necessary cognitive capabilities to detect disturbances and unwanted behavior. [21] gives an overview of appropriate monitoring approaches as enablers for self-healing methods in machine tools.

5.1. Self-Organization

Matt [33] introduced a theoretical framework in 2012 which relies on axiomatic design principles, and implements "[passive] self-healing mechanisms in agile production systems". The basic self-organizing strategies are described as redundancy, diversity and detection of system failures.

The modelling of the self-organizing energy harvester by Farnsworth & Tiwari in 2015 [15] is a prominent example of a self-organizing mechatronic system. It exhibits failure management systems allowing to regain functionality by redundancy for some specific and anticipated failure cases. In the same manner, Benkhelifa *et al.* [5] proposed a design of electronic circuits in 2013, in which faulty lines are eliminated and an alternative path fitting to the functionality is sought (Figure 19).

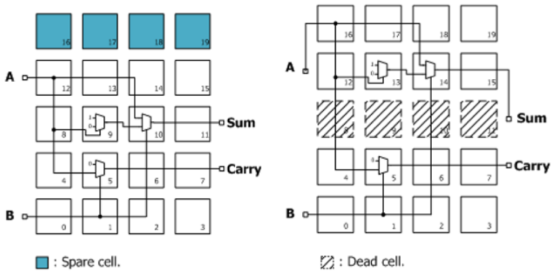


Figure 19 - Self-organizing electronic circuit as proposed by Benkhelifa et al. [5]

Another interesting case, falling under the definition of self-organizing systems, is the conception of a self-healing epoxy composite containing Hexamethylene Diisocyanate (HDI) capsules, proposed by Khun *et al.* in 2014 [27]. The application field is frictional abrasion and wear compensation by HDI-filled microcapsules, which form new layers on the wear track when disrupted. The system was developed with respect to the tribological surface properties, making it a useful application also for functional surfaces. A similar approach for the internal repair of epoxy coatings by the use of Tung Oil was proposed Samsdzadeh *et al.* in 2011 [46].

As the application of self-healing surfaces can be extended to many different research fields, research of different aspects and possible applications have been published over the course of the past 15 years. For an in-depth review of the fundamentals and earlier achievements of self-healing materials, the reader's attention should be drawn to [59].

Likewise, concerning the more recent developments on metal matrix composites including self-lubricating and self-healing aspects, Moghadam *et al.* [36] published in 2014 a comprehensive review of function and performance, as well as an outlook. They highlight the difficulty of healing metallic materials compared to polymers, given their material properties. As an approach pointing rather towards the self-healing category, they also consider the use of Shape Memory Alloys that recover after heating.

5.2. Self-Healing

First of it all, it should be noted that very few publications are currently available that directly concern the use of self-healing technologies in manufacturing systems. Therefore, an overview over recent publications in the field of self-healing technologies, and their possible uses in manufacturing will be provided.

In 2014, Jiang [26] proposed a bio-inspired self-sharpening cutting tool surface, which applies shark-teeth like architectures to hard turning of steel tools. The performance of the proposed structures yields longer tool life, and higher surface qualities than comparable benchmark solutions. The only other mentions of proper self-healing systems were not found directly as integrated parts of manufacturing systems, but can serve as exemplary blueprints for later applications in manufacturing systems: Murata *et al.* [39] demonstrated their fundamental research work on self-repairing systems, consisting of both component and functional healing. The work includes hardware designs of two- and three-dimensional mechanical units and algorithms for self-assembly and self-repair.

Bell *et al.* [4] confirm in 2013 the absence of a universally-accepted notion of self-healing and self-repair. They distinguish self-healing and self-repair in the way that self-repairing system can "partially or fully fix a given fault to continue operation", whilst self-healing systems are able to "bring themselves back to its initial state of operation after a fault has occurred". They also insist on the fact that different research fields (mechanics, electronics, software) bring different meanings to the terms, which leads to further confusions. No practical applications or prototypes are proposed. In a subsequent publication, Bell *et al.* [3] propose the concept of a self-rectifying 4-bar linkage system. The authors critisize the higher complexity, inevitably leading to a less reliable system, possibly hindering the harvest of the full fruit of self-healing techniques.

Levi *et al.* [29] proposed a lab-case of a swarm of reconfigurable robots with self-assembling and intended self-healing capabilities. They tried to prove the applicability of cognitive capabilities, including situation-awareness and decision-taking. While the mere re-organization of a swarm of robots falls under the definition of self-organization, the cognitive capabilities and the strive to repair and reconstitute the original state of swarm members corresponds to the definition of self-healing. However, many challenges in terms of complexity, both hardware and software alike, could not be overcome at the time of publication (2014). The authors mention that with learnings from the Human Brain Project (HBP), the gap to train artificial robot genomes after the human brain controls could be closed.

In general, "data-healing" could also be classified as a self-healing system in the context of manufacturing. Fault detection, isolation and recovery (FDIR) techniques have been applied in software applications for a while. The creation of redundancy and backups on memories help to recover (partial) data loss, while the system excludes faulty parts from further use. This principle can be applied also to general data processing units, e.g. in sensors, like Yang *et al.* [63] presented recently. The key is the self-creation of the necessary redundancies on available capacities. As the faulty parts of a memory usually cannot be recovered to its initial state, FDIR systems are ambiguous in the definition of self-organizing and self-healing systems.

To a certain extent, data loss is always recovered and brought back to an initial desirable state (self-healing), while the hardware aspects used can be described as self-organizing.

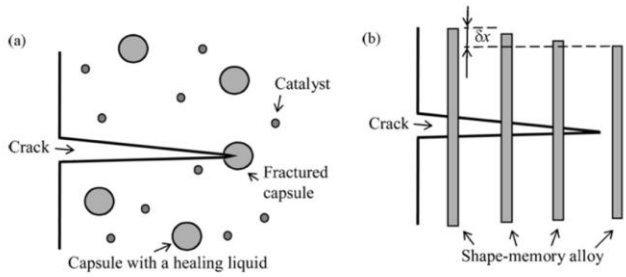


Figure 20 - proposal of a Self-Organizing (a) Liquid Capsule based Material Repair, and a Self-Healing (b) Shape Memory Alloy based Repair [42]

When referring to self-healing, surface reconstruction such as skin reconstruction or plant surface construction immediately comes to mind. Similar principles have been investigated: A possible application case for self-healing compensating wear-out of surfaces was published by Nosonovsky & Bushan [42] in 2010 and is shown in Figure 20. It contains a discussion of the fundamentals of bio-inspired self-healing technical surfaces, as well as most general self-healing mechanisms for surfaces.

Examples presented with a potential use in manufacturing are surface crack healing techniques using micro capsules containing a liquid or reinforcement of crack voids by shape memory alloy demonstrated in Figure 20. The first notion can be considered self-organization (as the liquid will be missing in other places, potentially weakening the structure), whilst the shape memory alloy reinforcement is closer to the definition of proper self-healing.

Another interesting aspect of self-healing aspects is introduced by Farnsworth *et al.* [14] in 2015: An autonomous maintenance system for through-life engineering. Here, the ideal aim is to conceive a system which predicts its need for maintenance and carries out the process autonomously. The emphasis lies on the autonomous execution of the repair task, as it would otherwise correspond to a mere condition monitoring or predictive maintenance system, which is the indispensable prerequisite for Self-Healing systems. This definition was first formulated by Amor-Segan *et al.* [1] in 2007, that is “the ability to autonomously predict or detect and diagnose failure conditions, confirm any given diagnosis, and perform appropriate corrective intervention(s)”. Another publication corresponding to this premise is the adaption of Prognostics and Health Management (PHM) principles to manufacturing environments, denoted as Engineering Immune Systems by Lee *et al.* in 2011 [28]. The approach of resilient and self-maintenance systems also seems like a promising approach for future research.

However, no system is fail proof, given the uncertainty which is at the root of most errors. This uncertainty is hard to cope with, given that current Self-Healing approaches include only anticipated and foreseeable failure cases. Farnsworth *et al.* [14] propose that the combination of their autonomous maintenance approach for robotics with other passive self-healing theories (e.g. Design for X) would yield favorable results. Design for X (DfX) is a term defined by [12] among others, designating a design philosophy focusing on a design around a certain parameter (set). Typical examples are Design for Assembly, or Design for Maintenance. In the specific case, Design for Maintenance, Design for Reliability, and Design for Self-Healing are mentioned, enabling self-awareness, preventive and reactive capabilities of the system. Again, it becomes clear that Self-Healing is currently limited by its specific and foreseen application context. One may argue that biological systems are also limited in their resistance to unforeseen disturbances, however, Associative Learning can help overcoming some of these boundaries. Additionally, Design for Self-Healing as mentioned by [14] implies other DfX, and lacks a concise definition. In future research scenarios, a profound and complete definition of this approach can be considered helpful.

5.3. Self sharpening tools

Several examples of adoption of properties and behaviours of biological systems exist. Famous are especially the properties of biologically inspired surfaces, where a good survey of technologies is collected in Malshe *et al.* [32]. An artificial reproduction of those surfaces yields the effect observed on the biological system. But contrary to biological systems the artificial ones are incapable of a self- healing or regrowth of worn parts of the surface. For example, superhydrophobic surfaces as shown in Figure 21, being manufactured by laser or EDM have the same structure as the lotus leaf, namely a nano structure superposed onto a structure with structural lengths in the 10 µm range, where the microstructure can also be discovered in Figure 21, while the nanostructure are the laser ripples inevitable with ultra short pulsed lasers. Interest exists to use this for machine tools to repeal coolant from the surfaces. Touching this surface damages the fine structure very fast, which decays the effect.

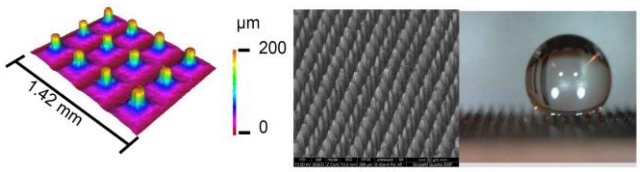


Figure 21 - Superhydrophobic surface manufactured by laser

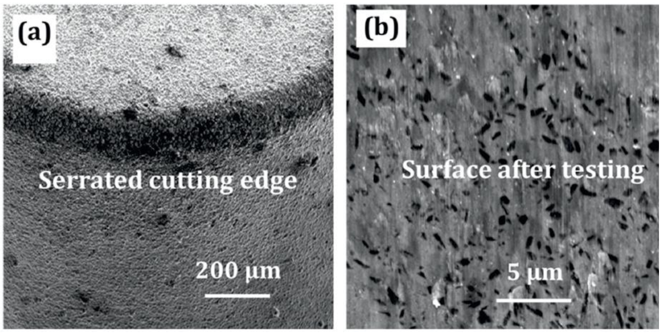


Figure 22 - permanently stable serrated tool surface with the help of CBN – TiN composition of the coating by [32]

All technically usable functional surfaces must be designed such, that the structure is reproduced directly by the expected wear attack, might it be mechanical or chemical. Such a functional surface is presented by [32] and in Figure 22. The wear reduction due to the serrated surface of the tool is reproduced by mechanical wear attack, as the embedded hard particles (CBN) repell the wear attack to the softer areas (TiN) in between and thus stay as new protruding tips above the surface. Whenever they become removed another particle emerging out of the bulk of the TiN-coating takes over the load. This is a new approach for geometrically defined cutting edges but the long used principle of self-sharpening of grinding wheels. This example shows clearly how technical systems must be equipped with a reserve of vitality to deal with wear that continuously removes parts from the surface.

Figure 23 - Archetype for self-sharpening tools: the beaver tooth

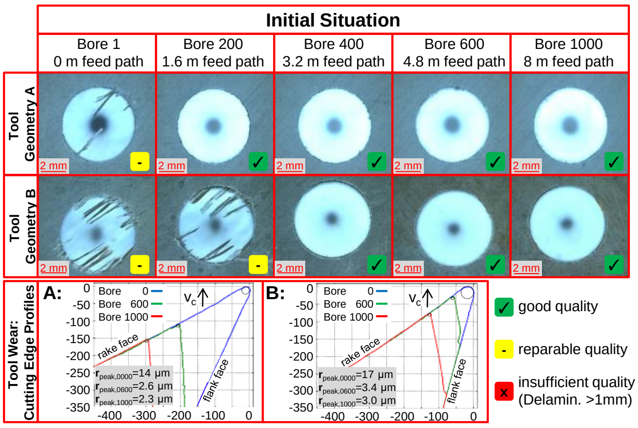


Figure 24 - Self-sharpening tools with cutting in CFRP [53]

For geometrically defined cutting edges of even more interest is the possibility to keep the cutting geometry despite of wear attack stable, which means mainly the micro geometry, namely the cutting edge radius. Here the archetype is the beaver tooth as shown in Figure 23. The beaver tooth contrary to technical systems I growing, which means that the worn geometry is replaced. So the material composition must be such, that wear reproduces the cutting geometry of the growing tooth, which is achieved by graded material, sharp inside and soft outside, so that the reset of the surface at the heavily loaded cutting edge is the same as in the less loaded rake and flank faces of the tooth. A similar approach is presented by Henerichs [24], Voss [54] and in [25] for tools for cutting of carbon fibre reinforced plastics (CFRP), which can be adopted for all abrasive material types. The carbon fibres are hard and due to fibre breakage sharp-edged and thus deteriorating cutting edges strongly. On the other hand, cutting, esp. drilling requires that the fibres are neither torn out nor stay uncut, which requires extremely sharp cutting edges, which are exposed and prone to wear. Figure 24 shows the principle, namely concentrating the wear onto the flank face, while the rake face is covered by a hard and thick diamond coating, by this making up a “graded” material. The benefit of this approach can clearly be revealed by drilling experiments, which is insofar of industrial interest as all CFRP parts in the aerospace industry are riveted together. Even after 1000 holes in unidirectional material IMA which means a total hole depth of 8 m the tool is still sharp enough to totally cut the fibres, while in the first 200 holes the bore hole quality measured by uncut fibres in the exit of the drill is insufficient. This can be enhanced by pre-wearing the tool through laser ablation, which serves twofold, reducing the cutting edge radius and weakening the diamond coating on the flank face, which serves again to concentrate the wear there. The advantage of the drilling process is that the actual position in feed direction of the cutting edge is without interest, which means that the reset of the cutting edge is of no relevance. Milling and turning requires accompanying measures, namely a wear compensation with the help of a wear model for the recalculation of the actual cutting edge position in the tool coordinate system and the respective tool corrective movements of the machine axes. And yet another topic needs to be discussed. As can be seen in Figure 24 the contact length of the flank face increases due to the fact, that wear distribution is not optimally realized and the geometrical setback not the same everywhere. This results in excessive growth of the feed forces. Besides the already discussed grading of the material for instance by means of additive manufacturing of the tool body also the geometry of the flank face can be designed such, that the inbuilt vitality reserve of the tool becomes larger, for instance with a stepped flank face.

6. SENSORS

6.1. Sensor based Cognition in Machines

The common trait of cognitive and biointelligent systems is their ability to replicate human capabilities, such as perception, reasoning, adaption and evolution. In order to shed light on the different facets of cognition in manufacturing, it is useful to distinguish between cognition on the machine level, referred to as Cognitive Technical Systems (CTS) [66], and cognition on factory level, referred to as Biological Manufacturing Systems (BMS) [52], or Cognitive Factories (CF) [65]. The underlying principle being similar, it does nonetheless make sense to separate these two aspects, as their outcome does differ significantly. The subsequent findings are related to cognition on machine level.

Cognitive Technical Systems (CTS): CTS are characterized by the presence of artificial sensors and actuators, as well as their integration or embedding into physical systems [8,66] as shown in Figure 25. Contrary to conventional technical systems, CTS are able to perceive, learn, plan and reason. A central capability of a cognitive control system is therefore to react autonomously to changing conditions as well as unexpected events [55].

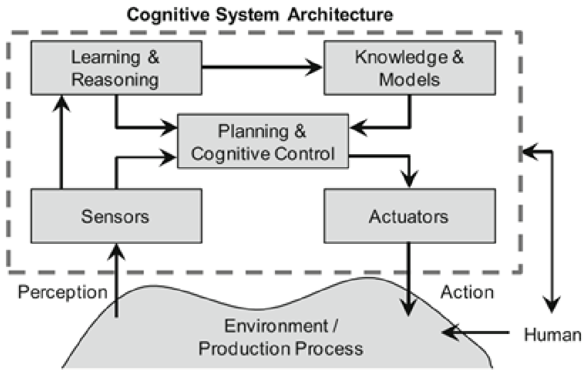


Figure 25 - Cognitive Technical System (CTS) Definition as introduced by Zaeh *et al.* [66]

Based on a cognition model fulfilling this criterion, Shea *et al.* [48] demonstrated a framework to implement cognitive capabilities in a milling process for individual parts. The framework considers both machine and factory level aspects. The on-machine aspects are a flexible fixture planning and re-configuration, as well as toolpath planning and evaluation. The publication has an explorative rather than an exploitative character, given that the performance of the system is neither reviewed nor compared to others.

6.2. Sensors in Abundance

Cognition describes the forming of knowledge through experience and thought, for which sensors provide the necessary data acquisition structure. A critical factor for the application of learning approaches is the information content of the data acquired by the system. As knowledge can only be extracted if the necessary information is present in a sufficiently large, comprehensive and minimum-resolution dataset, the appropriate use of sensors is the prerequisite of cognitive systems. Taking the human as the role model for cognitive capabilities, the ubiquitous use and combination of sensors of different natures (optical, acoustic, olfactory, haptic and others) comes to mind. In this fashion, the large array of sensors distributed across common machine tool architectures and connected to the Control Systems seems to correspond to this notion of Sensors in Abundance. However, Gittler *et al.* [20] point out that the availability of data on current-day machine tools is limited in various sights, and that additional sensor structures need to be deployed in order to allow for a more complete range of cognitive and analytical tasks. They propose a data acquisition architecture, in which internal, external and virtualized sensors of machine tools can be merged in order to ensure a thorough availability of data, advancing sensor architectures on machine tools towards the notion of Sensors in Abundance.

6.3. Exploitation of Optics

A particular example of ubiquitous data inflow is the usage of camera systems. The advantage of optics used in mechatronic systems is their ability to capture a large range of characteristics, the alignment of the information scope with the image boundaries, and the affordable and intuitive use in manufacturing environments. Given the common underlying principle of image transformation and feature recognition, which can be deployed from one system to another in cases of transfer learning, it has ever since been at the forefront of learning and cognitive research. Nagato *et al.* [40] proposed the approach of Figure 26 in 2017, in which a camera system identifying accept or reject parts on a manufacturing line is introduced. Failure-detection and recovery techniques are implemented, allowing the camera system to autonomously adapt to both environmental and specification changes without noticeable drops in the recognition rate. Due to modifications in the production environment (replacement or adjustments of components, changing lighting conditions, different camera installation alignments), the recognized images can suddenly differ a lot. In order to overcome this issue, the changing conditions are identified, and the analysis model is continuously updated. For sudden and very harsh changes, the intervention of an expert to requalify the underlying model can be implemented.

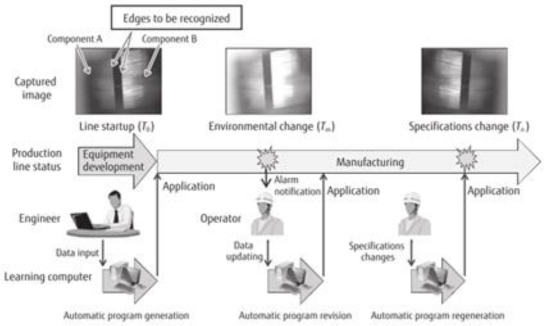


Figure 26 - Image Recognition System Architecture, conceived by Nagato *et al.* [40]

This system corresponds to the here within defined notions of Cognitive Technical Systems, learning with structured data, transfer learning, and Self-Organizing Systems. It adopts image recognition capabilities of other algorithms, whilst being trained on the specific job with images of accept and reject parts, to which the outcome is specifically known. Additionally, the system is capable of actively detecting a predefined set of changes, and it reacts autonomously to mitigate the effects of changes on the overall system behavior.

7. CONCLUSION

Biological systems have some properties making them clearly superior to today’s technical systems, which has been shown so far, which makes it desirable to enhance technical systems with beneficial properties of biological systems. This is in the long line of development of production machines a logical consequence. The increased understanding of systems behaviour and from this perpetual optimization has reached a state, where deterministic approaches become extremely complicated or reach their limits. Together with the unprecedented possibilities of data evaluation cognitive systems discussed already long ago become now possible and beneficial. This “biologicalization” of technical systems enables to overcome the vulnerability of highly developed technical systems and increase their robustness against faults as well external as internal disturbances. Especially for production systems that have to master their processes with highest repeatability and quality, such that only very limited group of operators are capable to fulfil this task can greatly profit from cognitive capabilities of biological systems. Releasing the operators from complexity can only be achieved if the production machine deals autonomously with the full complexity of the physics of machine, process and environment, which is only achievable with machine cognition and machine learning approaches. “Intelligence” to the machine necessarily requires to overcome the today’s separation between intelligent planning systems and stupid execution systems, where machines without a generalized CAM, like grinding and EDM might become the forerunners of biologicalization. To enhance this development process, a multitude of machines needs to be observed and thus Industrie 4.0 is necessary prerequisite for this approach. As the number of same machines in same operating conditions are sparse, modelling to make data comparable, to transform them to data for a master system plays a significant role. In every respect, it is worthwhile to systematically imagine which properties of biological systems are worth to be considered in production machinery, generating not only a new type of machine, but also a new type of machining processes and factory organization. An estimation of time scales and industrial readiness of different biologicalization approaches is given in Figure 27. An excellent overview on aspects of biological transformation of production is given in [9].

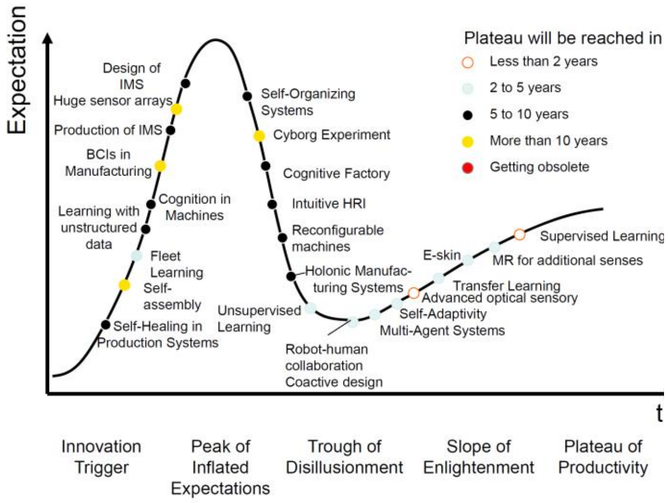


Figure 27 - Aspects of biologicalization illustrated within the Gartner Hype cycle.

8. ACKNOWLEDGEMENTS

The authors are grateful for information provided by P. Blaser and T. Haas and the financial support of the Commission for Technology and Innovation CTI, Switzerland as well as for the support of MTTRF, San Francisco.

9. REFERENCES

[1] Amor-Segan M, McMurran R, Dhadyalla G, Jones RP (2007) Towards the Self Healing Vehicle, Automotive Electronics, *2007 3rd Institution of Engineering and Technology Conference*.

[2] Arinez J, Ou X, Chang Q (2017) Gantry Scheduling For Two-Machine One-Buffer Composite Work Cell By Reinforcement Learning, in Proceedings of the ASME 2017 12th International Manufacturing Science and Engineering Conference, pp. 1–7

[3] Bell C, Farnsworth M, Tiwari A, Dorey R (2013) Theoretical design of a self-rectifying 4-bar linkage mechanism, Procedia CIRP, 11:385–389, DOI:10.1016/j.procir.2013.07.028.

[4] Bell C, McWilliam R, Purvis A, Tiwari A (2013) On the concepts of self-repairing systems Colin, The Journal of the Institute of Measurement and Control, 44/April:0–103, DOI:10.1063/1.2756072.

[5] Benkhelifa E, Pipe A, Tiwari A (2013) Evolvable embryonics: 2-in-1 approach to self-healing systems, Procedia CIRP, 11:394–399, DOI:10.1016/j.procir.2013.07.029.

[6] Blaser P, Pavlicek F, Mori K, Mayr J, Weikert S et al. (2017) Adaptive learning control for thermal error compensation of 5-axis machine tools, Journal of Manufacturing Systems, 44:302–309, DOI:10.1016/j.jmsy.2017.04.011.

[7] Brecher C, Hirsch P, Weck M (2004) Compensation of Thermo-elastic Machine Tool Deformation Based on Control internal Data, CIRP Annals - Manufacturing Technology, 53/1:299–304.

[8] Buss M, Beetz M, Wollherr D (2007) CoTeSys — Cognition for Technical Systems, of the 4th COE Workshop on Human, p. 10.

[9] Byrne G, Dimitrov D, Monostori L, Teti R, van Houten F, Wertheim R. Biologicalization: Biological transformation in manufacturing, CIRP Journal of Manufacturing Science and Technology 21, (2018): 1-32.

[10] Dabrowski C, Mills K (2002) Understanding Self-healing in Service-Discovery Systems, pp. 15–20.

[11] Denkena B, Dittrich M, Uhlich F (2016) Self-optimizing cutting process using learning process models, Procedia Technology, 26:221–226, DOI:10.1016/j.protcy.2016.08.030.

[12] Desai A, Mital A (2013) Design simplification and innovation through adoption of a time-based design for maintenance methodology, in Human Work Productivity: A Global Perspective, S. Kumar, A. Mital, and A. Pennathur, Eds. CRC Press, p. 253.

[13] Escobar CA, Morales-Menendez R (2017) Machine Learning and Pattern Recognition Techniques for Information Extraction to Improve Production Control and Design Decisions, in Advances in Data Mining. Applications and Theoretical Aspects, P. Perner, Ed. Cham: Springer International Publishing, pp. 286–300.

[14] Farnsworth M, Bell C, Khan S, Tomiyama T (2015) Autonomous Maintenance for Through-life Engineering Services, pp. 395–419, DOI:10.1007/978-3-319-12111-6.

[15] Farnsworth M, Tiwari A (2015) Modelling, Simulation and Analysis of a Self-healing Energy Harvester, Procedia CIRP, 38:271–276, DOI:10.1016/j.procir.2015.07.084.

[16] Frei R, McWilliam R, Derrick B, Purvis A, Tiwari A et al. (2013) Self-healing and self-repairing technologies, International Journal of Advanced Manufacturing Technology, 69/5–8:1033–1061, DOI:10.1007/s00170-013-5070-2.

[17] Fritz Studer AG (2012) StuderTechnology Integrated. [Online]. Available: www.grinding.com/uploads/media/StuderTechnology_Integrated_DE_04.pdf. [Accessed: 18-Oct-2017].

[18] Gaegauf F, Technologie schafft Wettbewerbsvorteile, Schweizer Präzisionstechnik (2011): 26-28.

[19] Gebhardt M. Thermal Behaviour and Compensation of Rotary Axes in 5-Axis Machine Tools, Ph.D. thesis, No.21733, ETH Zurich, (2014)

[20] Gittler T, Gontarz A, Weiss L, Wegener K (2018) A Fundamental Approach for Data Acquisition on Machine Tools as Enabler for Analytical Industrie 4 . 0 Applications, Conference Proceedings - 12th CIRP Conference on

Intelligent Computation in Manufacturing Engineering*, 2018, Gulf of Naples, Italy*.
[21] Goyal D, Pabla BS (2015) CIRP Journal of Manufacturing Science and Technology Condition based maintenance of machine tools — A review, CIRP Journal of Manufacturing Science and Technology, 10:24–35, DOI:10.1016/j.cirpj.2015.05.004.
[22] Haas T, Set Point Optimization for Machine Tools, Ph.D. thesis ETH Zürich, No. 24937, (2018).
[23] Haas T, Lanz N, Keller R, Weikert S, Wegener K (2016) Iterative Learning for Machine Tools Using a Convex Optimisation Approach, Procedia CIRP, 46:391–395, DOI:10.1016/j.procir.2016.04.033.
[24] Henerichs M, Bohrbearbeitung von CFK unter besonderer Berücksichtigung der Schneidkantengeometrie, Ph.D. thesis ETH Zürich 22629, (2015)
[25] Henerichs M, Voss R, Kuster F, Wegener K. Machining of carbon fiber reinforced plastics: Influence of tool geometry and fiber orientation on the machining forces, CIRP Journal of Manufacturing Science and Technology. **9**: pp. 136-145.
[26] Jiang W (2014) Bio-inspired self-sharpening cutting tool surface for finish hard turning of steel, CIRP Annals - Manufacturing Technology, 63/1:517–520, DOI:10.1016/j.cirp.2014.03.047.
[27] Khun NW, Sun DW, Huang MX, Yang JL, Yue CY (2014) Wear resistant epoxy composites with diisocyanate-based self-healing functionality, Wear, 313/1–2:19–28, DOI:10.1016/j.wear.2014.02.011.
[28] Lee J, Ghaffari M, Elmeligy S (2011) Self-maintenance and engineering immune systems: Towards smarter machines and manufacturing systems, Annual Reviews in Control, 35/1:111–122, DOI:10.1016/j.arcontrol.2011.03.007.
[29] Levi P, Meister E, Schlachter F (2014) Reconfigurable swarm robots produce self-assembling and self-repairing organisms, Robotics and Autonomous Systems, 62/10:1371–1376, DOI:10.1016/j.robot.2014.07.001.
[30] Lieber D, Stolpe M, Konrad B, Deuse J, Morik K (2013) Quality prediction in interlinked manufacturing processes based on supervised & unsupervised machine learning, Procedia CIRP, 7:193–198, DOI:10.1016/j.procir.2013.05.033.
[31] Luckow A, Cook M, Ashcraft N, Weill E, Djerekarov E et al. (2016) Deep Learning in the Automotive Industry: Applications and Tools, Big Data: International Conference on Big Data, pp. 3759–3768, DOI:10.1109/BigData.2016.7841045.
[32] Malshe A, Rajurkar K, Samant A, Hansen HN, Bapat S, Jiang W (2013) Bio-inspired surfaces for advanced applications, CIRP Annals – Manufacturing Technology 62/2, p. 607-628
[33] Matt DT (2012) Enhanced Survivability- a Framework of Self-Healing Mechanisms' Installation Using Axiomatic Design in Agile Manufacturing Systems, pp. 153–159.
[34] Mayr J, Egeter M, Weikert S, Wegener K (2015) Thermal error compensation of rotary axes and main spindles using cooling power as input parameter, Journal of Manufacturing Systems, pp. 542–549.
[35] Mayr J, Jedrzejewski J, Uhlmann E, Donmez A, Knapp W, et al. (2012) Thermal issues in machine tools, Annals of the CIRP 61/2, 2012, ISSN ISSN 0007-8506, DOI 10.1016/j.cirp.2012.05.008
[36] Moghadam AD, Schultz BF, Ferguson JB, Omrani E, Rohatgi PK et al. (2014) Functional metal matrix composites: Self-lubricating, self-healing, and nanocomposites-an outlook, Jom, 66/6:872–881, DOI:10.1007/s11837-014-0948-5.
[37] Monostori L (2002) AI and machine learning techniques for managing complexity, changes and uncertainties in manufacturing, IFAC Proceedings Volumes (IFAC-PapersOnline), 15/1:119–130, DOI:10.1016/S0952-1976(03)00078-2.
[38] Mou J, Liu CR (1995) An Adaptive Methodology for Machine Tool Error Correction, J. Eng. Ind 117(3), 389-399 (Aug 01, 1995) (11 pages) doi:10.1115/1.2804345
[39] Murata S, Yoshida E, Kurokawa H, Tomita K, Kokaji S (2001) Self-repairing mechanical systems, Autonomous Robots, 10/1:7–21, DOI:10.1023/A:1026540318188.
[40] Nagato T, Shibuya H, Okamoto H, Koezuka T (2017) Machine Learning Technology Applied to Production Lines : Image Recognition System, 53/4:52–58.
[41] Neugebauer R, Wabner M, Ihlenfeldt S, Friess U, Schneider F, Schubert F (2012) Bionics Based Energy Efficient Machine Tool Design. Procedia CIRP, 3:561–566.
[42] Nosonovsky M, Bhushan B (2010) Surface self-organization: From wear to self-healing in biological and technical surfaces, Applied Surface Science, 256/12:3982–3987, DOI:10.1016/j.apsusc.2010.01.061.
[43] Oquab M, Bottou L, Laptev I, Sivic J (2014) Learning and Transferring Mid-Level Image Representations using Convolutional Neural Networks, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1717–1724, DOI:10.1109/CVPR.2014.222.
[44] Reinhart RF, Steil JJ (2016) Hybrid Mechanical and Data-driven Modeling Improves Inverse Kinematic Control of a Soft Robot, Procedia Technology, 26:12–19, DOI:10.1016/j.protcy.2016.08.003.

[45] Ringsquandl M, Lamparter S, Lepratti R (2017) Advances in Production Management Systems. The Path to Intelligent, Collaborative and Sustainable Manufacturing, 514:302–310, DOI:10.1007/978-3-319-66926-7.
[46] Samadzadeh M, Boura SH, Peikari M, Ashrafi A, Kasiriha M (2011) Tung oil: An autonomous repairing agent for self-healing epoxy coatings, Progress in Organic Coatings, 70/4:383–387, DOI:10.1016/j.porgcoat.2010.08.017.
[47] Schischke K (2010) Energy-Using Product Group Analysis, Lot 5: Machine tools and related machinery, Task 2 Report. Fraunhofer IZM, Berlin.
[48] Shea K, Ertelt C, Gmeiner T, Ameri F (2010) Design-to-fabrication automation for the cognitive machine shop, Advanced Engineering Informatics, 24/3:251–268, DOI:10.1016/j.aei.2010.05.017
[49] Spescha D, Framework for Efficient and Accurate Simulation of the Dynamics of Machine Tools, Ph.D. thesis Clausthal University of Technology, Preprint, (2018).
[50] Shin S-J, Woo J, Rachuri S (2014) Predictive Analytics Model for Power Consumption in Manufacturing, Procedia CIRP, 15:153–158, DOI:10.1016/j.procir.2014.06.036
[51] Srivastava AN, Han J (2016) Machine Learning and Knowledge Discovery for Engineering Systems Health Management. CRC Press.
[52] Ueda K, Vaario J, Ohkura K (1997) Modelling of Biological Manufacturing Systems for Dynamic Reconfiguration, CIRP Annals - Manufacturing Technology, 46/1:343–346, DOI:10.1016/S0007-8506(07)60839-7.
[53] Voss R, Henerichs M, Harsch D, Kuster F, Wegener K (2017) Optimized approach for characterization of cutting edge micro-geometry in drilling carbin fibre reinforced plastics (CFRP), Internationa,l Journal of Advanced Manufacturing Technology 90, 2017: 457-472.
[54] Voss R (2017) Fundamentals of Carbon Fibre Reinforced Machining Ph.D. thesis ETH Zürich 24283
[55] Verein Deutscher Ingenieure (2012) VDI 6220 Blatt 1, /Pages:1–36.
[56] Wegener K, Bleicher F, Krajnik P, Hoffmeister H-W, Brecher C. Recent Developments in Grinding Machines, CIRP Annals Manufacturing Technology 66/2, (2017): 779-802.
[57] Wegener K, Weikert S, Mayr J. Age of Compensation – Challenge and Chance for Machine Tool Industry, IJAT 10 (2016): 609-623.
[58] Weiss L, Wegener K (2007) Trends im Werkzeug-maschinenbau. MB-Revue, Zürich.
[59] Wool RP (2008) Self-healing materials: a review, Soft Matter, 4/3:400, DOI:10.1039/b711716g.
[60] World Machine Tool Survey (2016), Gardner Research; https://www.gardnerweb.com/cdn/cms/2016%20WMTS%20Report.pdf. Retrieved 2018-06-06
[61] Wuest T, Irgens C, Thoben K-D, Wuest T, Thoben K-D (2014) An approach to monitoring quality in manufacturing using supervised machine learning on product state data, Journal of Intelligent Manufacturing, 25:1167–1180, DOI:10.1007/s10845-013-0761-y
[62] Wuest T, Weimer D, Irgens C, Thoben K-D (2016) Machine learning in manufacturing: advantages, challenges, and applications, Production & Manufacturing Research, 4/1:23–45, DOI:10.1080/21693277.2016.1192517.
[63] Yang J, Chen Y, Zhang L (2017) An Efficient Approach for Fault Detection, Isolation, and Data Recovery of Self-Validating Multifunctional Sensors, IEEE Transactions on Instrumentation and Measurement, 66/3:543–558, DOI:10.1109/TIM.2016.264265.
[64] Yang H, Ni J (2005) Adaptive model estimation of machine-tool thermal errors based on recursive dynamic modeling strategy, International Journal of Machine Tools and Manufacture, 45/1:1–11.
[65] Zaeh M, Beetz M, Shea K, Reinhart G, Bender K et al. (2008) The cognitive factory, in Changeable and Reconfigurable Manufacturing Systems, pp. 355–371.
[66] Zaeh M, Ostgathe M, Geiger F, Reinhart G (2012) Adaptive Job Control in the Cognitive Factory, in Enabling Manufacturing Competitiveness and Economic Sustainability, H. A. ElMaraghy, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 10–17.
[67] Zhang Y, Qian C, Lv J, Liu Y (2017) Agent and Cyber-Physical System Based Self-Organizing and Self-Adaptive Intelligent Shopfloor, IEEE Transactions on Industrial Informatics, 13/2:737–747, DOI:10.1109/TII.2016.2618892.
[68] Zirn O (2008) Machine tool analysis, Ph.D. thesis, ETH Zurich,. doi:10.3929/ethz-a-005825192.

10. BIOGRAPHY

Konrad Wegener is Head of the Institute for Machine Tools and Manufacturing (IWF) at the Swiss Federal Institute of Technology Zurich (ETHZ). His research is focused on machine tools, manufacturing processes and process chains, as well as the development, evaluation and optimization of production machinery.

Thomas Gittler works on manufacturing-related Industry 4.0 topics during his current doctoral studies at the Institute for Machine Tools and Manufacturing (IWF) of the Swiss Federal Institute of Technology Zurich (ETHZ). He also accompanies the digital transformation of Swiss machinery manufacturers.

Lukas Weiss is a senior researcher with an industrial background leading a group focused on conceptual issues, innovative materials, energy efficiency and digitalization of machine tools at a research institute related to ETH Zürich.

Stream A

Virtual Dashboards in Pilot Production Environments

E. Lutters[1,2], J. de Lange[1], R. Damgrave[1]
[1]Dep. of Design, Production and Management, University of Twente, The Netherlands
[2]Fraunhofer Project Center at the University of Twente, Enschede, The Netherlands

Abstract
Development cycles encompass an overwhelming arsenal of data, tools, methods and techniques. They relate to many different types of stakeholders, with different perspectives, objectives and interests. Also, there are many differences in fields of expertise, languages/jargon, system levels, cultures and opinions that complicate multi-stakeholder decision making. Even if unequivocal information and/or Digital Twins are available, shared and unambiguous understanding in decision making is not obvious. Virtual Dashboards allow stakeholders to identify with any subset of the perspectives involved, because the insight in the information is presented in an autarchic manner. This implies that the information content is filtered and presented according to perspective(s) of the specific stakeholders involved. Virtual Dashboards build on the realm of information, models, scenarios, simulations, tools and techniques available, allowing stakeholders to address specific subjects or aspects of e.g. a production environment. To make the Virtual Dashboards as instrumental as possible, different ways of making the information insightful are combined. This ranges from smart information sets, via simulations, infographics, 3D models and visualisations to full-blown Virtual Reality or Augmented Reality applications. Depending on the way in which the information is accessed, Virtual Dashboards are especially appropriate to connect to and control remote locations. Moreover, a Virtual Dashboard can ostensibly integrate multiple environments into one coherent entirety.

Keywords
Virtual Dashboard, Digital Twin, Pilot Production Environment

1 INTRODUCTION

What has come to be known as the “Fourth Industrial Revolution” is being shaped predominantly by production digitalisation and networking. Buzzwords such as the “Internet of Things and Services” and “Cyber-physical Production Systems” promise increased networking of autonomous and self-optimising production machines and intelligent products which can be customised to manufacture highly individual outcomes. On the one hand, much of this remains a vision – for the time being. At the same time, however, industry and academia employ many initiatives that already do realise networked (sub-)systems. These engender overwhelming amounts of data while relying on growing numbers of process-, information- and knowledge structures and management systems. Accordingly, steady-state production environments are expected to grow to states of complexity that are extremely difficult to establish, to understand and to change, especially if e.g. self-optimisation, re-configurability or machine learning are involved [1, 2]. Particularly if the development process of a production environment is considered, the amount of available information and interdependencies might well exceed the capabilities of any development team. Even more, development cycles increasingly incorporate multiple companies, locations, stakeholders and countries, thus again challenging the ability to establish and maintain adequate overview.

Growing with the increasing elusiveness of production environments has been the set of tools, techniques and approaches to manage the development cycles of such environments. In this, the focus predominantly is on ways to better exploit, establish and position the processes that are envisaged to constitute the development cycle. However, unduly attention for these processes would distract from the actual focal point of any development cycle: the evolvement of, in this case, the production environment under consideration [3]. At the same time, there is an increasing tendency to capture data on assets, processes or systems by establishing digital replicas or digital twins, based on e.g. IoT approaches. In themselves, however, such digital replicas do not necessarily contribute to the development cycle, unless the data they represent can be converted into information that is meaningful and relevant for the development team. Moreover, in many cases, the sheer amount of data undermines the usefulness and applicability thereof. With this, the many stakeholders in development cycles of production environments need to blend the abundance of available process-oriented directions and the foreshadowed eruption of potentially useful data into a meaningful entirety. This is done in the many multi-criterion, multi-stakeholder decisions that together constitute the development cycle of a production environment, by actualising the evolvement of the information content defining that environment.

This publication sketches the use of so-called virtual dashboards to allow (groups of) stakeholders to purposefully, effectively and efficiently focus on the information that is underlying the decisions they have to reach, while considering both time-dependency and different levels of aggregation. The publication introduces pilot production environments as the scope of application and then addresses the digital system reference and synthetic environments as the foundation for these virtual dashboards.

2 PILOT PRODUCTION ENVIRONMENTS

A pilot production environment or pilot plant is a facility that allows a company to develop, test, improve and upscale (parts of) a production environment while not hampering primary processes and avoiding investments where possible [4]. Therefore, in this research, pilot plants consist of virtual entities where possible and physical entities where required. Such physical entities in a pilot plant need not necessarily be at one single location. In the context of this research and the underlying industrial case studies, the scope of a pilot plant is the manufacturing of discrete physical products, from small-batch production to dedicated line production. In its constitution, a pilot plant is not assumed to be critical in the primary process of a company. This implies that the flow of products through the pilot plant need not be continuous, reliable or uninterrupted. Additionally, physical components or semi-finished products may follow routes that are more informative for the team that develops the pilot plant than that they would be optimised or beneficial for the envisaged production environment. At the same time, virtual entities in the pilot plant will obviously merely render simulated outcomes, thus only influencing or interfering with the way in which physical entities can behave.

Therefore, the ‘pilot plant as a whole’ merely exists by virtue of the platform that represents the separate physical components and virtual components as well as their mutual interactions. Both types of entities and the interactions can be represented according to the perspectives of the stakeholders involved. Therefore, a virtual component is not ‘just’ a model of a machine or sub-system in the pilot plant; it rather consists of a set of coherent, matching and aligned simulations and models. This is especially relevant for the many situations where new production processes, resources, machine tools, materials or product designs are innovative beyond currently available expertise and experience. Each of these simulations or models covers a subset of all perspectives involved, addressing e.g. logistics, sustainability, mechanical process simulation or ergonomics [4].

It goes without saying that an effective pilot plant requires a flawless overview of all the data, information, stakeholders and perspectives that play a role in the development cycles of the production environment and the products it will engender. Moreover, sensible ways of interaction with the simulated pilot production environment need to be provided. As already mentioned in section 1, the sheer availability of data is a necessary but not sufficient precondition. Foremost, developers of production environments require a way to bring structure to the (potentially) available data, even before they can fully appreciate what data will lead to meaningful information in their development cycles. In literature, may approaches investigate the use of digital twins (e.g. [5-7]), but as these represent already instantiated products or environments, digital twins can only cover part of the information structure required. After all, in the development cycle of production environments (and products alike), there is a clear difference between to-be and as-is states of development, whereas also what-if states play essential roles. For this reason, the following section elaborates on how a digital system reference can be instrumental in the development cycle of pilot production environments.

3 DIGITAL SYSTEM REFERENCE

The digital twin is commonly known as a key enabler for digital transformation, though literature does not provide a common understanding on this term. It is used differently over disciplines, focus areas, level of integration and technology with many different interpretations [8]. Hence, this publication does not aim to endorse or propose one specific definition or structure; it rather formulates a digital system reference framework allowing developers of pilot production environments to interact with the data/ information content of the development cycle.

Any such system must be able to evolve with the pilot production environment over its entire lifecycle; from initial concept, via exploration phases to final handover of the resulting production environment. Consequently, the system does not only need to be able to deal with the amalgamation of virtual and real entities, but also with the integration of existing/instantiated components and components that are only envisaged. This immediately implies that a digital system reference should not only refer to data on instantiated products, production environments or part thereof, but should also capture the design intent in the to-be stage.

For developers of pilot production environments, the purpose of the digital system reference is that it can be questioned, adjusted, varied and probed while providing simulations that are independent of, but always represent, the tangible product, asset or production environment. The digital reference aggregates findings and measurements from individual products or machines into insights that underpin and drive development cycles, leading to robust development and optimisation of products, processes and production environments. Figure 1 shows the overall digital system reference.

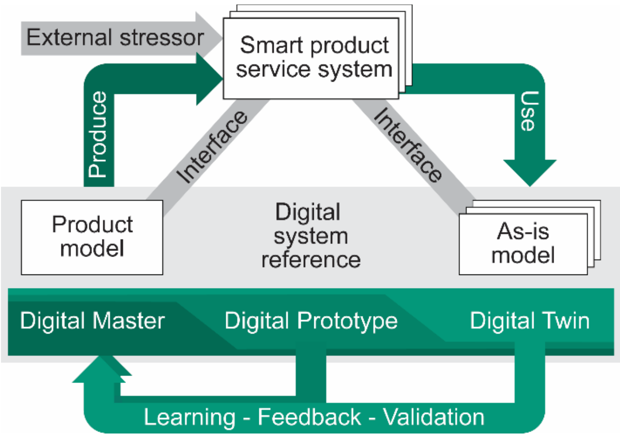


Figure 1 - Digital system reference

3.1 Environment

Being generic in nature, the digital system reference can relate to (parts of) products, assets and production environments. For this reason, figure 1 refers to 'Smart product service system' as the encompassing term for the system subject. In the context of this publication, the notion refers to the pilot production environment that is the goal of the development cycle. As section 2 describes how part of that environment can be real and part can be virtual, this environment exhibits behaviour that is either the consequence of design decisions (captured in a to-be product model) and can feed or underpin subsequent development decisions while being impacted by external stressors. Multiple (partial) instantiations of the pilot production environment may exist simultaneously, from which it can feed both to different as-is models while also allowing for aggregated input to such models.

3.2 Digital triplet

To distinguish between the as-is and to-be models mentioned, the digital system reference employs a digital triple: the digital master, prototype and twin. To characterise the roles of these different part, the following depictions are used:

- digital twin: the conglomerate of data, information, models, methods, tools and techniques to represent current states of an instantiated system coherently and consistently.
- digital master: the envisaged state of a system and its components as the captured outcome of the development cycle.
- digital prototype: the envisaged state of a (extrapolated/predicted) system, based on models, simulation and aggregated experience, related to the inputs by the 'to-be' digital master and the 'as-is' digital twin(s).

In a more descriptive approach: any production environment is the result of a development cycle, where myriad decisions by multiple stakeholders converge and are aligned into an evolving depiction of the design of that pilot production environment. This to-be model captures the design at any point in time, considering uncertainties and anticipating changes while simulating behaviours of the system. This digital master thus depicts the blueprint of the production-environment-to-be. Obviously, while under development, many alternatives and variants will be considered, which all need exploration, evaluation and assessment, for example by means of 'what-if' analyses. Such intermediary trials or concepts are not yet accepted as integrated part of the digital master; therefore, they can separately be captured in digital prototypes that dissociate 'design freezes' from 'creative endeavours'. Digital prototypes thus have a temporal character, allowing for constructive findings to be integrated in the digital master, while vain attempts fade away. In both cases, however, the rationale of decisions can endure. With evolving to-be models and digital prototypes available, more physical entities can be instantiated, with the immediate ability to measure characteristics, performance and behaviour. Here, actual feedback on realistic and practical behaviour is available, rendering a fingerprint of the specific instantiation. This represents the typical digital twin as it is often described.

3.3 Interaction

Distinguishing between the three elements of the triplet allows for a structured and accurate depiction of how these elements interact. The digital twin, for example, collects actual data that allows for accurate comparisons to the predictions that can be made based on the digital master. This is instrumental in determining the validity of the to-be models, but also in making adjustments. The same yields true for the virtual prototypes that can test circumstances/situations and even can explore the response of the to-be model by stressing it up to and beyond the anticipated threshold of parameters involved. This allows for making adjustments, but certainly also for dedicated (machine) learning in modelling efforts. What is especially relevant in these interaction and feedback mechanisms is that there is no pre-defined perspective or viewpoint involved. This implies that different stakeholders can study and learn from different simulations or 'what-if' scenarios. Also, sensor data from actual measurements can influence decisions from different viewpoints. As an example, validation and feedback can be related to perspectives as different as logistics, quality, (anti)fragility [9], cost estimations, and environmental impact.

4 SYNTHETIC ENVIRONMENT

The data/information modelling basis depicted in section 3 may provide a sound foundation for reaching decisions in development cycles of pilot production environments, but unfortunately the efforts required to accomplish effective and efficient making are far from trivial. After all, given the observations on the numbers of processes and the amount of data/information that is available in development cycles (see section 1), the biggest challenges for all stakeholders involved will relate to:

- navigating the information content to adequately address the current state of affairs;
- envisioning impacts of (un)intentional changes and internal as well as external stressors;
- placing oneself in different or shared viewpoints and responsibilities.

Whereas the virtual dashboards (as will be introduced in section 5) are intended to allow stakeholders to meet precisely these challenges, these dashboards are not the automatic and obvious consequence of the underlying information content and digital system reference. The relation between the information content and the virtual dashboards needs explicit development efforts, for example because objectivity of the information content and subjectivity of stakeholder perspectives require alignments. This development effort is often seen as a repetitive one-off exercise, reinventing the wheel for new situations. However, research on Synthetic Environments aims to further facilitation of multi-criterion, multi-stakeholder decision making against the background of extensive data and information realms.

Synthetic Environments (SE) are design environments that bring together real and virtual components to allow for adequately experiencing shared information [10, 11]. They range from small setups, focusing on e.g. working with a new machine, to large systems for conjoint development of factory layouts. SEs are composed of a wide variety of tools, techniques, information, hardware and software components. With the advent of more affordable VR/AR solutions, building SEs increasingly is a configuration issue rather than an innovative design effort. This makes SE development less investment-intensive, more flexible and agile, but not necessarily less complex. Ensuring that an SE meets all functional specifications and requirements with adequate quality requires SE configuration to be structured, vigorous and predictable, but also flexible and adaptable, while doing justice to all the different stakeholders involved.

As such, establishing an SE for pilot plants intends to bring the information content to life, allowing any stakeholder to interact with and (conjointly) decide on aspects that are relevant for this stakeholder's perspective. This does not mean that other viewpoints are disregarded but rather that stakeholders can focus on selected priorities in the overall pilot production plant. With that, the end result of establishing an SE for pilot production plants is a coherent set of virtual dashboards, allowing the different stakeholders to fathom the input, rationale, consideration and foreseen consequences of their decisions.

5 VIRTUAL DASHBOARD

In its simplest form, a virtual dashboard can be seen as a control room for a pilot production environment that does not (yet) exist. Such a control room aggregates input from the shopfloor and allows for interventions at that shopfloor. However, a virtual dashboard exceeds the capabilities of a traditional control room by far. For example, a dashboard may aggregate results from simulations of non-existing assets or show responses to planned or unplanned stressors in the production environment. In this, the virtual dashboard shows resemblance to a flight simulator, allowing for purposeful interactions with real and imposed conditions. Yet, the capabilities of virtual dashboards exceed the preconceived setup of a flight simulator, as virtual dashboards change perspectives, viewpoints or level of aggregation.

As a simple example, a fully functional set of virtual dashboards may allow a stakeholder to interact with a future production environment in different roles, for example as an operator, a quality/safety manager, a process or production engineer, or as a forklift driver. Representing these different roles and the corresponding perspectives allows for situational analyses, what-if analyses and insights that, together with carefully planned modelling and simulation efforts, are the basis for underpinned decision making.

Shifting perspectives also addresses the ability to use non-traditional ways of assessing activities in the production environment. Next to analysing the result of using a specific process or a specific machine, or already modelling an approach for predictive maintenance on a machine, it is equally possible to 'travel with a product' through the production environment while aggregating the impact of subsequent process steps. Also, virtual dashboards allow for zooming in and out. This implies that a dashboard can give access to the details of a process model for a specific production step (even if the machine for that process has not yet be fully developed) and can simultaneously focus on the overall logistic performance of the entire envisaged environment. This entails that virtual dashboards also purposefully integrate levels of aggregation, for example by simultaneously interacting with a process, material or machine parameter and with the impact such parameters have on e.g. overall logistics and expected cost or quality. Moreover, a virtual dashboard can interact with assets as if they were part of an integrated production environment, although they in reality are at separate locations or only exist as a model.

Consequently, stakeholders can choose e.g. to 'travel with a product through the production line', to 'observe process conditions on a machine that show signs of wear', to 'focus on the overall production throughput', to 'relate production bottlenecks to product features' or to 'compare defects between machines and between product types'. Underlying assumption in all these possibilities is that neither the information content, nor the Synthetic Environment that engendered the virtual dashboard are limiting the way in which interactions are possible. In other words, the information content must be filtered and presented fully according to perspective(s) of the specific stakeholders involved, without the bias of any pre-assumed hierarchy. Such a virtual dashboard must therefore be autorarchic in nature: rather the required perspective and filtering are leading than any predefined hierarchy. This autorarchic approach renders the essential added value; in terms of the control room, it means for example that the control room changes depending on who is in it and who enters it under which circumstances.

Current research efforts aim to gain experience such that a standard architecture for developing virtual dashboards by means of Synthetic Environments can be developed. Given the complexity and breadth of the field of pilot production environments and especially the volatility of developments related to e.g. digital twins, IoT and Industry 4.0 in general, this is an impactful, arduous and effortful undertaking. With that, the development of the architecture will span multiple years and will require many cases as both input and as test environment. This implies that this publication can not yet preview such an architecture, but it can depict some experiences and outcomes of one of the case studies that are or have been executed.

6 PRACTICAL APPLICATION

The practical application of virtual dashboards depicted here has been established in co-operation with a company that aimed to extend production capacity by opening up a new and innovative production environment. Many engineering companies were involved (from architectural to building contractors and machine designers); each company and sub-contractors tended to use different models and information storage approaches. At the same time, the characteristics of the production environment caused direct and significant interactions between the main machine(s) in the building (some multiple storeys in size) and the geometry and construction of the building. In total, more than ten companies were involved in the development cycle, whereas their involvement was also not necessarily concurrent, and communication was often bilateral. This made the company reach out to build an evolving virtual dashboard for the development cycle of the production environment.

In this context, the main driver for creating the virtual dashboard(s) was the risk involved in misinterpretations of requirements, models and foreseen consequences of design decisions. This justified a considerable effort in the setup of a Synthetic Environment to establish an adequate deck of virtual dashboards as a starting point.

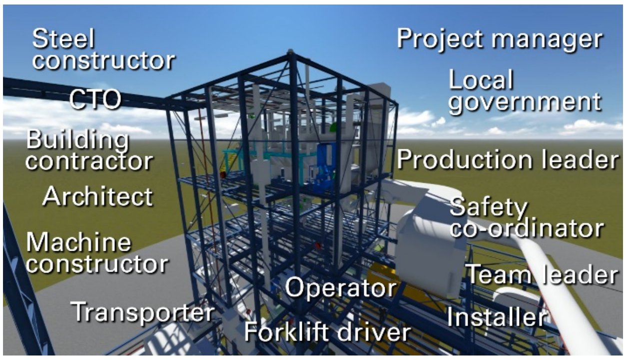


Figure 2 - Illustration of one of the case studies

In the project a digital master was established, bringing together the different types of models of the different stakeholders. As anticipated, amalgamating these models was far from effortless. However, even the processes involved in constructing the overall digital master already revealed inducements on how the virtual dashboard would contribute. The digital master distinguished between construction of the machine/production environment and the usage aspects of operating the plant. This steered the SE development cycle and resulting virtual dashboards. Figure 2 illustrates a section of the visualised digital master and some of the stakeholders involved.

Based on the virtual dashboard developed, several sessions were organised to have different stakeholders interact with the envisaged production environment. Partially scripted, the stakeholders received assignments or scenarios. Other session parts allowed for free interactions, with researchers observing stakeholders' interactions and the employment of different perspectives. A main outcome was that stakeholders indeed recognised issues in their own model/view more easily if stakeholders' perspectives are (implicitly or explicitly) overlaid. This resulted in conjointly identifying issues that would have complicated the building process and in uncovering at least one issue that rendered the to-be situation infeasible.

More striking even, however, was the observation that stakeholders implicitly formed clusters and implicitly and unknowingly started using virtual dashboards to formulate digital prototypes. Not only did they start exploring potential changes or improvements, but they additionally formulated missions for other stakeholders to address. In this, they scrutinised details, and implicitly explored previously unaddressed or unprepared territory. They started to specify and elaborate subsequent steps in the development cycle, giving rise to new sub-projects on e.g. reachability of locations and assets or costs of intermediate stock, each leading to different uses of the virtual dashboard. Moreover, digital prototypes were devised to compare different solution principles and alternatives. It was especially noteworthy that establishing virtual prototypes was hardly hampered by the differences in perspectives and fields of expertise of the stakeholders involved.

Next to the building of virtual dashboards by means of the Synthetic Environments and preparing/ moderating the sessions where the dashboards were used, specific effort in the project was directed in feeding back the learnings to the digital master. Here, a self-evident connection between the virtual dashboards and the structure of the digital master and digital prototypes became apparent, which was quite instrumental in capturing focus points for further elaboration and addressing the design rationale that underpinned relevant decisions.

7 CONCLUDING REMARKS

Establishing an architecture that interrelates the digital system reference, Synthetic Environments and virtual dashboards for pilot production environments is a long-term endeavour. Validating such an architecture is even more so. However, in considering the architecture as a continuously evolving entity, all experience and every case study can contribute, with specific dynamics and impacts. Also, the quick and significant changes in e.g. IoT technologies can be taken into account in the underlying digital twins. Likewise, advances in simulation techniques change the way in which digital prototypes interrelate with the overarching to-be digital master. This does not exempt this research project from making that architecture and the corresponding working methods explicit. However, as the project addresses pilot production environments that have significant lead times, such an architecture can not yet be presented in a convincing manner. Up to now, research on Synthetic Environments and experience gained in the case studies has demonstrated that there is an underlying rationale in establishing virtual dashboards. Also, from the case studies it is clear that virtual dashboards do have constructive and structured impact on development trajectories and underlying decision making. Consequently, ongoing research focuses on gaining more experience and transforming that experience in improved and broader underpinning of the methods used. There also is a focus on making the architecture explicit, but even more on rendering the approaches more instrumental and more accessible for industry – even if the architecture is (still) evolving.

8 REFERENCES

[1] Papazoglou, M.P., Elgammal, A., Krämer, B.J., 2018. Collaborative on-demand Product-Service Systems customization lifecycle, *CIRP Journal of Manufacturing Science and Technology*.

[2] Váncza, J., Monostori, L., Lutters, D., Kumara, S.R., Tseng, M., Valckenaers, P., Van Brussel, H., 2011. Cooperative and responsive manufacturing enterprises, *CIRP Annals,* 60(2), p. 797-820.

[3] Oude Luttikhuis, E.J., de Lange, J., Lutters, E., ten Klooster, R., 2015. Evolving Product Information in Aligning Product Development Decisions across Disciplines, *Procedia CIRP,* 29, p. 573-578.

[4] Lutters, E., 2018. Pilot Production Environments driven by Digital Twins. in *29th SAIIE Annual Conference*. Stellenbosch (SA).

[5] Macchi, M., Roda, I., Negri, E., Fumagalli, L., 2018. Exploring the role of Digital Twin for Asset Lifecycle Management, *IFAC-PapersOnLine,* 51(11), p. 790-795.

[6] Qi, Q., Tao, F., Zuo, Y., Zhao, D., 2018. Digital Twin Service towards Smart Manufacturing, *Procedia CIRP,* 72, p. 237-242.

[7] Tao, F., Zhang, M., Liu, Y., Nee, A.Y.C., 2018. Digital twin driven prognostics and health management for complex equipment, *CIRP Annals,* 67(1), p. 169-172.

[8] Kritzinger, W., Karner, M., Traar, G., Henjes, J., Sihn, W., 2018. Digital Twin in manufacturing: A categorical literature review and classification, *IFAC-PapersOnLine,* 51(11), p. 1016-1022.

[9] Kennon, D., Schutte, C.S.L., Lutters, E., 2015. An alternative view to assessing antifragility in an organisation: A case study in a manufacturing SME, *CIRP Annals,* 64(1), p. 177-180.

[10] Damgrave, R.G.J., Lutters, E., 2018. Enhancing development trajectories of synthetic environments, *CIRP Annals,* 67(1), p. 137-140.

[11] Damgrave, R.G.J., Lutters, E., 2018. The rationale of PSS as an inspiration for Synthetic Environments, *Procedia CIRP,* 70, p. 356-361.

9 BIOGRAPHY

Eric Lutters obtained his PhD. degree in Mechanical Engineering at the University of Twente, The Netherlands. At that university, he currently holds a position as an associate professor in the faculty Engineering Technology. He also is a professor extraordinary at the Industrial Engineering department at Stellenbosch University.

Jos de Lange obtained his PhD. Degree in Industrial Design Engineering at the University of Twente. Currently, he is assistant professor in the Department of Design, Production and Management (faculty Engineering Technology) at that university.

Roy Damgrave obtained his PhD. degree in Industrial Design Engineering at the University of Twente. Currently, he is assistant professor in the Department of Design, Production and Management (faculty Engineering Technology) at that university.

Digital Production Order Processing Support System Using Real Time Data

P. Schäfers[1], A. Mütze[1], P. Nyhuis[1]

[1] Institute of Production Systems and Logistics (IFA), Leibniz University Hannover, Germany.

Abstract

This paper presents a concept for a digital production order processing support system. The basic idea is to use production feedback data in real time to support the fulfillment of the tasks of production control. As a consequence, logistical performance can be improved and logistical cost can be reduced. The concept consists of the three modules data acquisition, data processing and information provision on the shopfloor. For each module, the fundamental points are discussed in theory and a solution for practical application is proposed. In summary, the concept helps to realise static control procedures and, beyond that, enables the implementation of dynamic control procedures that aim on real time optimization of logistical target achievement and use algorithms for this purpose. The potentials associated with those procedures are discussed. In order to validate the concept, it has been implemented in a learning factory. The experiences made are described.

Keywords

Production, Order Processing, Digitalization

1 INTRODUCTION

In the current age of digitalization, a number of technologies and data-driven concepts around key words like Industry 4.0, cyber physical production systems and smart factories are subject of discussion in science [1]. Some researchers do formulate the vision of self-control of production [2]. For producing companies, the ongoing technology-driven change is associated with the hope of increasing efficiency. Experts see the opportunity to reduce production costs by 10-20 % and to increase productivity by up to 50 % [3].

On the other hand, completely different challenges can be observed in industrial practice, especially with small and medium-sized enterprises. In some cases there is only little feedback data available from production and even the realisation of relatively simple control procedures fails [4]. The reasons mentioned for this are not only bad discipline and a lack of knowledge regarding logistical cause-effect relationships, but also inadequate communication and missing information to meet the requirements [4]. Mistakes in production control result in high costs for WIP, long throughput times and a bad due date compliance towards the customer.

Looking at digitalisation and data utilisation there seems to be a gap between visions and theoretical concepts on the one hand and practical implementation in industry on the other hand. For this reason, a simple concept is presented in the following that aims at supporting the flow of orders through production and making control decisions based on real time data.

2 FUNDAMENTALS OF PRODUCTION CONTROL

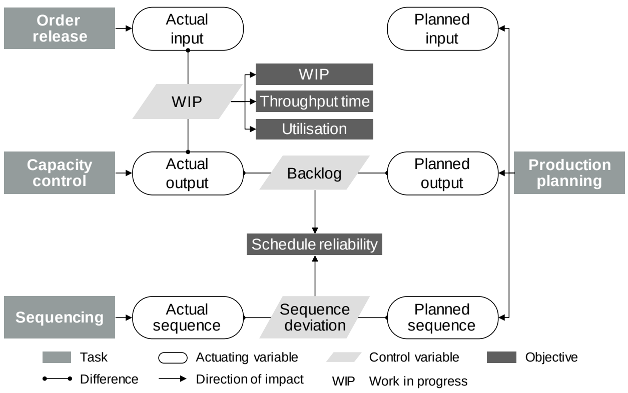


Figure 1 - Production Control Model [4]

Figure 1 shows the model of production control developed by Lödding [4]. The model is very well suited to explain the cause-effect-relationships in production control. In addition to a generally valid definition of the tasks of production control, a connection to the logistical objectives of production is established via actuating and control variables.

According to [5], the logistical objectives of production are schedule reliability, throughput time, capacity utilisation and WIP. Some of these variables run in opposite directions, so that simultaneous optimization is not possible. For this reason, companies must consciously position themselves in fields of tension created by opposing logistical objectives [6]. This requires knowledge of the cause-effect relationships, whereby logistical models such as the production control model create transparency.

The tasks of production control are order release, capacity control and sequencing at the workstations [4]. Order release influences the actual input of production. Capacity control, on the other hand, influences the actual output of production. The difference between these two variables is the WIP, which is an objective by itself and affects the objectives throughput time and utilisation [7]. The difference between planned output and actual output is the backlog. The difference between the planned sequence and the actual sequence is the sequence deviation. The backlog and the sequence deviations affect the objective schedule reliability [8].

3 THE CONCEPT AND ITS MODULES

As explained in the previous section, the fulfillment of the production control tasks influences the logistical target achievement. For this reason, a concept for a digital production order processing support system has been developed that supports the fulfillment of the tasks of production control and enables new approaches in production control. In order to validate the concept, it has been implemented in a learning factory.

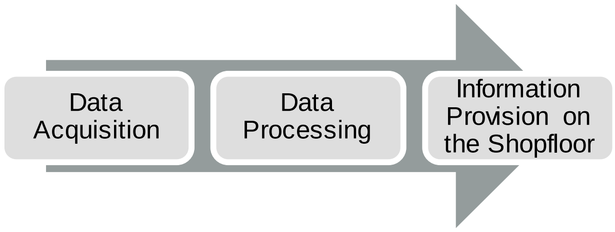


Figure 2 - Modules of the production order processing support system

Figure 2 shows the three modules of the concept. Those are data acquisition, data processing and information provision on the shopfloor. In the following, the individual modules will be presented in detail and the practical solution chosen for the prototype implementation in the learning factory will be presented.

4 DATA ACQUISITION

In order to control production on the basis of the current status of the system, it is necessary to depict this status as accurately as possible. In this context, two central questions have to be answered:

- Which data should be recorded at which points?
- Which technology should be used for data acquisition?

The recording of time stamps has proven its worth in order to depict the flow of orders through production [5]. The time at which an order passes a certain point in production or a certain event occurs (for example, the end of a processing operation) is logged. As a general rule, the more feedback data collected, the better the transparency. However, each data acquisition is also connected with expenditure, which must be justified by a benefit. If too much effort is required, the operatives carrying out the data acquisition may adopt a negative attitude, which may result in inaccurate execution of the feedback process. For each feedback point, it must therefore be weighed carefully whether it adds value for later evaluations and the derivation of control decisions. A distinction can be made between essential and optional feedback points, which is explained below.

Figure 3 shows the operation specific throughput element [9]. The throughput element defines the components of the throughput time of a production order at workstations. According to the definition, the consideration begins after the completion of processing at the upstream workstation. The product remains idle for a time period. Subsequently, the order is transported to the next workstation and remains in idle time until it is processed. In most cases the order will be part of a pool of orders waiting in front of the workstation competing to be selected. After selection, it may be necessary to set-up the workstation before processing can actually begin. The processing of the production order begins once the set-up is complete. This type of theoretical and generally valid depiction of the throughput of a production order may be used for the target-oriented positioning of feedback points in production systems.

The letters A, B, C, D, E and Z in Figure 3 mark possible feedback points. Feedback point E tracks the end of an operation. This feedback point is essential. Feedback point Z marks the beginning of the throughput element for this operation. According to the definition, time stamp Z is equivalent to time stamp E of the previous operation. Hence, time stamp Z must not be recorded separately, but can be copied. The feedback points E and Z already provide transparency about the throughput time of the individual operations. Feedback points A, B, C and D make the depiction of the production order throughput even more precise. Based on feedback point C it is possible to split the throughput time into the inter-operation time and the operation time. Feedback points A and B make it possible to further differentiate the inter-operation time. Feedback point D allows to distinguish between the setup time and the processing time.

For the prototype implementation in the learning factory the feedback points E, C and B have been selected. E is indispensable for determining which operations of an order have already been completed. In combination with the planned routing, the information can also be used to determine which workstation the order is currently located at. C

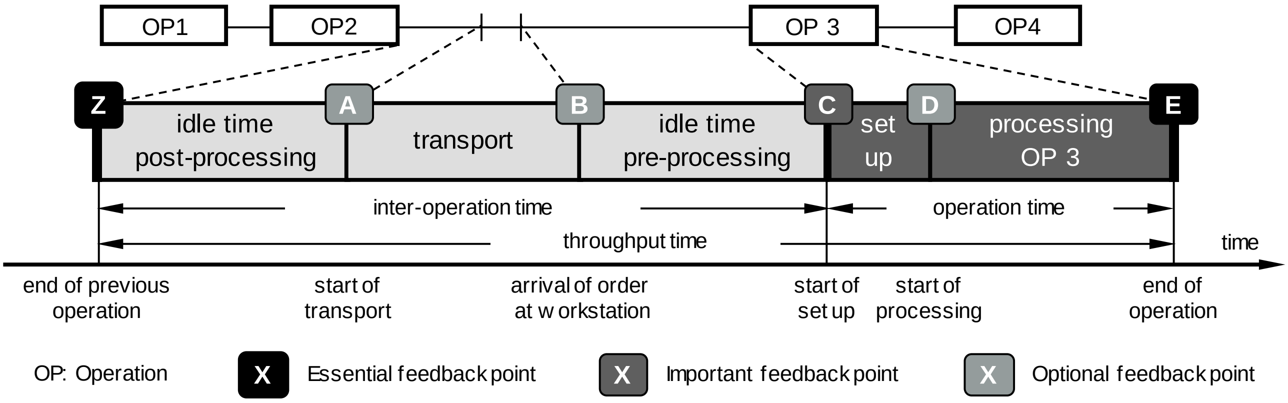


Figure 3 - Possible feedback points marked in the throughput element [based on 9]

enables you to identify which orders are waiting to be processed at a workstation and which order is currently being processed. B tells you whether an order has already arrived at the respective workstation or whether the logistician still has to take action. A is not recorded, because the transport distances in the learning factory are short and in this use case this feedback point does not add value for production control. D is comparatively difficult to record automatically and it does no add value for production control in the application under consideration as well.

To capture the time stamps at the feedback points RFID technology has been chosen. The advantage of this technology is that the time stamps are recorded with almost no operational effort for employees in production. In the prototype implementation in the learning factory, one production order is always linked to one loading box. All loading boxes are equipped with RFID cards. In addition, RFID scanners are installed at the selected feedback points in production. Whenever a loading box passes a feedback point, a time stamp is written into a central database.

The feedback data recorded following this approach is of high granularity and high quality. Granularity results from the fact that several feedback points are recorded at each workstation. Quality is based on the digital acquisition and real time transmission of information and the integration of the feedback points into the material flow. Consequently, this data provides an adequate foundation for processing the data and supporting production control.

5 DATA PROCESSING

With regard to data processing, two questions are central:

- Where should the data be stored?
- How should the data be further processed?

The decision was made, to store the recorded time stamps in a central database. In this way, the collected data is available to all users and there is transparency regarding the current state of production. All systems access this data and store their results there.

It is the intention of the overall concept to support production order processing. So this intention also applies to data processing. On the one hand, the basics of error-free production order processing must be ensured (for example the production of the right number of the right product or the right routing). On the other hand, the fulfilment of production control tasks should be supported (order release, capacity control, and sequencing). A focus has been put on the production control tasks of order release and sequencing. To simplify matters, the task of capacity control was not taken into account in the first step. The capacity of the system and the routing were regarded as rigid. Regarding the task order release, it was decided to implement the procedure of due date oriented order release (DDO) in the first step. Regarding the task sequencing, the procedures FIFO and Earliest Due Date (EDD) were selected. These procedures are widely used in industry [4]. Especially when DDO or EDD are used the advantage of the digital production order support system over printed order documents is evident. For example, the search for appointment information in the papers by the worker is no longer necessary. Potential sources of error are avoided. Please refer to [4] for an explanation of the procedures.

Various routines have been programmed. One example is checking whether production orders have reached their release time. Another example is the virtual sorting of production orders waiting to be processed in front of a workstation. A dynamic ranking list for processing is created for every workstation. The list contains all orders that are waiting in front of the workstation. With reference to Figure 3, this means that Feedback Point B has been captured. The orders are sorted according to a priority criterion. For FIFO this is the actual input date. For EDD this is the planned output date. The list is updated when an order enters the buffer or when an order is selected for processing and thus leaves the buffer. It is relatively simple and also planned to implement further sequencing procedures with their priority criteria.

Now the difficult question had to be answered how the information generated from the feedback data could be made available on the shopfloor. After all, a pure evaluation and processing of data does not bring any added value if the generated information does not reach its addressees.

6 INFORMATION PROVISION ON THE SHOPFLOOR

Concerning information provision the following three points have to be clarified:

- Who is the addressee of the information?
- What information should be made available?
- What technology should be used to provide the information?

The employees at the workstations and the employees of the internal transport were identified as addressees of information. Workers must be provided with all the information they need to process production orders. As mentioned before, the tasks of order release and sequencing are to be supported. The employees of internal transport are to be supported by the provision of routing information. So-called Electronic Shelf Labels (ESL) were selected as the technology for information provision. This decision has been taken because the ESL can be updated in real time via radio. They can display not only text, but also pictures. In addition, they are not limited to black and white, but can also display some colours. This allows highlighting information being particularly important at a certain point in time. Compared to other displays ESL are relatively affordable. Obviously, the costs are higher than with the use of printed order documents. Another unfavourable point is that the displays need a few seconds to update the information displayed.

Figure 4 shows the prototype implementation in the learning factory. There, each loading box is equipped with two ESLs - one for the worker and one for the logistician.

Figure 4 - Loading boxes with two ESL each

The ESLs can be labelled with different layouts. The choice of the layout depends on the addressee, the task and the PPC configuration (see Figure 5). The label for the worker shows the general information "order number", "product", "quantity" and "current station" as well as specific information to support the worker during order processing. In order to support the due date oriented order release, the due date is indicated by the label in addition to the general information. If the date is reached, the corresponding field turns red and thus signals to the worker that the order should be released into production. To support the task of sequencing, the orders waiting at a workstation are placed in a logical order by the system depending on the selected sequencing rule, displaying the respective order criterion. On the ESL associated with the most urgent job, the field with the order criterion is coloured red so that it is easy to select the appropriate job from the queue. The logistician receives other information on his ESL compared to the worker. These are the order number, the current station and the next station. With this information, the logistician can see on the ESL which route the loading box and the connected order should take through production.

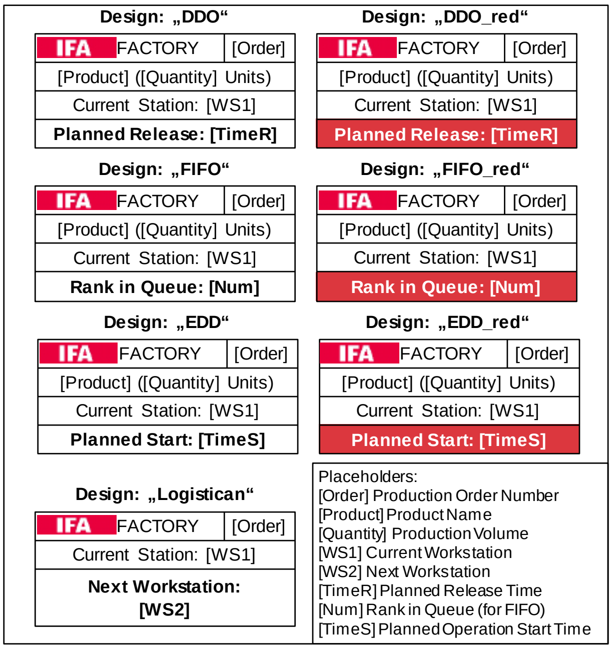


Figure 5 - Different ESL Designs

The update of the ESL is triggered by defined events. The ESL system and the RFID system are connected via the central database. Scanning a production order at a defined feedback point, triggers an update of the information shown on the ESL, taking into account the PPC configuration. Figure 6 shows the update process for the sequencing rule EDD. It shows which information the ESLs display after being scanned at different feedback points. If an order is scanned at the last feedback point (E) of one workstation (here: Milling 1), the time stamp is identical to the first time stamp (Z) of the next workstation (here: Milling 2). While the loading box waits for its transport to this workstation, the ESL for the worker is labelled with the information required at this workstation. At the same time, the system automatically compares the current order with all other orders waiting for

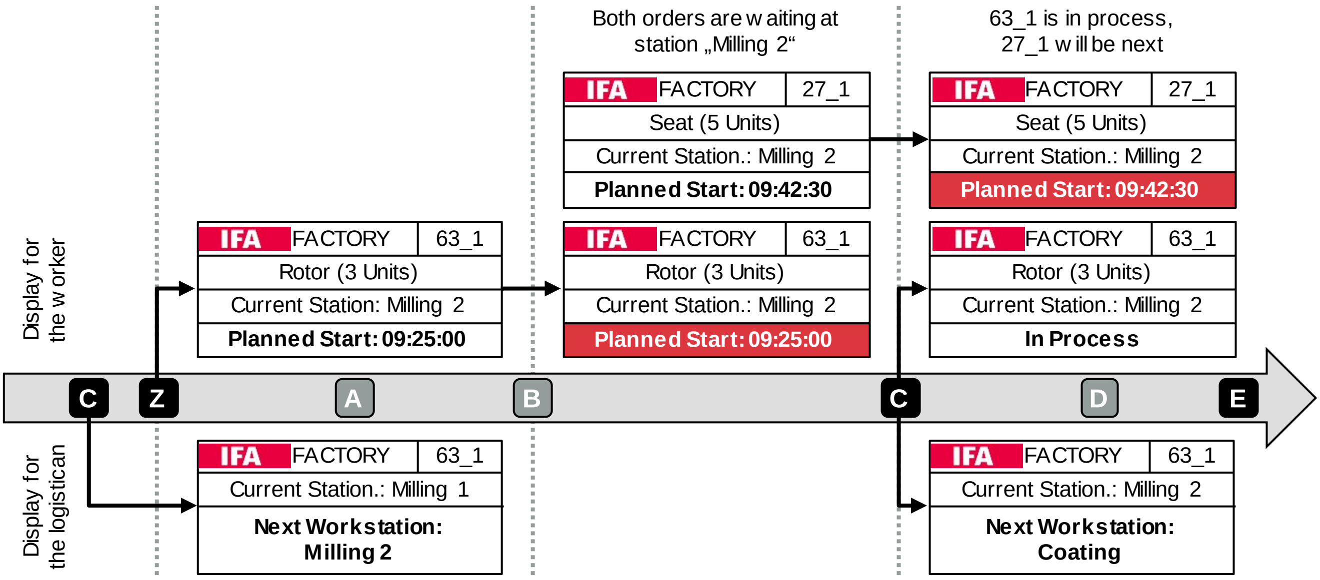


Figure 6 – ESL update process triggered by RFID scans at feedback points

processing at the workstation and checks which of all waiting orders has the earliest planned start date, whose ESL is coloured red. In Figure 6, this is order 63_1 with a planned start date at 09:25:00. The next feedback point triggering an update of the ESLs is point "C", which records the start of processing. Here the ESL for the worker as well as the ESL for the logistician are updated. The ESL for the worker is set to the status "In Process", so that it is always clear which order is in process and which orders are still waiting for processing. At the same time, the system checks all remaining orders waiting for processing and determines the one with the earliest planned start date, whose ESL is coloured red. In Figure 6, this is order 27_1 with a planned start date at 09:42:30. The ESL for the logistician is also synchronised and labelled with the routing information for the transportation process. This step was deliberately placed on feedback point C and not on feedback point E, as the update process of the ESL takes some time and by using this point, the update process is definitely completed when processing of the order is finished at the workstation. If processing of the order is completed at the workstation and the order is detected at the feedback point "E", the ESL for the worker is labelled for the next workstation and the cycle begins again.

7 THE SYSTEM ENABLES DYNAMIC PRODUCTION CONTROL

In addition to the previously depicted implementation of static control procedures, the system also enables the implementation of dynamic approaches that lead to control decisions made on the basis of the current system status. Firstly, this requires the exact and real time depiction of the respective system status (e. g. WIP in the system or backlog). Secondly, the recorded data must be evaluated in real time and converted into control decisions on the basis of defined rules. Thirdly, the information regarding the control decisions must be made available on the shopfloor in real time and in a manner appropriate to the addressees. All these requirements are met by the system presented in this paper. Figure 1 makes it clear that the three tasks order release, capacity control and sequencing are differentiated in production control. For the tasks order release and capacity control different procedures can be found in literature and are widely used in industry which propose control decisions on the basis of the current system status [4]. For order release, these are the procedures that regulate the WIP, such as CONWIP, Workload Control or Load Oriented Order Release [4]. In capacity control, this is mainly the Backlog Oriented Capacity Control [4]. Here, however, it should be questioned whether these procedures can be upgraded using the very high data granularity made possible by technologies such as RFID and a high number of feedback points along the order flow through production. Looking at sequencing, another picture emerges. The methods commonly used in practice, such as FIFO, EDD and setup-optimized sequencing [4], take into account the production orders waiting in front of a workstation, but not the overall status of the system. In line with this observation, several researchers are currently working on approaches for dynamic routing and dynamic sequencing. In particular, adaptive systems as well as learning algorithms (e.g. reinforcement learning) are tested and implemented in companies [10] [11]. The approaches are based, among other things, on the tool of simulation and use the improved quantity and quality of the feedback data. Dynamic routing and dynamic sequencing are also seen as possible extensions of the system presented in this paper. Dynamic routing requires short-term decisions regarding the path of orders through production. Since only the current and the next workstation are shown on the ESL for the logistician this could be realized trouble-free. Dynamic routing allows the distribution of orders to capacities to be done dynamically and the load situation to be smoothed out. Dynamic bottlenecks that occur, for example, due to malfunctions or necessary rework, can be avoided. Concerning sequencing decision could be made on the basis of real-time information. For example, orders could be accelerated if they are particularly urgent for a customer or a downstream workstation threatens to run empty. In return, orders that would have to wait at the next workstation anyway could be delayed. It is also conceivable that setup processes could be eliminated and, thus, productive performance could be raised, since there is transparency which orders are waiting in front of a workstation and which orders will reach the workstation in the near future. In any case, such approaches for dynamic sequencing could be put into practice with the system presented in this paper.

8 CONCLUSION AND OUTLOOK

This paper presents a concept for a digital order processing support system. The concept has been developed by a coherent consideration of the three modules data acquisition, data processing and information provision on the shopfloor. For each module, a solution for practical application is proposed. This comprises data acquisition using RFID technology, data storage in a central database, data processing through defined routines and the dynamic user and situation-specific provision of information regarding the production orders on the shopfloor using ESL. For validation, the system has been implemented in a learning factory. In each training session in the learning factory, several game rounds are carried out in which the same set of customer orders is processed. KPIs such as the number of completed orders or schedule reliability are recorded. The PPC configuration is improved step by step starting from a chaotic state. The success of measures can be evaluated on the basis of the KPIs. The comparison of KPIs recorded in game rounds with and without the use of the system shows that order processing can be effectively supported because, for example, premature or late release of orders into production or sequence deviation can be prevented. In addition, the training participants rated the limited and specific provision of information via the ESL very positively. Beyond that, it is intended to test an adapted version of the digital production order processing support system in industrial practice. Talks are currently underway with a company that produces customer-specific precision components. A big barrier for usage in industrial practice is the creation of links to established IT systems. In the future, further PPC procedures should be implemented in the system. Potential and research needs are seen with regard to the development of algorithms for dynamic production control - especially for dynamic routing and dynamic sequencing. In summary, the concept helps to realize static control procedures and, beyond that, enables the implementation of dynamic control procedures. A closed loop has been established between data acquisition on the shopfloor, IT-based data processing and the implementation of measures back on the shopfloor.

9 REFERENCES

[1] Liao, Y., Deschamps, F., Loures, E. D. F. R., Ramos, L. F. P., 2017, Past, present and future of Industry 4.0, Int. Journal of Production Research, 55:12, 3609-3629.

[2] Kagermann, H., Wahlster, W., Helbig, J., 2013, Umsetzungsempfehlungen für das Zukunfts-projekt Industrie 4.0, Forschungsunion.

[3] Bauernhansl, T., 2014, Die vierte industrielle Revolution, In: Bauernhansl, T., ten Hompel, M., Vogel-Heuser, B. (eds.), Industrie 4.0 in Produktion, Automatisierung und Logistik, Springer Vieweg, Wiesbaden.

[4] Lödding, H., 2013, Handbook of manufacturing control, Springer, Berlin, Heidelberg.

[5] Wiendahl, H.-P., 1997, Fertigungsregelung, Carl Hanser Verlag, München, Wien.

[6] Schmidt, M., Schäfers, P., 2017, The Hanoverian Supply Chain Model, Production Engineering, 11 (4-5), 487-493.

[7] Nyhuis, P., Wiendahl, H. P., 2006, Logistic production operating curves, CIRP Annals, 55 (1), 441-444.

[8] Lödding, H., Nyhuis, P., Schmidt, M., Kuyumcu, A., 2014, Modelling lateness and schedule reliability, Production Planning & Control, 25 (1), 59-72.

[9] Heinemeyer, M., 1974, Die Analyse der Fertigungsdurchlaufzeit im Industriebetrieb, Offsetdruck Böttger, Hannover.

[10] Frazzon, E. M., Kück, M., Freitag, M., 2018, Data-driven production control for complex and dynamic manufacturing systems, CIRP Annals, 67 (1), 515-518.

[11] Kuhnle A., Lanza G., 2019, Application of Reinforcement Learning in Production Planning and Control of Cyber Physical Production Systems, In: Beyerer J., Kühnert C., Niggemann O. (eds.), Machine Learning for Cyber Physical Systems. Springer Vieweg, Berlin, Heidelberg.

10 BIOGRAPHY

Philipp Schäfers (*1987) studied Mechanical Engineering at Karlsruhe Institute of Technology. Subsequently, he completed a MBA programme. Since 2014 he is working as a research associate at IFA of Leibniz University Hannover.

Alexander Mütze (*1994) studied Engineering and Business Administration at Leibniz University Hannover. Since 2018 he is working as a research associate at IFA of Leibniz University Hannover.

Peter Nyhuis (*1957) is Professor and heading the Institute of Production Systems and Logistics (IFA) of Leibniz University Hannover since 2003.

Digitalization in Quality Assurance

R. Jochem [1],

[1] Chair of Quality Science, Institute for Machine Tools and Factory Management,
Technische Universität Berlin, Germany

Abstract

The increasing amount of sensor data generated in production systems together with the ever-growing interconnectedness of all systems result in new opportunities for quality assurance. Process data can be accurately monitored, analyzed and connected to assess product quality already during the production process. While being announced as a theoretical concept, the practical transition of Industry 4.0 to autonomous flexible production plants with complete data exchange between all components is now in full swing. But the technical changes provide far more opportunities than just modernizing the classical schemes, it opens the door to a whole range of digitally assisted new ways to rethink production. A small micro factory has been built, that incorporates all aspects of Industry 4.0 as a demonstrator and creates the opportunity to test and study new concepts in a safe environment as well as providing the platform for teaching concepts. The contribution will outline the history and further development of our demonstrator, draw the connection between classical and future production and show novel approaches from process data to self-optimization with focus on product quality. A focus will be laid on data analysis and the steps from process data to knowledge of the product quality. Furthermore, assistance system named QS-Services is presented to support the user in understanding process data. Following the Define-Measure-Analyze-Improve-Control Methodology, the application guides the user through the individual steps. Upon completion, participants will be able to describe novel approaches to quality assurance in industry 4.0.

Keywords

Digitalization, Quality Assurance, Industry 4.0

1 INTRODUCTION

Quality, time and money are essential factors to ensure the competitiveness of businesses. In order to maintain the top position of German industry among the industrialized countries, a work group being instituted by the German government announced the fourth industrial revolution in 2013 [1], [2]. The aim is to increase the rate of innovation, therefore enhancing the development of new processes, technologies and business models. Today, Industry 4.0 is one of the decisive drivers of numerous developments in Germany and beyond [3], [4], [5].

With the digitalization and networking of production systems, the amount of data available for quality assurance and quality management increased immensely [6]. Assistance systems offer the possibility of intuitive and efficient control of production processes through acquiring, analysing and evaluating data in real time. In the long term, a fully networked processes chain together with an integrated management system will optimize production machines in real time [7].

Industrial production is highly automated today, there are correspondent sensor data available for each regulation. Furthermore, the interconnectedness of machines is growing more and more, more and more data are generated, which are available at central places [8].

Instead of using the data only for the maintenance of the production they could also be viewed out of the perspective of quality assurance. Through real-time observation information about the quality of a product can be achieved already during the production process as well as quality knowledge is generated then. This will actively be led back into the process, because a qualitatively high result will be achieved with the regulation of the process parameters. Learning algorithms are installed, which adapt the process flow automatically to an ideal result. At the same time deviations are identified early and waste and downtime are minimized. Alongside long-run analysis and pattern recognition give evidence about the machine´s condition, downtimes can be avoided through intervention in time.

Since the 80ies the catapult has been established as standard learning instrument for the acquisition of the Design of Experiments (DoE) methods and got worldwide fame. Because of the broad acknowledgement and acceptance even of deciders the catapult is extremely well suited for the implementation of the new approach of Quality in the Industry 4.0 and to accompany and support the change of industry.

2 SMART MICRO FACTORY AS ENABLING INFRASTRUCTURE FOR DIGITAL QUALITY ASSURANCE

To study the changes and opportunities of Industry 4.0, a completely autonomous production plant has been built on a lab scale. The small micro factory (SMF) incorporates all aspects of Industry 4.0 as a demonstrator and creates the opportunity to test and study new concepts in a safe environment as well as providing the platform for teaching concepts.

The experimental setup to produce valid and reliable answers while also optimizing the cost to benefit ratio is given by the so-called design of experiments method. To teach this, the catapult experiment is well established. In our lab setup, the classical catapult, where the experimenter tries to throw an object a given distance by manually varying force, angle and length of the catapult arm, has been transferred into the fully automated cyber-physical system 'catapult 4.0'. An app guides the user through the necessary steps and monitors his progress, while the experiments are performed automatically. A camera system tracks the trajectory of thrown object and the flight stability is analysed. The landing position is accurately measured by a rotating laser scanner and the object is picked up and placed back into the catapult by a five axes industrial robot arm mounted on a linear axis.

Around the setup of this experiment, a modular and autonomously working manufacturing plant has been constructed to produce the object that is thrown. It consists of 3d printers, a 3d scanner and a tumble finisher. As with the experiment design parameters, an app guides the customer through the process of designing an object and helps deciding on its quality in terms of accuracy versus production time. The small micro factory handles the individual production process autonomously. During the entire time, all process parameters are accurately tracked, and quality of the outcome is being assessed.

For instance, layer height is a dominant factor for product quality in the 3d printing process, the first step in our production line. If sensors detect an error, it is possible to analyse the cause and effect in real-time using the help of statistical algorithms. In the future, all sensor data will be analysed, and cross referenced with known data patterns. Should patterns occur, that indicate deviations from the target quality, the smart factory can adapt the production process and either correct the edges by including the tumble finisher in the production process or reprint the part. If data analysis does not result in a known case, the surface quality of the produced object is measured in the 3d scanner using the structured light technique. This comparison between target and actual quality is enabling the factory to learn the connection between quality and process parameters, during future operations similar data patterns can be linked to states of quality.

The transport of the object between the individual machines is managed by the robot arm on the linear axis. To create the link between printed object and digital information, every object is stored on a work piece holder. These holders use RFID tags to store information about previous and planned steps in the production cycle and create the link between physical object and digital footprint allowing for smart products. The smart micro factory consists of separate production units. By plugging in the connectors of the modules into a central module, this establishes the capabilities of the production line and adapts the production orders (see figure 1). Through constant data analysis, the smart micro factory is optimizing the quality outcome of every individual step as well as the overall performance of the entire production process.

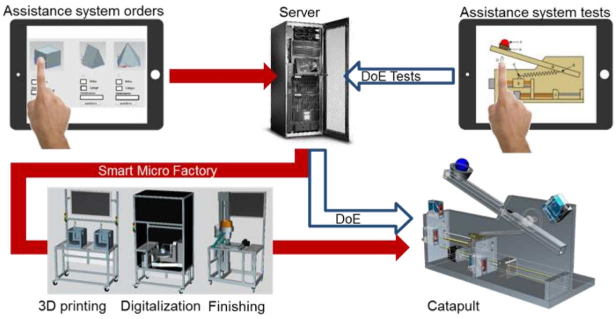


Figure 1 – Production Process within the Smart Micro Factory.

Alongside the automated production, the quality engineer receives information on his mobile device enabling him to react and improve the chain of production or to build the basis of communication to relevant persons. Quality dashboards demonstrate the links between process data and predicted outcome. While being a project in itself to demonstrate the next steps in the future of production, the smart micro factory has become the basis for other projects and collaborations that build on its opportunities, as well as training courses in an industry 4.0 environment.

A focus will be laid on data analysis and the steps from process data to knowledge of the product quality. Furthermore, practical implications and future projects are being presented. Upon completion, participants will be able to describe novel approaches to quality assurance in industry 4.0.

3 QS-SERVICES SUPPORTING DIGITAL QUALITY ASSURANCE TASKS

Accordingly, an assistance system named QS-Services has been developed to support the user in understanding process data (see figure 2).

Following the Define-Measure-Analyse-Improve-Control Methodology [9], the application guides the user through the individual steps. By asking more and more specific question, the user interactively characterizes a problem in the define phase. Adding additional parameters, such as tolerance limits have a positive influence on the algorithms and enable more specific recommendations.

During the development of QS-Services, experimental tests were conducted to determine and correlate significant influencing factors on product quality. In the measure phase the user is assisted in defining relevant sensors and tracking sufficient process data. Statistical algorithms as well as methods for quality improvement form the basis of the analyze phase in QS-Services. To improve the production process, a detailed report is containing in depth analysis of previously entered issues and recommendations to solve identified problems in future production cycles. By this, the responsible quality engineer or machine operator is enabled to adjust the process settings at an early stage and actively influence quality in a positive way.

It is to be emphasized that the results of the implementation of measures in QS-Services are documented and stored for further usage, forming the control phase. Therefore, a control loop is implemented, that enables the system to learn continuously and thus derive targeted preventive measures even more precisely.

QS-Services provides a framework to identify causes and derive measures in a transparent way, even without in-depth knowledge in the field of statistical data evaluation. In particular, the analysis of data correlations with regard to product quality is a major challenge. To be able to make statements about the product quality, laboratory tests have been conducted and methods developed in order to derive statements about interaction of process parameters. QS-Services differentiates between the analysis of individual parameters and parameters in correlation to find solutions within increasingly complex production environments.

Due to the findings of QS-Services, it is possibly to intervene directly in the production process. Complaint and error management costs are lowered, and organisations are enabled to further increase their competitiveness through introducing targeted preventive measures to eliminate errors. The assistance system describes a first step towards predictive quality [10].

In a sample implementation the assistance system has been integrated in an Industry 4.0 production environment. The Quality Science Lab (QSL) [11] is a production system consisting of independent cyber-physical units. As key technology in Industry 4.0 an off-the-shelf 3D-printer has been equipped with additional sensors and data processing units. Prints of low quality are identified and process parameters are analyzed with the help of QS-Services. Further implementations with industry partners are in preparation.

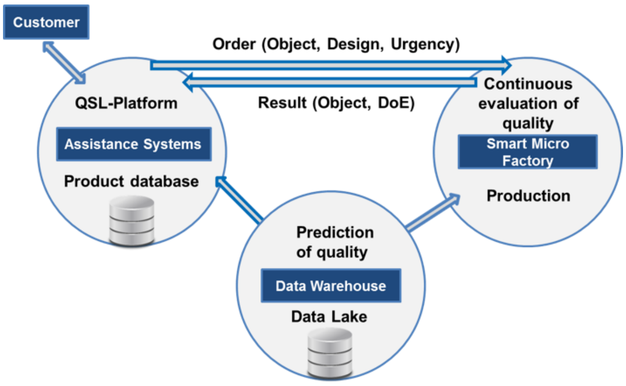


Figure 2 – QS-Services provided by an assistance system in the QSL-Plattform.

Presently QS-Services focuses on statistical methods, in next steps machine learning technologies will be implemented. Comparison of this classical and modern approaches will lead to insights about applying data science techniques like cluster analysis and neuronal networks in a production environment. Based on the current research, QS-Services provides a framework to be further developed and to encompass ever more complex scenarios. With the implementation of further technologies and their continuous improvement, the reliable prediction of product quality is possible. With the realization of industry 4.0 in a comprehensively networked, digitalized production environment this will give quality engineers the opportunity to intervene and adapt process parameters already during production process. Through monitoring and analyzing process parameters, negative outcomes can be avoided and all customer requirements can be fulfilled.

4 CONCLUSIONS

The intelligent analysis of real-time print data enables automated quality-related decisions to be made whilst the company is in operation. The mere observation and interpretation of the process parameters alone allows for sound statements about the quality of the product. The quality is produced and predicted in this way. However, this is only one of the possibilities of quality-related activities in industry 4.0. Considerable developments can be expected in this environment.

During the implementation of the sample applications in the QSL it became apparent that there are no standards for the communication between the systems as well as for data transfer and intelligent data analysis. As a result, the communication interfaces and data analyses in the sample application had to be individually programmed and implemented. This is certainly an inhibiting factor for a broad application in industry, although its configuration opens up great potential. An extensive view must also be taken of the changes in quality management in manufacturing companies due to the quality knowledge generated in real time for product development. The changes that will result from this form a further research focus of the Department of Quality Science at the TU Berlin, in addition to the change in quality assurance through intelligent and automated data analysis. The demonstrator of the Quality Science Lab will provide support in all these areas. He serves as a platform for the development of various concepts and methods of quality in industry 4.0 and furthermore makes research tangible and visible.

5 REFERENCES

[12] Adolph, A. et al., 2015, Deutsche Normungs-Roadmap, Industrie 4.0, Version 2, DIN e. V.:8

[13] Plattform Industrie 4.0, 2018, online available from http://www.plattform-i40.de/I40/Navigation/DE/Plattform/Plattform-Industrie-40/plattform-industrie-40.html, accessed 11th December 2018

[14] Kagermann, H.; Wahlster, W.; Helbig, J., 2013, Deutschlands Zukunft als Produktionsstandort sichern Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0 Abschlussbericht des Arbeitskreises Industrie 4.0, acatech – Deutsche Akademie der Technikwissenschaften e.V.:1

[15] Bauernhansl, T. et al., 2014, Industrie 4.0 in Produktion, Automatisierung und Logistik, Springer Fachmedien Wiesbaden:9

[16] Roth, A, 2016, Einführung und Umsetzung von Industrie 4.0, Springer-Verlag Berlin Heidelberg:3-6

[17] Alexandro, A. et al., 2015 , Implicit Parallelism through Deep Language Embedding, Conference on Very Large Data, Hawaii:35-46

[18] Kiem, R.: Qualität 4.0, 2016, QM, MES und CAQ in digitalen Geschäftsprozessen der Industrie 4.0, Carl Hanser Verlag GmbH & Co. KG:5-6

[19] DIN und IIC kooperieren bei Industrie-4.0-Standards, online available from https://www.qz-online.de/news/normen-richtlinien/artikel/din und-iic-kooperieren-bei-industrie-4-0-standards-1204118.html, accessed 11th December 2018

[20] Jochem, R., Herklotz, H., Geers, D., Giebel, M., 2015, Six Sigma leicht gemacht: Ein Lehrbuch mit Musterprojekt für den Praxiserfolg, Symposion Publishing:24

[21] Hagerty, J., 2016, 2017 Planning Guide for Data and Analytics, Gartner Inc.:13

[22] Seiffert, K., et al., 2016, Traditionelles neu gedacht: Das Katapult als Lernzeug für die Qualität in der Industrie 4.0, published in Refflinghaus, R; Jochem, R., Kassler Schriftenreihe Qualitätsmanagement, Fachgebiet Qualitäts- und Prozessmanagement an der Universität Kassel, Band 6:43-59

6 BIOGRAPHY

Prof. Dr.-Ing. Roland Jochem is Professor for Quality Science at the Technical University of Berlin (TU Berlin). After working as Head of the Department 'Process Management' at Fraunhofer IPK, he moved to Bosch & Siemens Home Appliances as a Process Manager. Before his appointment as Professor for Quality Science at TU Berlin in 2010, he was the holder of the Quality Management Chair at the University of Kassel since 2005. He is also Extraordinary Professor for Quality Management at the University of Stellenbosch in South Africa since 2009. He has authored over 70 publications, including 8 books as editor/author.

Berlin Institute of Technology, Pascalstraße 8-9, 10587 Berlin

roland.jochem@tu-berlin.de

Analytical User Story Clustering supporting Requirements and Synthesizing Efforts for the Digital Transformation

T. Gittler[1], R. Lorenz[2], L. Weiss[1], K. Wegener[1]
[1]Institute of Machine Tools and Manufacturing (IWF), ETH Zurich, Switzerland
[2]Production and Operations Management (POM), ETH Zurich, Switzerland

Abstract
The digitalization of manufacturing processes, also related to as the digital transformation, poses a large number of challenges to manufacturing companies. As digitalization projects aim at highly integrated processes and data flows, companies carry out extensive analyses and considerations. Since isolated solutions undermine the full harvest of benefits and integration ambitions, the development and roll out of digitalization solutions require the involvement of many parties across many different hierarchical levels. However, especially larger corporations still face the issues of developing redundant data storages and solutions, or missing substantial aspects in the requirements engineering of large scale digitalization projects. As a part of agile development efforts, user stories have proven their effectiveness for expressing and comparing functions. In this study, we present a user story centered approach to the collection of requirements and synthesizing efforts of digital transformations. It embraces the findings of the previously applied brownfield redundancy recognition principles, and extends them to an asynchronous multi-site requirements collection of a large corporation's greenfield approach. This enables the involvement of a larger number of collaborators than for conventional requirements engineering techniques. Furthermore, the approach allows for a data-driven analysis and categorization of collected requirements into appropriate categories according to overall business goals. The approach is applied in a case study to develop a manufacturing execution system within several plants of a large manufacturing company.

Keywords
Digital Transformation, Requirements Engineering, Manufacturing Execution Systems, User Stories

1 INTRODUCTION

The introduction and setup of integrated and continuous data flows in manufacturing companies, also referred to as the digital transformation, poses a challenge for small and large manufacturing enterprises alike [2,10]. Isolated or stand-alone tools connectivity or software solutions do not manage to harvest the full benefits of digitalization ambitions [14]. Therefore, the digital transformation often implies comprehensive and encompassing process rollouts for full coverage of related data streams. In order to consider and align all the concerned endpoints for data exchange throughout an entire organization, profound and embracing analyses are necessary. This consideration of various processes, machinery installations and interfaces requires the involvement of several interdisciplinary units and personae across different organizational levels, which can be a cumbersome task to organize and execute. This holds especially true for enterprises organized in locally separated units and production plants. The challenge thereby becomes the extraction of requirements from different divisions of an organization to develop a large-scale digital solution without creating redundant functionalities.

In the underlying paper, a user stories centered approach to the collection of requirements and synthesizing efforts of digitalization transformations is presented. We thereby build on existing theory and extend the redundancy recognition principle discussed by Lorenz et al. [7] to a greenfield approach. We then apply this approach in a case study for the implementation of a manufacturing execution system (MES) at a large machinery equipment manufacturer. The resulting clusters of requirements help to organize and aggregate user stories for further communication, evaluation and quality assurance purposes in the digital transformation process.

2 THEORETICAL BACKGROUND

Digital transformation receives a high priority within most of the strategic discussions of manufacturing companies [6]. Firms need to develop new capabilities and skills to implement digital solutions. The development of such solutions shares rather common attributes with software development than with traditional engineering practices: for instance, to enable digital solutions in production, agile development practices have become a best practice [4]. Digital transformation connects digital assets such as machines with each other (horizontal) and provides access to this data (vertical). However, as this bidirectional connectivity requires the involvement of many experts from different disciplines, traditional software development practices cannot overcome the challenges alone. Yet, it can certainly provide fundamental approaches.

2.1 Digital Transformation & Digital Factory

The vision of the digital factory is the all-time, overall availability of all necessary and relevant information for each and every specific person and situation, which may include real-time digital replications of shop floor operations and transactions, allowing to preempt, simulate and virtualize existing or future processes [14]. The potential benefits of the digital factory are leaner operations, rapid setups, and swift accommodation of product or process changes in manufacturing [3]. This comes with the issues of the connection between processes and structures both horizontal and vertical, the commitment and contribution of all concerned and relevant parties, as well as an extensive network to span around people, assets and products [2]. The literature sometimes refers to this phenomenon as the smart factory, however, the terms digital factory and smart factory are not synonymous. Contrary to the digital factory describing mainly connectivity topics, the smart factory not only encompasses data transfer protocols and data representation and semantics, but also the (automated) understanding and analysis of the data [12]. As the presented approach focuses on the overall assessment of needs and solution development for a thorough information flow, we will incorporate mainly aspects of the digital, and limited aspects of the smart Factory. The digital factory and the notion of MES are closely linked. The digital factory is the way to permanently optimize operational processes in the factory through real-time availability of all relevant data and information to all units and persons involved [5]. The MES is at the core of this effort, as it makes use of data in order to support and execute critical decision making processes based on information from products, machines and workers.

The Digital Factory and its MES systems are no off-the-shelf standard products, representing a challenge especially for specialized for Small and Medium Enterprises (SMEs). One of the main reasons is an incomplete understanding of their future production ambitions [10]. This might be due to a limited assessment of both, the current production-related problems and future challenges and ideas, which can be a cumbersome and time-consuming process - during which challenges and priorities may shift. The present study suggests a straightforward and quick approach to determine the current situation and to express future needs within a single framework. This allows to incorporate different units and levels of an entire manufacturing organization.

2.2 User Stories for agile requirements engineering

Requirements engineering deals with the collection and analysis of customer requirements for a product, service, or process [16]. Agile requirements engineering is subfield that aims at a fast and flexible elicitation process. Most of the agile requirement projects apply artifacts to collect their requirement, under which user stories are the most common [11]. User Stories are an effective way to express requirements intuitively and with little formalization. According to [15], they are composed of an identifier (who), a function request (what), and the target of the request (why). The following form based on Wirdemann is not an official definition, but will be used as a basis for user interaction in the following sections: “As a user, I want some sort of function, so that I achieve some sort of target”. This form provides little structure, but is simple, convenient to convey and easily recordable. User stories can be applied to diagnose existing solutions, respond to challenges, and to express ideas for future applications [7]. However, this implies the postponement of complexity to the analysis phase, when all user stories are examined and processed.

The origin of user stories stems from software and application development. Especially in agile development operations, user stories are employed to gather the necessary customer insights for the development of adequate solutions [8]. Even though [1] describe user stories as a high level approach in agile requirements engineering, the digital transformation can adopt agile principles, as the digital transformation exhibits similarities to software development. Yet, the focus lies on supplier and function selection, on the assurance of overall information coverage, as well as on the avoidance of redundancies or heterogeneous solutions. It is therefore an approach comprising low level to high level requirements.

The diagnosis of existing solutions, as well as the application in digital transformation cases in manufacturing has been proven and described in detail by [7]. The authors suggest a three-step brownfield procedure starting with the elicitation of the user stories, similarity determination, and road mapping. In the elicitation step, user stories are generated from different sources. The collected user stories are then labeled with the Key Performance Indicators (KPIs) or target improvements they affect, which allows to determine similarities between different user stories analytically.

For the analysis of similarities, there are different algorithms allowing to calculate distances between two objects, where small distances imply high similarity. The Euclidean metric is used for distances in multidimensional spaces. It is rotation invariant and hence implies linear dependence of all attributes, which is not a suitable representation in all cases. [7] chose a Jaccard metric, as it attributes a higher importance to factors equaling zero (compared to the Euclidean metric), and additionally allows to process categorical besides binary variables. However, Jaccard is a similarity index, of which the inversion is not a valid distance expression for all possible cases. Moreover, Jaccard is not suitable for non-binary values. Even though the approach by [7] performs well in checking for similarities and filtering redundancies, it is not applicable to all types of vectors. Hence, we see this development of an overall structure of unstructured requirements, as well as the extension to a greenfield approach, the research gap for this study.

3 USER STORY FRAMEWORK

The following sections present a user-stories based collection of requirements and subsequent processing within a digital transformation project to establish an MES. Both the theoretical fundamentals, as well as their application and outcome in an actual digitalization approach for a global machinery equipment manufacturer are discussed.

The case study takes place at a global manufacturing company situated in central Europe. The case company decides to introduce an MES for its multiple sites across different countries and continents. The overall goal is to enable decisions based on full information at any given point, location and organizational level in real time. This includes many activities, most notably order information provision, production progress feedback, machining and quality measurement protocols, material track and trace, digital worker signatures, alerts, as well as asset health. This requires in return a profound analysis for subsequent streamlining of the currently used IT systems, interfaces and applications. However, the asynchronous evolvement of processes across the different manufacturing sites requires a mutual effort to cover all particularities. The challenge is therefore to find a standardized, asynchronous and potent approach allowing to collect and compare a large number of different requirements from a large number of key players.

Following [7], we chose a user stories based approach in order to collect, analyze, categorize and communicate a maximum set of expectations and needs. This enables a broad participation and subsequent treatment of a large number of records.

3.1 Collection

The collection of user stories is challenging, especially due to the involvement of many parties. For the present application case, this held especially true, as the elicitation process comprised multiple geographically separated production sites. The effort associated herewith however yielded different favorable outcomes in digital transformation processes: The information inflow of an indefinite number of key players, as well as the early communication to a large number of parties in the transformation process. Moreover, it provided an unbiased view of all parties towards future digitalization ambitions.

In order to accommodate these aspects, a mobile-optimized web interface (Figure) was developed and made accessible to all members of the organization across different sites. The web interface consisted of four fields, in which users could insert records which were subsequently stored on a web server for subsequent analysis (see Figure). A quick briefing to all relevant key players as well as an access link via a QR code was provided. The key players stemmed from all units affected by factory operations, such as quality management, production controllers, assembly line workers, work plan engineers, machine operators, IT application owners, process engineers, product developers, sales and directors. For an easier labeling and clustering process of the data, we divided the considered hierarchical levels into five roles: Workers, Production Controllers, Team Leaders, Value Chain Responsibles and Executives. In addition to the common three elements of a user story, users of the web interface had to select the appropriate functional area of an MES according to their function. This selection of entries was based on the Verein Deutscher Ingenieure (VDI) guideline 5600 [13].

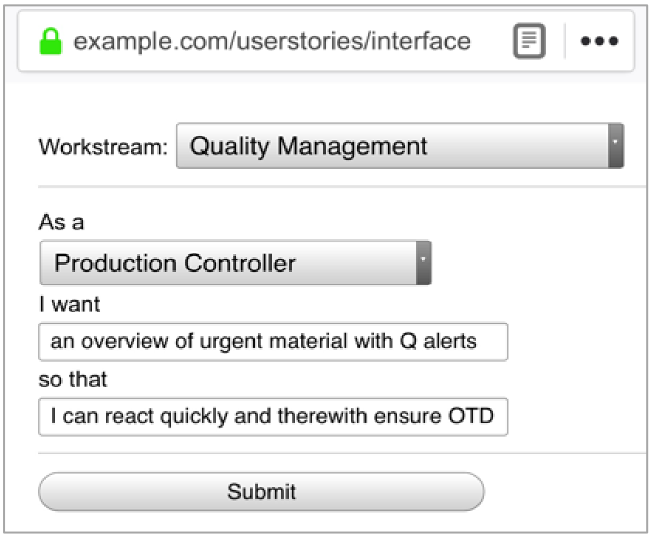


Figure 1 - Web Interface comprising the necessary fields for the Collection of User Stories

The requirements elicitation was open for a duration of three months, during which more than 600 entries were collected. 550 entries remained valid after a first manual check by the project team for erroneous input (duplicates, incomplete submissions, and out of scope entries).

3.2 Analysis

From the complete and unstructured record of requirements extract meaningful and compact information needs to be extracted for further use. This can be the communication of requirements to suppliers, to internal stakeholders, or for future project safeguarding and assessments, for which a list of multiple hundred entries is unsuitable. A partially automated approach is helpful, in order to avoid a time-consuming manual processing of all user stories. However, this requires a certain form of structure in the array of records. In order to create the necessary structure for an automated analysis, all of the entries need to be labeled appropriately. Upon complete labeling of all user stories, the clustering is realized via the calculation of a distance matrix and the application of an unsupervised clustering algorithm. Different principles are possible, of which the most suitable are presented and compared in their output.

3.2.1 Labeling of Records

After completion of the requirements collection, the entries inhibited little structured aspects. An option to prevent this, is to demand some sort of labeling of the entry by the submitting person. However, this requires substantial prior education of the participants in order to ensure the necessary consistency, and partially contradicts the simplicity and straightforwardness of the user story approach.

The labeling approach in the application case was of hybrid form, which means that both general contributors, as well as a project team member provided the labels. As previously stated, users of the web interface had to select both functional area of the MES, as well as the user role from five options. However, the most important label for the subsequent analysis was the user story impact factors: For each record in the user story array, an impact factor was assigned. The impact factors consist of a binary vector, designating in each component the fulfilled requirement of a user story regarding one or more defined enterprise goals. The range of goals was derived from the overall business goals of the company at which the exercise was conducted. They were in descending order:

- Quality Defect (Q) ↓
- On-Time Delivery (OTD) ↑
- Cost ↓

Additionally, the project team identified overall transparency and availability of complete information (T&I) to be an internal goal. We assigned it a lower priority than the overall business goals, since transparency was not seen as a direct value driver for customers. All aspects were split into subcategories, in order to allow for a concise analysis (

Table 1):

Q	Worker Info & Instructions	Material Track & Trace	
OTD	Throughput Time	Schedule Quality	Reactivity
Cost	WIP	OEE	
T&I	Info Availability	KPIs & Reports	

Table 1 - Division of Overall Business Goals into Subfactors for User Story Labelling

Each recorded user story was subsequently examined, and a binary impact for each of the factors on the right-hand side of

Table 1 was manually assigned by the project task force, resulting in a binary vector of 9 elements according to the subcategories in Table 1 for each user story (see Table 2).

	WIP	Schedule Quality	KPIs	Worker Info	⋮
User Story 1	1	1	0	1	…
User Story 2	0	1	1	0	…

Table 2 – Binary Vector for each User Story Record via Manual Assignment of Impact on Overall Business Goals

The binary assignment was chosen for the sake of simplicity, given that the manual assignment for a large number of records can be cumbersome. Depending on the algorithms used in the further analysis, also non-binary or weighted factors can be used in order to obtain a more precise result.

3.2.2 Distance Calculation for Records

The aim of the labeling and analysis approach is to sort and divide the record of user stories, and to create a clear and concise overview of categories and congregations of all collected requirements.

The calculation of distances between user stories was based on how the user stories fulfill defined enterprise goals. The function used to calculate the distances is a Manhattan metric. The Manhattan metric defines a distance between two points as the sum of the absolute values of its individual coordinate differences:

$$D(a,b) = \sum_i |a_i - b_i| \qquad (1.0)$$

The reason to choose the Manhattan metric over a Jaccard metric as described by [7], is that it is extendable to also non-binary vectors, which makes it more versatile. It accepts discrete, binary and non-binary values as entries. Additionally, the Manhattan metric has proven efficient applications in other clustering algorithms, wherefore it is also readily available in different libraries and analytical applications. The calculation of the distance is invertable, and the distance of a record with itself equals zero. According to the Gaussian Sum Formula, a number of $\frac{n^2-n}{2}$ calculations need to be carried out, in order to construct the $n \times n$ distance matrix for a record of a length of n entries, comprising the distances of each n-th entry, with the $n-1$ other entries.

3.3 Categorization

For easier understanding and communication during the selection of suppliers for softwares and applications, as well as the development and validation of the future solution, the need for a subsequent categorization is given.

In order to obtain a meaningful categorization, the distance matrix constructed via Manhattan metric was fed to a clustering algorithm. The following section will compare two common approaches: a hierarchical bottom-up clustering resulting in a dendrogramm, and a variant of a k-Means algorithm, in which entries are assigned to a predefined number of clusters, and rearranged until convergence to an optimal distribution.

For the clustering, the k-Medoids as a variant of the k-Means was chosen, as it is not limited to a specific distance limit or format. The k-Medoids algorithm converges to an optimized distance distribution by continuously re-determining the cluster centers, until the sum of the distances of all clusters is minimized. Similar to a k-Means algorithm, the final number of clusters needs to be chosen beforehand [9]. In order to determine the appropriate number of clusters, the relation of attributes, which constitute the binary vector, needs to be determined: More precisely, interdependent and dependent attributes need to identified. One can start by choosing one attribute arbitrarily, and subsequently construct a tuple of all dependent attributes. As soon as one dependency tuple is complete, this exercise is repeated until all of the attributes have been assigned to one or multiple tuples. The relationship tuples:attributes is an $n{:}m$ relationship, meaning that a tuple can contain multiple attributes, and an attribute can be contained in different tuples. In the given case, $n = 3$ tuples were identified, with the justification explained below:

$$T_1 = (WIP, Schedule\ Quality, Information\ Availability) \quad (2.0)$$

$$T_2 = (Material\ Track\ \&\ Trace, Reactivity\ KPIs\ \&\ Reports, Information\ Availability) \quad (2.1)$$

$$T_3 = (OEE, Throughput\ Time, Material\ Track\ \&\ Trace, Worker\ Info) \quad (2.2)$$

For the tuple T_1, the interdependency is that Work in Progress (WIP) can be optimized by an enhanced Schedule Quality. Schedule Quality can be improved by consideration of actual times, their deviations and available resources of the shop floor (set up, machining and transit times, tool and suppository availabilities), which requires information availability. However, the decision whether or not a tuple should be separated can be indistinct: In general, one tuple should be split into separate tuples as soon as there is at least one unidirectional relationship between the elements. A unidirectional relationship means that two or more elements of the tuple depend directly on another element of the tuple. This would be the case if tuples T_2 and T_3 were combined, given that both have a dependency on the attribute Material Track & Traceability. In case of uncertainties of the assignment of tuples, the outcome of the clustering for $n \pm 1$ clusters should be verified additionally.

4 RESULTS AND DISCUSSION

The solution of the hierarchical approach tends to be more homogeneous for large numbers of clusters with high similarities. The k-Means algorithm searches for the most homogeneous overall distribution, as it includes all points within a cluster [9]. It requires to select a priori the number of clusters, which significantly impacts the result. Yet, the overall homogeneity is considered a priority for the case study, and the preliminary selection of cluster numbers can be achieved by the described analysis of variable dependence of the binary vectors. We identified three clusters in which the user stories were distributed over three clusters of 135, 156 and 277 user stories for the k-Medoids algorithm. The distribution across the different clusters is homogenous and well-balanced, the same holds true for the allocation to different organizational levels within a cluster (black beams in the top image of Figure).

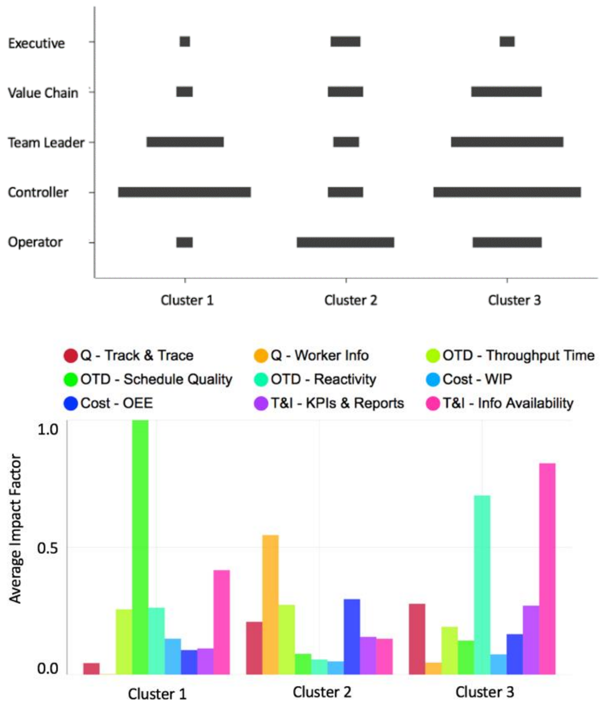


Figure 2 – k-Medoids: Distribution of User Stories across Clusters and Organizational Levels (top – width of the black beams proportional to number of User Stories), as well as their Impact on Overall Business Goals (bottom – coloured bars are average impact factor value per User Story in that cluster)

The same analysis was also carried out via a hierarchical clustering (Manhattan Metric, Figure), in which the two nearest neighbours are always linked. The hierarchical clustering yielded a distribution of user stories over clusters of 93, 114 and 361 requirements. The dendrogram for the hierarchical clustering demonstrates the selection of three clusters to be appropriate (Figure), as the upper node of the dendrogram consists of three larger branches on similar levels. However, it becomes clear that the hierarchical clustering introduces heterogeneities by the aggregation of larger sub clusters. Even for homogeneous individual sub clusters, the combination of them results in heterogeneities, as indicated by the distances of aggregated sub clusters on higher levels (Figure). Since the mean affection of overall business goals for the k-Medoids algorithm exhibits a more concise and separation and homogeneous distribution across clusters and organizational levels, this approach was confirmed to be the most suitable.

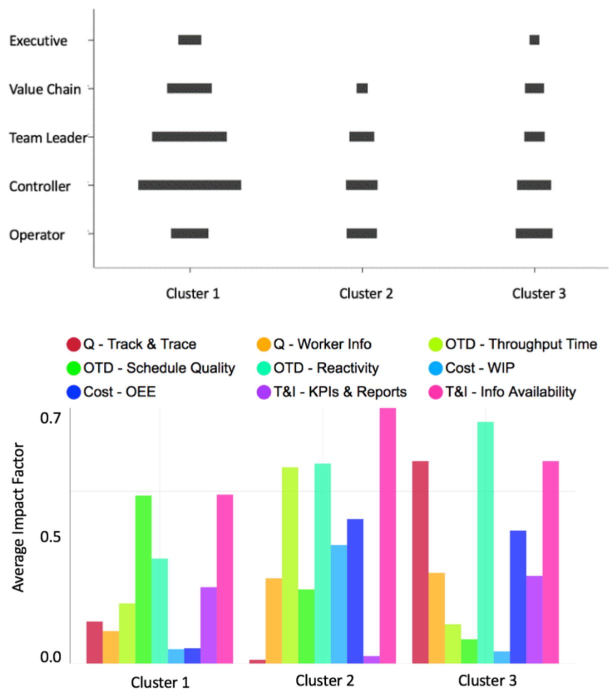


Figure 3 – Hierarchical Clustering: Distribution of User Stories across Clusters and Organizational Levels (top – width of the black beams proportional to number of User Stories), as well as their Impact on Overall Business Goals (bottom – coloured bars are average impact factor value per User Story in that cluster)

The k-Medoids clustering shows superior performance, compared to the hierarchical clustering approach, when considering the homogenous distribution of User Stories across clusters and organizational levels, and the distinct separation of User Story fulfilments measured in contribution to defined enterprise goals. This clear separation regarding the enterprise goals, is a desirable outcome allowing to segregate the entire digital transformation into smaller, more concise aspects and work packages, expressed in clusters.

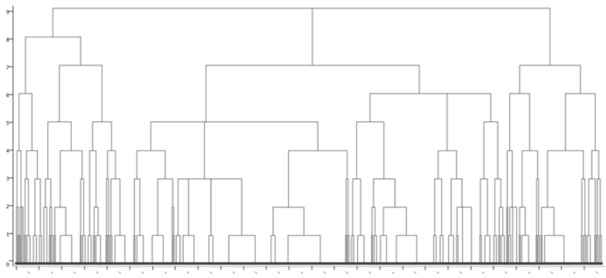


Figure 4 – Dendrogram of the Hierarchical Clustering

Each cluster is a functional collection contributing to the fulfilment of one or several business goals. The importance of each cluster can be determined, based on whether it fulfils high or low priority overall business goals. In the given case, the Q and OTD related clusters are assigned a superior priority, as dictated by the case study company's priorities. Hence, the user stories in these clusters are highlighted as crucial to future suppliers. In this context, the collection of user stories provides a simple and convenient way of recording requirements, with a subsequent analysis based on metrics rather than opinionated influence. The clustering results provide a clear overview of functions and challenges to tackle first, and indicate which collection of functions should be prioritized in order to allow for a quick harvest of benefits in digital transformation processes. Additionally, the clusters demonstrate how specific requirements of different organizational levels are interconnected: They clearly show the coherence of requirements on different levels that contribute to the same enterprise goal. This is the foundation for the development and rollout roadmap, as it prevents developing isolated horizontal solutions without any vertical integration. The given approach is therefore considered a robust fast-track from requirements collection to synthesis roadmap. The application as a validation tool is yet to be demonstrated and proven.

5 CONCLUSIONS

This study describes an approach capable of conducting requirements engineering for digital transformation processes in manufacturing industries in a straightforward manner. The asynchronous and intuitive collection of user stories allows to involve a large number of contributors in the requirements collection for digital transformation efforts, and enables a data-driven analysis and separation of collected requirements into appropriate categories. The allocation of user stories to clusters representing business goals and related functionalities, can be used for the internal and external communication of requirements, as well as for supplier evaluation, design and development roadmap purposes. The approach has proven efficient and robust to bias in Lorenz et al.'s [7] and this study, which appears to make it advantageous over traditional approaches. Of the different algorithms for the processing of all elicited entries, a suitable framework was selected and tested in a real requirements elicitation process. Its applicability has been approved by the members of the project team and deemed superior over traditional approaches prone to biased influence. Especially the stringent and calculation-based approach make it less vulnerable against short-term and undifferentiated influences of few opinion leaders with an organization – most notably for digitalization projects that may take between months and years to complete. Further research can investigate the automated assignment of user stories to overall business goals to replace the manual labelling effort.

6 REFERENCES

[1] Balasubramaniam R, Lan C (2008) Agile Requirements Engineering Practices: An Empirical Study, *IEEE Software*, Jan/Feb:60–67.

[2] Bracht U, Masurat T (2005) The Digital Factory between vision and reality, *Computers in Industry*, 56/4:325–333.

[3] Deuse J, Weisner K, Hengstebeck A, Busch F (2015) Gestaltung von Produktionssystemen im Kontext von Industrie 4.0, in *Zukunft der Arbeit in Industrie 4.0*, Berlin: Springer, pp. 99–109.

[4] Kagermann H (2014) Chancen von Industrie 4.0 nutzen, in *Industrie 4.0 in Produktion, Automatisierung und Logistik: Anwendung ·Technologien ·Migration*, Wiesbaden: Springer, pp. 603–614.

[5] Kletti J (2005) *Manufacturing Execution System - MES*. Springer Science & Business Media.

[6] Koch V, Geissbauer R, Kuge S, Schrauf S (2014) Chancen und Herausforderungen der vierten industriellen Revolution, *Pwc*, p. 52.

[7] Lorenz R, Lorentzen K, Stricker N, Lanza G (2018) Applying User Stories for a customer-driven Industry 4.0 Transformation, *in 16th IFAC Symposium on Information Control Problems in Manufacturing INCOM*, pp. 48–53.

[8] Paetsch F, Eberlein A, Maurer F (2003) Requirements Engineering and Agile Software Development, *in Proceedings of the 12th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, pp. 308–313.

[9] Park HS, Jun CH (2009) A simple and fast algorithm for K-medoids clustering, *Expert Systems with Applications*, 36/2 PART 2:3336–3341, DOI:10.1016/j.eswa.2008.01.039.

[10] Radziwon A, Bilberg A, Bogers M, Madsen ES (2014) The smart factory: Exploring adaptive and flexible manufacturing solutions, *Procedia Engineering*, 69:1184–1190.

[11] Schön E-M, Thomaschewski J, Escalona MJ (2017) Agile Requirements Engineering: A systematic literature review, *Computer Standards & Interfaces*, 49:79–91.

[12] Shariatzadeh N, Lundholm T, Lindberg L, Sivard G (2016) Integration of Digital Factory with Smart Factory Based on Internet of Things, *Procedia CIRP*, 50:512–517.

[13] Verein Deutscher Ingenieure (2007) VDI Guideline Manufacturing Execution Systems (MES) 5600. .

[14] Wegener K, Kunz A, Bochmann L, Bänziger T (2016) Industrie 4.0 für den Maschinen- und Anlagenbau, pp. 121–145.

[15] Wirdemann R (2017) *Scrum mit User Stories*. Carl Hanser Verlag.

[16] Zehnter C, Burger A, Ovtcharova J (2012) Key-Performance-Analyse von Methoden des Anforderungsmanagements, Karlsruher Institut für Technologie (KIT).

7 BIOGRAPHY

Thomas Gittler works on manufacturing-related Industry 4.0 topics during his current doctoral studies at the Institute for Machine Tools and Manufacturing (IWF) of the Swiss Federal Institute of Technology Zurich (ETHZ). He also accompanies the digital transformation of Swiss machinery manufacturers.

Rafael Lorenz is a doctoral student at the chair of Production and Operations Management at the Swiss Federal Institute of Technology. His research areas are the sources of process innovations and their implementation throughout global networks.

Lukas Weiss is a senior researcher with an industrial background leading a group focused on conceptual issues, innovative materials, energy efficiency and digitalization of machine tools at a research institute related to ETH Zürich.

Konrad Wegener is Head of the Institute for Machine Tools and Manufacturing (IWF) at the Swiss Federal Institute of Technology Zurich (ETHZ). His research is focused on machine tools, manufacturing processes and process chains, as well as the development, evaluation and optimization of production machinery.

Simulation-Based Product Development Framework for Cutting Tool Geometry Design

A. Devotta[1], T. Beno[2], M. Eynian[2]
[1]Sandvik Coromant AB, Sandviken, Sweden
[2]University West, Trollhattan, Sweden

Abstract
Cutting tool geometry design has traditionally relied on experimental studies; while engineering simulations, to the level of industrial deployment, have been developed only in the last couple of decades. With the development of simulation capability across length scales from micro to macro, cutting tool geometry development includes engineering data development for its efficient utilization. This calls for the design of a simulation-based approach in the design of cutting tool geometry so that the engineering data can be generated for different machining applications (e.g. digital twin). In this study, the needs for engineering model development of different stages of cutting tool design evaluation is assessed. To this end, some of the previously developed engineering models have been evaluated for evaluation of chip form morphology in industrially relevant nose turning process, work piece material behavior modeling and damage modeling for the prediction of chip shape morphology. The study shows the possibility for the developed models to act as building blocks of a digital twin. It also shows the need for engineering model development for different aspects of cutting tool design, its advantages, limitations, and prospects.

Keywords
Product design, Simulation, Finite element method

1.1 Introduction

Modeling and simulation are one of the key components in engineering advancement as it is directly linked to the codification of decades of knowledge for the betterment of future generations. Modeling and simulation are applied both in the design of products and manufacturing processes. In machining, from Taylor's tool life model to modern numerical modeling of the cutting process, engineering models which predict important process outputs have had a significant impact in the development of cutting processes. Numerical modeling of machining process is constantly being used in cutting tool design. With advances in the modeling of plastic deformation and fracture using nonlinear finite element method and increase in the understanding of material behavior during deformation at strains, strain rates, and the temperature is seen in metal cutting, the integration of numerical modeling in the cutting tool geometry development is inevitable.

Integration of finite element modeling within product development needs the development of tools and methods beyond the development of finite element (FE) model itself [1]. The report on research and development in simulation-based engineering [2] has mentioned that among other factors, the experimental validation of models remains a challenge in practical situations.

Numerical modeling of chip formation gives the opportunity to evaluate the machining process with improved temporal and spatial resolution compared to empirical and analytical modeling. A large amount of research work in machining processes have used force and temperature predictions to evaluate FE models [3]. In addition to force and temperature prediction, chip morphology prediction is of prime importance from a cutting tool design viewpoint. The term, 'chip morphology' refers to the description of chip flow, chip shape and chip curl [4]. Chip shape is primarily classified into the continuous chip, segmented chip, and discontinuous chip. Numerical models which can predict chip shape has been and still is an active area of research [5]. Chip shape characterization has been done traditionally using metallographic analysis and use of advanced characterization techniques like scanning electron microscopy (SEM) studies, electron backscatter diffraction (EBSD), X-ray diffraction, and transmission electron microscopy (TEM) are gaining prominence in the last decade. These methods are crucial for analyzing the metal cutting process for investigative purposes but are prohibitive in terms of time and cost which are crucial within a commercial product development process.

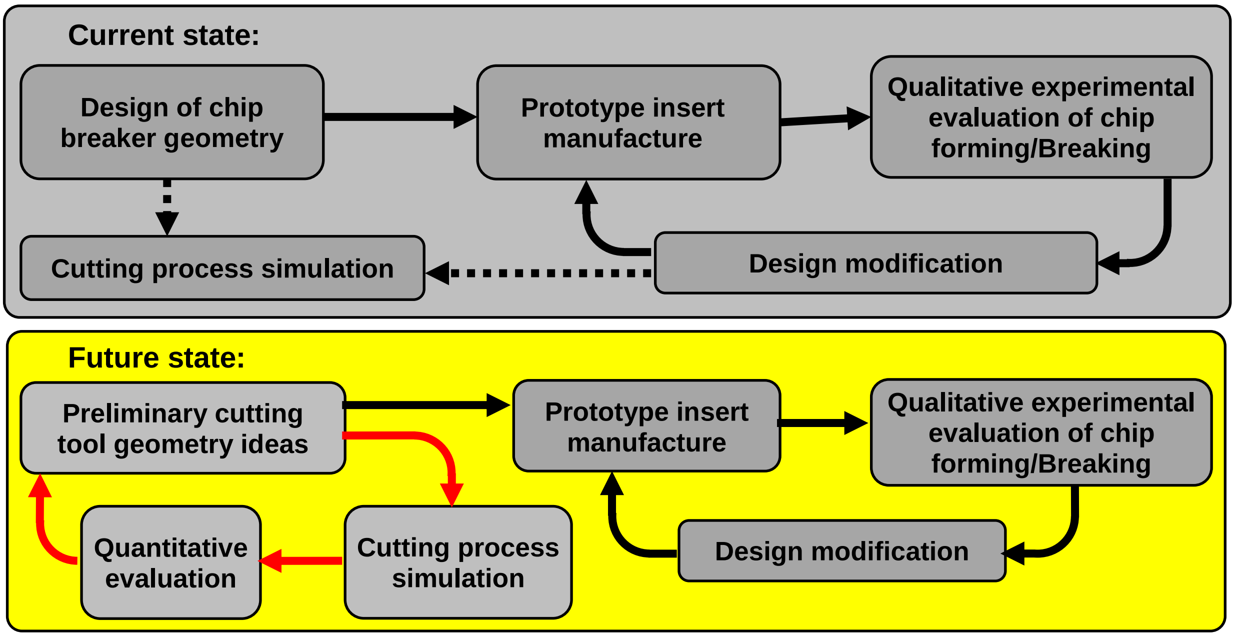


Figure 1 - Simulation based product development framework

From a cutting tool design perspective, chip flow on the rake face is important as it has a direct consequence on the force, tool wear, workpiece surface finish and process robustness [6]. Therefore, chip curl which describes the chip flow on the rake face is to be used as a parameter to evaluate an FE model. Chip curl as an evaluation method within academic research is limited compared to chip shape owing to its complex geometry. Nakayama et al [7] defined the different geometrical variations of chips with reference to the chip flow angle, feed rate and depth of cut. Kharkevich et al [8] developed a more fundamental model to model chip curl from a geometric perspective and developed models to evaluate chip curl during machining offline. Most of the studies have only used chip curl qualitatively. Quantitative evaluation has been limited mostly to orthogonal cutting. Quantitative evaluation of chip curl in nose turning process has been reported scarcely in comparison [9],[10].

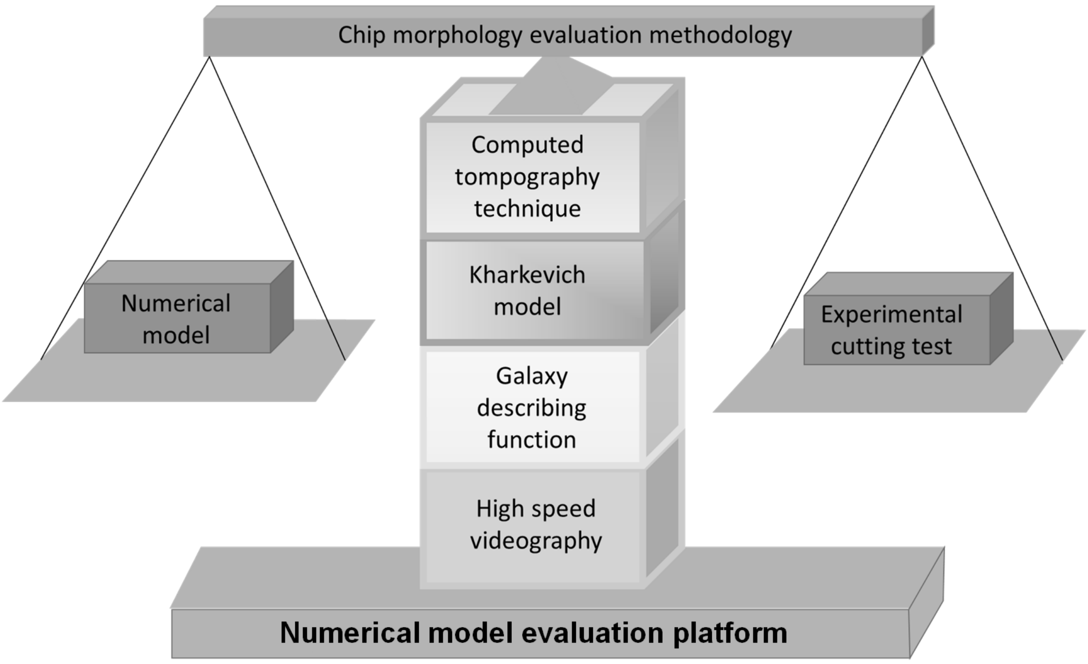
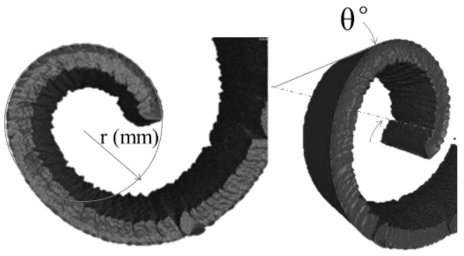


Figure 2 - Numerical model evaluation platform with specific reference to chip morphology [4].



r (mm) - Initial chip radius; θ^{o} – Twist angle

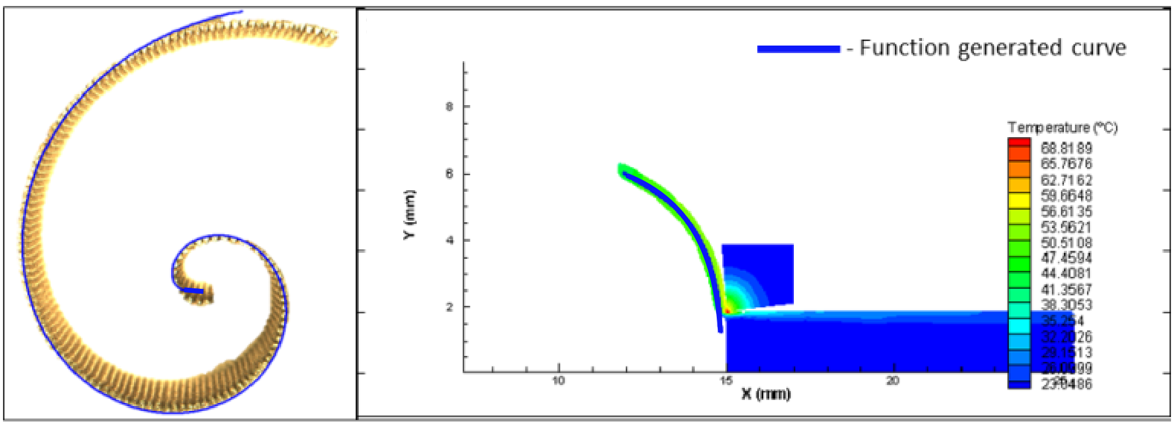
Figure 3 - Measurement of initial chip curl parameters using computed tomography [14].

Chip formation in machining is modeled using a finite element method which can simulate large plastic deformations and fracture, a material model as a function of strain, strain rate, temperature, microstructure, heat transfer, and friction. To incorporate such models within product development requires a considerable level of effort and should warrant significant economic benefits. The ability to integrate with other components of product development systems such as CAD / CAE is also important. Gardner et al [11] have shown that DEFORM, Advantedge, and Abaqus are three softwares available for chip formation models with varying levels of maturity in terms of ease of set up, material modeling capability, adaptive meshing capability and overall control. Within product development, ease of simulation set up capability, data transferability and overall process control are important to be used by tool design engineers.

Within this work, the development of a framework to employ chip formation models within product development of cutting tools is presented. Methods required for experimental evaluation of finite element models in terms of chip morphology is discussed. Chip formation simulation in machining of an engineering steel using the presented frame work is also carried out to evaluate its efficacy.

1.2 Simulation-based product development framework

Cutting tool product development involves the design of a cutting tool insert and tool holder. The design of cutting tool insert involves the design of different functional surfaces, rake face, relief face and edges connecting the rake face and the relief face and can be termed as the design of a chip breaker. A chip breaker design is governed by geometrical constraints and manufacturing constraints. New chip breaker development predominantly uses experimental studies for evaluation aided to a large extent by the knowledge gained by design engineers spanning several decades as shown in the current state of Fig. 1. Numerical modeling based cutting process simulation is utilized in the present stage to investigate specific problems and is not integrated entirely into the product development process. This is highlighted in the current state of Fig. 1 by the dotted lines. The challenges with full integration of simulation in the product development is three-fold. One is the absence of simulation evaluation methodologies. The second reason is the level of maturity in data transferability between different CAD/CAE systems. The third reason is the very high demand for simulation technologies in terms of its prediction capabilities. The future state shown in Fig. 1 presents the possibility to integrate cutting process simulation into the product development process. This requires improvement in all the three aspects mentioned. The future state which envisages a simulation-based product development process aids incorporation of knowledge more effectively into the next generation of products and processes. In the following section, some of the works related to bridging the need in two of the three above mentioned reasons, cutting process simulation, and quantitative evaluation are analyzed. Data transferability between CAD/CAE systems is reserved for future work.



Experimental : φ_s = 2.839 **Simulated : φ_s = 2.975**

Figure 4 - Chip curl modeling using a galaxy describing function [15].

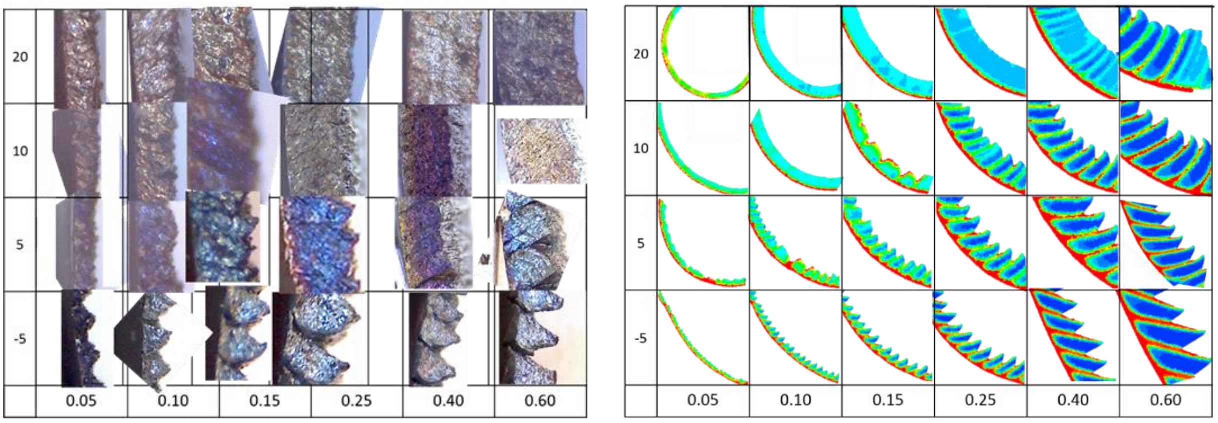


Figure 5 - Chip segmentation prediction validated using experimental investigation in machining of AISI 1045 steel [19].

2 SIMULATION EVALUATION METHODOLOGIES:

The evaluation methodologies required for implementation of numerical models should use measurable criteria [12] whereas, in the traditional product development process, success is measured at a higher level such as product profitability. In the traditional cutting tool product development process, chip morphology is used as a measurable criterion, albeit a qualitative one. The chip morphology is measured through studying the variation with increasing feed rate and depth of cut for a chosen cutting speed. This is presented in the form of a chip chart in technical guides [13]. In this work, the possibility to employ quantitative methodologies to evaluate chip morphology in evaluating simulation results against experimental validation is assessed. This would provide the ability to advance simulation technologies with specific reference to be used in product development. Fig. 2 shows the need to develop such a platform which utilizes several measurement methodologies. This platform would enable the evaluation of the simulation predictions with experimental results for chip morphology studies. This platform is to be continuously updated with developments in the field of metrology. The following section provides some of the measurement methodologies developed in previous works that can be used as a part of the numerical model evaluation platform.

2.1 Computed tomography for digitalization of chip morphology

Chip morphology, due to its complex geometry does not lend itself for characterization through conventional methods. In a previous work [14], a measurement methodology was developed to measure chips obtained from machining tests. This provides the possibility to compare chips between simulation and experiments. The detailed methodology is described in [14]. This method lends itself to be used in a variety of machining process ranging from orthogonal cutting to gear cutting. The short set up time, reduced measurement uncertainty and a direct 3D CAD model output provides the ability to be used in product development. Fig. 3 shows the CAD models obtained through computed tomography of physical chips. The CAD models enable chip visualization in relation to the cutting tool geometry, digitalization for future use and also CAE analysis. Fig. 3 shows experimental chip can be categorized by geometrical parameters such as initial chip radius and twist angle which provides better insights for the design engineers.

One of the current disadvantage with this method is the huge file size of these CAD models. Developments in transferrable CAD modeling would provide the possibility to use this method more often.

2.2 Chip modeling using geometrical curves

With the earlier method, being available for any cutting conditions, the method by its inherent nature is an offline method. At the early stages of product development for special materials, a methodology for chip morphology characterization for a geometrically simple cutting process such as orthogonal cutting is warranted. A methodology employing an analytical equation describing the chip up curl was developed in [15] and shown in Fig. 4. The function can predict the varying chip curl. The simplicity of the model lends itself to be incorporated into an online measuring system. Fig. 4 shows both experimental chip and simulation predicted chip measured using the analytical equation. The analytical equation was developed by Ringermacher et al [16] for categorization of spiral galaxies. This helps in quantifying the chip morphology easier than the traditional metallographic analysis. Fig. 4 shows that the chip's initial radius

2.3 Chip morphology modeling in nose turning process

With computed tomography method being used to capture the chip morphology offline, a methodology employing a mathematical description of the chip curl in 3D and high-speed videography is used to measure the chip morphology online during the experimental investigation and is presented in [17] and also presented here in Fig. 5. The methodology provides the possibility to evaluate simulation results in practical metal cutting processes. The method provides a very accurate description of the chip flow, chip up curl and chip side curl thereby aiding the design engineer to evaluate design iterations in a quantitative method. The method employs the mathematical formulas developed by Kharkevich et al [8] to calculate the chip geometry parameters in the 'chip in tool' reference frame from the chip geometry parameters in 'chip in hand' reference frame. The 'chip in tool' reference frame chip geometry parameters is chip flow angle, chip up curl radius and chip side curl radius. The 'chip in hand' reference chip geometry parameters is diameter, pitch, and slant angle as described in [17]. To evaluate the methodology, chip flow angle is measured by high-speed videography and the chip up curl and chip side curl radius are measured by computed tomography technique developed in the previous study [14].

3 NUMERICAL MODELING OF MACHINING PROCESS

Numerical modeling of machining processes has attained a level of maturity where simulations can be run without much intervention for mesh degradation. This provides the possibility to use chip formation simulations to predict chip curl for different chip breaker design alternatives. This requires robust material models that are built for the material used during the experimental investigation.

Thirdwave Advantedge [18] provides a database of workpiece material and tool properties and a simple simulation setting up procedure aiding in its deployment in a product development process. The software also provides the possibility to deploy custom material models. The different aspects of numerical modeling of machining process are described with good clarity in [19]

The accuracy of the cutting process simulation is in its accuracy of the description of thermo-mechanical properties of workpiece material and tool material. Upgrading the simulation capabilities is a continuous process and the simulation tool should lend itself to regular upgrading. A commercial software with a reliable support provides this possibility.

One of the important insights that can be obtained from machining simulations that aids in product development is the prediction of transition from continuous chip to segmented chips for different chip breaker geometries. With recent advances in the understanding of the influence of the loading conditions (Stress triaxiality and lode angle parameter) on material failure (chip segmentation), the ability to predict chip segmentation is improved. The work presented in [20] shows that the influence of rake angle on chip segmentation can be captured by modeling the damage failure as a function of stress triaxiality and temperature and not only as a function of temperature. The lode angle parameters influence has been studied by other authors for limited cutting conditions. Future work in this area is being planned to test this hypothesis for varying cutting conditions.

4 CONCLUSION

In this work, the need for simulation-based product development framework to use in cutting tool development was presented. Challenges with the present method and a future desired state are presented highlighting areas that need to be improved. Simulation evaluation methodologies, data transfer between CAD/CAE and high demand for simulation accuracy are presented as some of the most important challenges. Some of the work developed towards the development of simulation evaluation methods are presented and developmental activities regarding simulation accuracy are also presented. The presented method provides the possibility to deploy a first-generation simulation-based product development and identifies the areas to be continuously improved towards a more robust simulation-based product development process within cutting tool design.

5 ACKNOWLEDGMENTS

The authors kindly acknowledge the financial support from Sandvik Coromant and the Knowledge Foundation through the Industrial Research School SiCoMaP, Dnr 20110263, 20140130

6 REFERENCES

[1] J. Kortelainen, "Towards a simulation-based product process," 2015.

[2] S. C. Glotzer, *International assessment of research and development in simulation based engineering and science*. Imperial College Press, 2011.

[3] J. Mackerle, "Finite element analysis and simulation of machining: an addendum," *Int. J. Mach. Tools Manuf.*, vol. 43, no. 1, pp. 103–114, Jan. 2003.

[4] A. M. Devotta, "Characterization & modeling of chip flow angle & morphology in 2D & 3D Turning process," University West, 2015.

[5] Y. C. Zhang, T. Mabrouki, D. Nelias, and Y. D. Gong, "Chip formation in orthogonal cutting considering interface limiting shear stress and damage evolution based on fracture energy approach," *Finite Elem. Anal. Des.*, vol. 47, no. 7, pp. 850–863, 2011.

[6] R. Ghosh, O. W. Dillon, and I. S. Jawahir, "An investigation of 3D curled chip in machining Part 2: Simulation and validation of FE techniques," *Mach. Sci. Technol.*, vol.

2, no. 1, pp. 117–135, Aug. 1998.

[7] K. Nakayama and M. Arai, “Comprehensive Chip Form Classification Based on the Cutting Mechanism,” *CIRP Ann.*, vol. 41, no. 1, pp. 71–74, Jan. 1992.

[8] A. Kharkevich and P. K. Venuvinod, “Basic geometric analysis of 3-D chip forms in metal cutting.,” *Int. J. Mach. Tools Manuf.*, vol. 39, no. 6, pp. 965–983, 1999.

[9] R. Li, A. J. Shih, and O. Article, “Finite element modeling of 3D turning of titanium,” *Int. J. Adv. Manuf. Technol.*, vol. 29, no. 3–4, pp. 253–261, Jun. 2006.

[10] S. Buchkremer, F. Klocke, and D. Lung, “Finite-element-analysis of the relationship between chip geometry and stress triaxiality distribution in the chip breakage location of metal cutting operations,” *Simul. Model. Pract. Theory*, vol. 55, pp. 88–103, Jun. 2014.

[11] D. Joel, A. David, J. D. Gardner, A. Vijayaraghavan, and D. A. Dornfeld, “Comparative Study of Finite Element Simulation Software,” 2005.

[12] L. T. M. Blessing and A. Chakrabarti, *DRM, a Design Research Methodology*. .

[13] Sandvik Coromant, *Modern metal cutting: a practical handbook*. 1994.

[14] A. Devotta, T. Beno, R. Löf, and E. Espes, “Quantitative characterization of chip morphology using computed tomography in orthogonal turning process,” *Procedia CIRP*, vol. 00, pp. 299–304, 2014.

[15] A. M. Devotta, T. Beno, and R. Löf, “Modeling of Chip curl in Orthogonal Turning using Spiral Galaxy describing Function,” in :, 2016.

[16] H. I. Ringermacher and L. R. Mead, “A new formula describing the scaffold structure of spiral galaxies,” *Mon. Not. R. Astron. Soc.*, vol. 397, no. 1, pp. 164–171, 2009.

[17] A. M. Devotta and T. Beno, “Characterization of Chip Morphology in Oblique Nose Turning employing High Speed Videography and Computed Tomography Technique,” *Int. Conf. Compet. Manuf. – COMA ‘16”in Stellenbosch, South Africa 2016*, pp. 249–254, 2016.

[18] “http://www.thirdwavesys.com/advantedge/.” .

[19] P. J. Arrazola, T. Özel, D. Umbrello, M. Davies, and I. S. Jawahir, “Recent advances in modelling of metal machining processes,” *CIRP Ann. - Manuf. Technol.*, vol. 62, no. 2, pp. 695–718, Jan. 2013.

[20] A. M. Devotta, T. Beno, R. Siriki, R. Löf, and M. Eynian, “Finite Element Modeling and Validation of Chip Segmentation in Machining of AISI 1045 Steel,” *Conf. 16th CIRP Conf. Model. Mach. Oper. CIRP C. 2017 ; Conf. Date 15 June 2017 Through 16 June 2017*, vol. 58, pp. 499–504, 2017.

7 BIOGRAPHY

Ashwin Devotta is an Industrial doctorate student at Sandvik Coromant, Sweden and University West. His research focuses on numerical modeling of cutting process with specific interest in material modeling and product development

.

Prof. Tomas Beno is a Professor of Industrial Production at University West, Trollhättan, Sweden. His research interests include machining and process planning for CAM programming.

Mahdi Eynian received his Ph.D. from the University of British Columbia, Canada in 2010 and joined University West in Trollhättan, Sweden where he is a docent now. His interests include simulation, dynamics of structures and machining dynamics.

Nano-Indentation deformation Studies of Machined $AlMg_1SiCu$ Alloys via Molecular Dynamic Simulation

O. A. Olufayo[1], V. Songmene[1], M.K.O. Ayomoh[2],

[1]Department of Mechanical Engineering, École de Technologie Supérieure, ÉTS, 1100 Notre-Dame St. West, Montréal (QC), Canada.

[2]Industrial and Systems Engineering Department; University of Pretoria, Pretoria, South Africa.

Abstract

This paper presents results on a simulated deformation study of crystalline $AlMg_1SiCu$ alloys using classical molecular dynamics (MD) simulation. The experimental case study focussed on high speed machining chips of Aluminium alloys 6061-T6. During high speed machining, the workpiece material undergoes large deformation at high strain rate and elevated temperature at both the primary and secondary shear zones. Under some specific machining conditions, this can lead to a fine grain microstructure change at the primary and/or secondary shear zones. A nano-indentation representation of this microstructure was replicated by coding in LAMMPS (Large Scale Atomic/Molecular Massively Parallel Simulator). The granular workpiece was designed and subjected to indentation studies. In this study, the influences of temperature and indenter diameter on deformation properties of alloys were investigated. The simulated deformation studies were carried out by nano-indentation at both room and elevated temperatures. A simulation box size of dimension (200 Å × 150 Å × 100 Å) and consisting of 133,465 atoms was used to design and develop the alloy. Furthermore, a canonical (NVT) ensemble with periodic boundary conditions along the non-loading directions was used. The obtained results have been used to further discuss the underlying behavior of the alloys at the macroscopic level.

Keywords

Molecular dynamics, High-speed machining, Nano indentation, $AlMg_1SiCu$ alloys, Deformation

1 INTRODUCTION

Aluminium is a commonly used metal in industries due to its light weight and relatively low cost. Numerous forms of Aluminium alloys exist with improved physical and chemical properties. Traditional grades of aluminium alloys, such as the AA6061, AA6055 are produced through conventional foundry and smelting processes which creates coarse grain structures and large grain sizes. The manufacturing process of such alloys is largely responsible for its achievable microstructure. Distinct thermomechanical processes which consist of a combination of thermal and deformation processes have given rise to interactive microstructural features with improved material qualities such as improved yield strength, toughness, resistance to corrosion and stress.

While cutting high temperatures are achieved at the cutting tool tip. Plastic deformation occurs at the secondary shear zone and due to elevated temperatures creates a dynamic recrystallization of the grain microstructure. The rheological behavior during chip formation when machining has been shown to indicate the presence of dynamic recrystallization phenomena. Senecaut et. al [1] in their research indicated that Johnson Cook constitutive law does not include such phenomena in its rheological assessment in chip formation.

However, throughout the chip formation process, ultrafine grains microstructure is formed during machining at the intersection of the deformation zones close to the surface of the chip. This is formed due to the high shear strain generated as the tool moves over the surface of the workpiece. In combination with an elevated temperature level, dynamic recrystallization occurrence results in an ultra-fine grain microstructure. This takes place near the surface of the workpiece mostly along the deformation shear bands [2]. Table 1 shows the core elements of aluminium 6061.

Core elements (%)				Optional (%)				
Al	Si	Cu	Mg	Fe	Mn	Cr	Zn	Ti
96.0	0.8	0.4	1.2	0.7	0.15	0.35	0.25	0.15

Table 1 - Element composition of $AlMg_1SiCu$

Increased machining speed during cutting is associated with elevated temperatures in cutting operations. Machining of light metals such as aluminium is often performed using high speed machining due to improved process and to achieve desired surface results.

Mechanical/physical properties	AlMg1SiCu
Hardness, Brinell	95
Ultimate Tensile Strength	310 Mpa
Tensile Yield Strength	276 Mpa
Modulus of elasticity	68.9 gpa
Machinability	50 %
Aging Temperature	160 °c
Solution Temperature	529 °c
Density	2.7 g/cc

Table 2 - Physical properties of $AlMg_1SiCu$ [3]

High speed machining

High-speed machining is commonly used in the production of aluminium alloys parts due to its high manufacturing accuracy and efficiency, reduced cutting force and improved surface quality. Material shear slip deformation and tool chip friction are core machining characteristics that occur during high-speed machining [4]. In high-speed orthogonal cutting, the cutting edge is often perpendicular to the cutting direction hence, the shear angle is as presented in figure 1. During this process, high temperature concentration is often found at the primary shear zone and at the interface between the face of the cutting tool and the chip due to the high-abrasive forces and chip rubbing effect at tool-chip interface.

$$\phi = \arctan[a_c cos\gamma_0 / (a_o - a_c sin\gamma_0)] \quad (1)$$

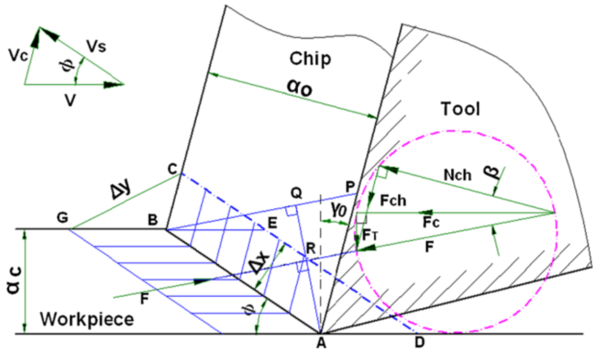


Figure 1 - High-speed orthogonal cutting model of chip formation [4]

where ∆x is the shear slip plane length (μm) indicating the distance of two adjacent slip lines, α_0 is the chip thickness (μm), Δy is the width of the first deformation zone (μm), ϕ is the shear angle (deg), the length of line AP (L_f) is tool-chip contacting length (μm), β is the friction angle (deg), the length of line AB (L_{AB}) is the length of shear plane (μm), FC is the cutting force (N) and FT is the thrust force (N) parameter.

Nano-deformation studies in alloys

There is a limited amount of tools available and capable of exploring the phenomenon of deformation in metallic alloys. Atomistic simulation is an adequate technique for investigation of materials behavior at the nanoscale. This methodology has consequently been employed in nano-scratching and nano-indentation studies.

Nano-scratching and nano-indentation phenomena occurs within a very small region in a workpiece and for a limited timeframe. A clear representation of the tool-workpiece interaction at such an atomic scale is not implicitly stated in research studies. The need for a supplemental insight into the underlying interactions at the atomic scale would further enhance the understanding of material deformation.

Based on the tedious and costly exercise that experimental investigations at such a scale would carry, atomistic simulations thus provides a viable tool whereby microstructural interactions at the nanoscale can be investigated. Molecular dynamics as an atomistic tool has been greatly used to study atomic motion in nano-indentation [5-8]. MD is a computer simulation tool which utilises time-based statistical mechanics to study atomistic interrelation for condition prediction and analysis [9]. Core predictions in MD are based on Newton's second law of classical mechanics and interatomic potential values to estimate atomic positions, velocity and acceleration. Newton's law is given by:

$$F_i = \mathrm{m}_i \mathrm{a}_i \quad (1)$$

Where F_i is the force acting on an atom, m_i and a_i stands for the mass and acceleration of the atom.

Influence of parametric selection for MD

The selection of interatomic potential is a critical step in MD design. In fact, the accuracy of MD results are contingent to the suitability of the model design for the molecular problem, in addition to the selection of its interatomic potentials [10].

There exist a large variety of potential energy functions commonly used in MD simulations which are particularly suited for different materials. For instance, Morse potential is preferably used for cubic metals, Tersoff potential for covalently bonded materials and for interface simulations, Lenard-Jones potential for rare gases and Embedded Atomic Potential (EAM) for a wide range of metals [11]. EAM potential incorporates an approximation for the many-atom interactions neglected by other potential schemes. Modified Embedded Atomic (MEAM) potential represents an extension of this potential which further includes the directionality of the bond angle to accommodate covalent systems. MEAM is ideal for metallic alloys with varying lattice structures and covalent materials such as silicon and carbon [12] .

The need to assess the value of the used interatomic potential could generate unreliable results. The development and testing of interatomic potentials is an intricate and elaborate process not often researched. However, the National Institute of

Standards and technology (NIST) has documented a list of potential files which have been tested, assessed for quality and published as adequate representations of force fields between elements [13]. In their study, Jelinek et. al [14] developed a MEAM for the pair interactions between Al, Si, Mg, Cu and Fe from the combination of their respective MEAM potential. In their study they compared the pair potentials with previously published MEAM parameters and validated experimental tests based on energies of defects, equilibrium volumes, elastic moduli and heat of formation. They concluded by indicating that the MEAM formalism permits any of the listed potentials to enable the prediction of multi-component alloy properties and adequate for the representation of alloy simulations. Figure 2 shows Comparison of MEAM potential to DFT based on B1 Volume.

A nano-indentation representation of this microstructure was replicated by coding in LAMMPS (Large Scale Atomic/Molecular Massively Parallel Simulator). The granular workpiece was designed and subjected to indentation studies and the influence of temperature, indenter diameter, loading rate, element composition on mechanical properties of alloys was investigated. Also, simulated deformation studies were carried out by nano-indentation at room temperature and at elevated temperatures.

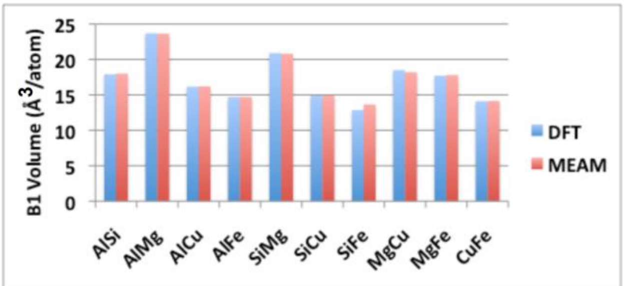


Figure 2 - Comparison of MEAM potential to DFT based on B1 Volume [15].

In this study, the simulated deformation studies of crystalline $AlMg_1SiCu$ alloy by classical molecular dynamics simulation (MD) is presented. MD potential force energies in the alloy are modelled using established MEAM potential energy functions to better represent the conditions occurring during nano-indentation tests. An experimental case study focused on fine grained microstructure area formed through deformation, during high speed machining of Aluminium alloys 6061-T6 is presented. The influence of element composition and mechanical properties of alloys was also investigated.

2 EXPERIMENTAL AND SIMULATION SETUP

2.1 Experimental case-study

The experimental nano-indentation setup focused on the milling of Aluminium 6061-T6. Due to the high speed in machining and the elevated temperature, large deformations at high strain rate both at the primary and secondary shear zones are expected on the workpiece. This condition is responsible for the fine grain microstructure localized at the primary and/or the secondary shear zones. Nano-indentation tests with the spherical indenter point with radius of 50nm as found in industrial nano-indentation testers was utilised on this structure. Figure 3 shows the load-displacement curve for a nano-indentation test.

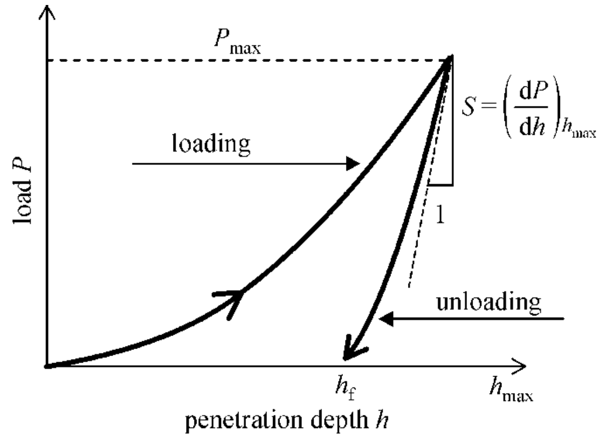


Figure 3 - chematic of load-displacement curve for an instrumented nano-indentation [16]

2.2 MD simulation setup

Various software platforms exist which can be used for the development of molecular simulations mainly based on their suitability and user-friendliness. LAMMPS (Large Scale Atomic/Molecular Massively Parallel Simulator) was used in this study to simulate the nano-indentation of ultra-fine grained microstructure of AlMg1Sicu alloy. This is an open source platform with a versatile range of toolbox options, integrative ease to third party platforms and a wide-range user support for various elements. Open visualisation tool (OVITO) was also utilised as a scientific visualization tool of the atomistic simulation (Figure 4).

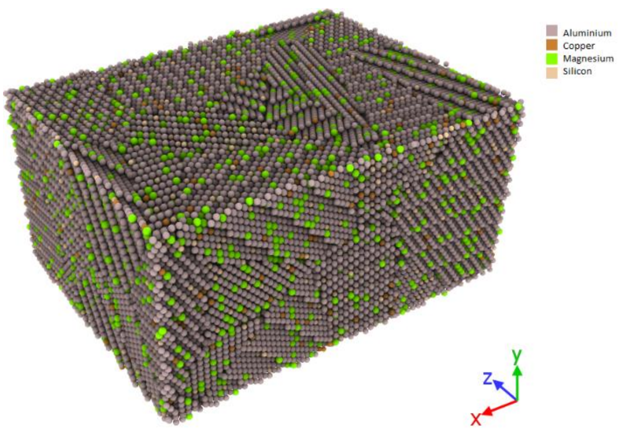


Figure 4 - MD workpiece design setup.

To mimic a portion of the alloy fine-grained tool chip occurrence, the workpiece was designed as an aluminium lattice with a percentage composition of its alloying elements to represent the composition of $AlMg_1SiCu$. Figure 4 shows the representation of the workpiece in the percentage composition of its elements (Table 1).

The process flow chart of the MD simulation is shown in Figure 5. In this figure, the procedure commences with the conception of the voronoi structure to represent fine grains of less than (10µm) as shown in figure 6.

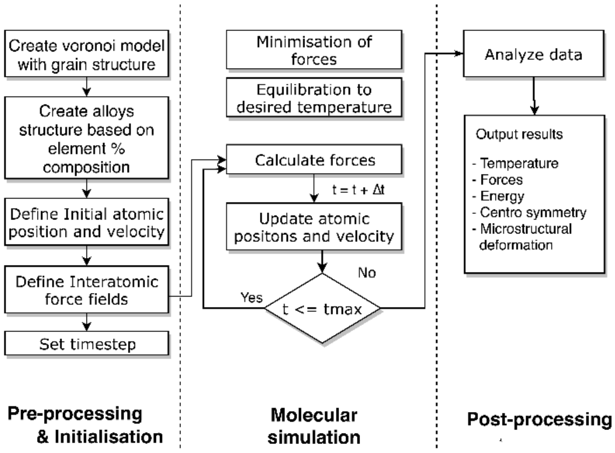


Figure 5 - MD process flowchart (Adapted from [17]).

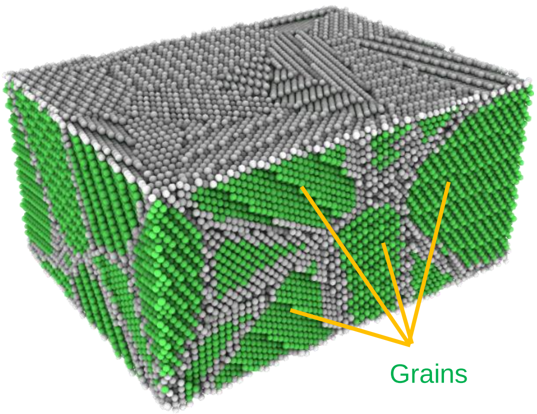


Figure 6 - Simulated workpiece with grain structure.

A valuation of grains number over size for the voronoi molecular setup is shown in Figure 7. In this setup a total of 13 grains were randomly generated with sizes around few nanometers.

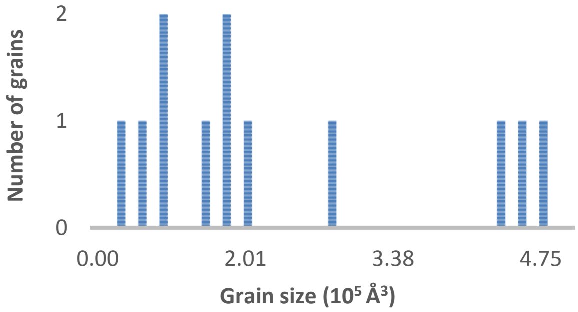


Figure 7 - Graph of number of grain over size.

2.2.1 Atomistic model design

A reproduction of the atomistic build of the medium strength AlMg1SiCu alloy was created. This aluminium alloy composition was selected due to a form of its fine grained microstructure which adequately mimics the literature [18]. The study only considers core alloying elements due to interatomic force potential limitations. Therefore property enhancing elements which are found in trace amounts such as titanium, Zinc, Iron, Chromium, and Manganese were not included. The atomistic model consist of a rectangular sample consisting of this mixture. The system consists of a total of 133,465 atoms. Table 3 and 4 show the percentage/atomic composition of element atoms as well as parameters setup for the simulation.

Workpiece material	No. of Atoms	%
Aluminium	108,107	98.1
Copper	4,004	0.3
Magnesium	14,681	1.1
Silicon	6,673	0.5

Table 3 - Percentage and atomic composition of elements

Workpiece material	$AlMg_1SiCu$
Workpiece lattice (a_1)	0.428nm
Workpiece dimension	$200a_1 \times 100a_1 \times 150a_1$
Tool structure	Rigid
Tool radius	30 a_1, 40 a_1, 50 a_1
Tool shape	Spherical
Number of grains	13
Depth of indentation	16nm
Force constant	10 kg^3/nm
Bulk temperature	293K
Time steps	1 fs

Table 4 - MD simulation for $AlMg_1SiCu$

2.2.2 Interatomic potential

As earlier indicated, this system is based on established interatomic potentials to adequately represent the alloys. The standard interatomic potential for modified embedded atomic method

(MEAM) potential for Al, Si, Mg, Cu and Fe alloys was used [14].

2.2.3 Energy minimization

In molecular dynamics, energy minimization prepares a molecular system by relieving any conflicting interactions in the initial configuration. A minimization function iteratively adjusts atomic coordinates in an aim to reduce energy gradient decent flow. Iterations are terminated when either the energy/force tolerances are attained or when the maximum iterations threshold conditions for the system energy is reached. Figure 8 shows the minimization chart of the total energy of the system. The stopping criteria triggered during simulation test is linked to the energy tolerance.

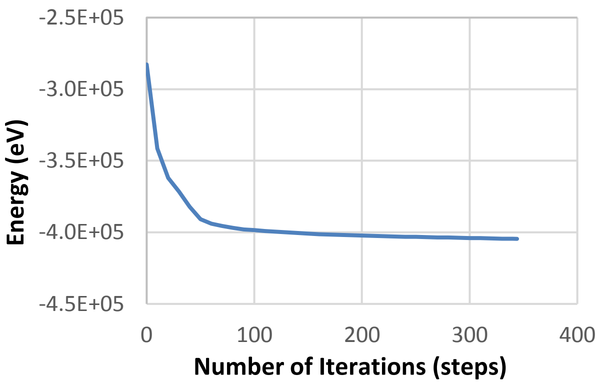


Figure 8 - Total energy minimization.

2.2.4 Equilibration of temperature.

The model was relaxed from its artificially assigned initial conditions to a more natural dynamic equilibrium state. This was performed by running the system setup at a predefined temperature for a fixed amount of time to permit achievement of atomic velocities equilibrium at the specified temperature. Equilibration permits the system to mimic the actual representation of metal materials. The Nose-Hoover non-Hamiltonian style time integration algorithm from the canonical (NVT) ensemble was used during equilibration of the system to generate the new positions and velocities. Figure 9 shows an equilibrated form of the simulated metallic alloy with its new structure.

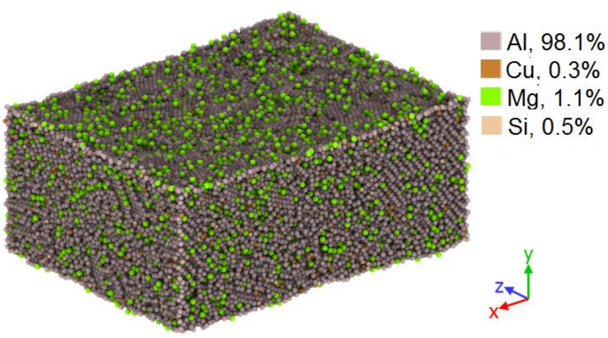


Figure 9 - Equilibrated workpiece with percentage element composition.

2.2.5 Boundary conditions

The global simulation box was set to be periodic for the x-axis, y-axis and z-axis hence allowing atoms to interact across the boundaries. However, a boundary condition was set at the base of the y-axis to prevent atomic flow in this direction.

2.2.6 Indenter specifications

The indenter originated from the LAMMPS toolkit and it acts as a rigid structure. This indenter is characterised with the property to repel all atoms within the group it touches. A spherical indenter was utilised to replicate the effects from an experimental case study.

A spherical indenter exerts a force of magnitude $F(r)$ on each atom it comes in contact with. The force is depicted as:

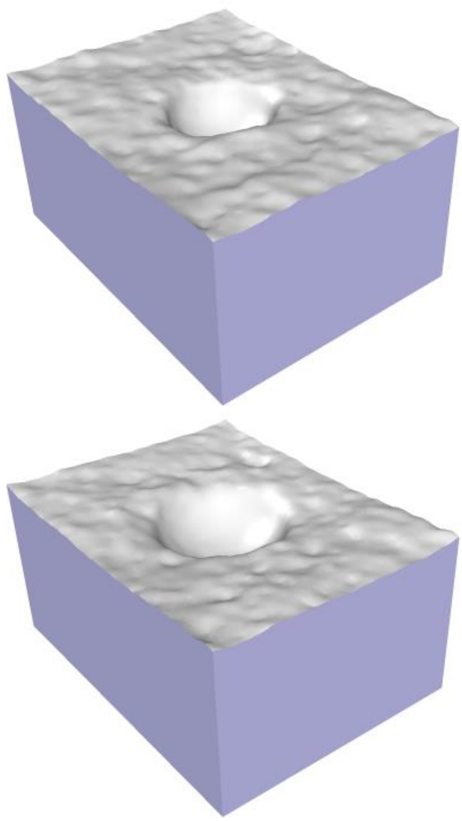
$$F(r) = -K\,(r - R)^2$$

Where K is the specified force constant, r is the distance from the atom to the center of the indenter, and R is the radius of the indenter. The force is repulsive and F(r) = 0 for r > R.

3 EXPERIMENTAL RESULTS AND DISCUSSION

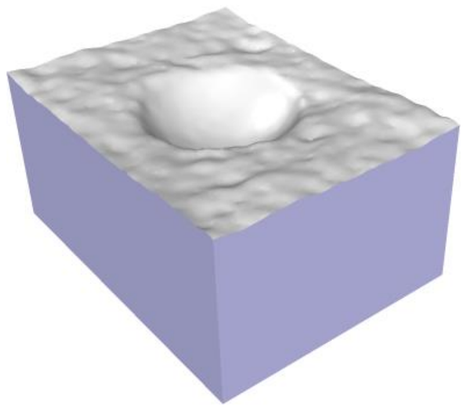
3.1 Effect of indenter size

Various indenter sizes have been used in the simulation. These sizes are 30, 40 and 50 Angstroms in diameter. Figure 10 shows the various indenters utilised.



(a) 30Å

(b) 40Å



(c) 50Å

Figure 10 - Different indenter sizes utilised.

From the smallest indenter size, a reduction in cutting forces is needed to indent the workpiece. This is evident based on the surface area and the resistance to material compression obtained. However, indenter sizes often affect the pop-out effect during indenter unloading under dynamic indentation. This presents a brisk ejection of the indenter out of the material [19]. This phenomenon is often associated with Si, and is believed to be as a result of the extension of the material volume by the phase transformation occurring in the workpiece. Figure 11 shows a pronounced push out seen at 50Å indentation.

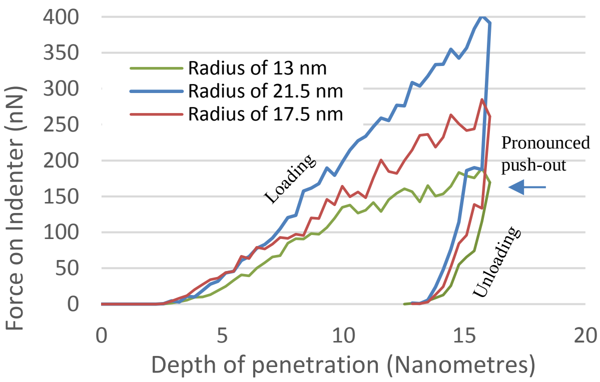


Figure 11 - Load-displacement curve for varying sizes of spherical indenter radius.

3.2 Influence of temperature on load-displacement curve

At elevated temperatures, plastic indentation depth is seen to increase. It is depicted with a reduced indentation force. This indicates that elastic recovery at such temperatures are reduced. Effects of temperature on aluminium alloys has been extensively studied in literature [20-22]. Young's modulus at elevated temperature on aluminium alloys is known to almost linearly decrease with increasing temperature [23]. This condition is associated with the effect of thermal softening on the material.

Figure 12 shows that a higher indentation force of 180nN is needed at a temperature of 300K than 140nN at 600K. Push-in/push-out fluctuations during loading were due to the brisk atomic dislocations during dynamic loading.

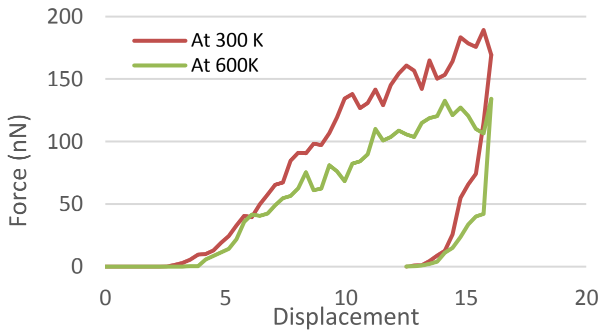


Figure 12 - Load-displacement curve at varying temperature.

3.3 Force and displacement vectors

The nano-indentation model was run for 60,000 timesteps, and the cutting forces, temperature and energy of the system were recorded. Simulation conditions were at a loading rate of 0.5 A/1000 timesteps for indentation depth of 7, 10, 13 and 16nm. Figure 13 shows the nano-indentation process as the spherical tool travels through the workpiece.

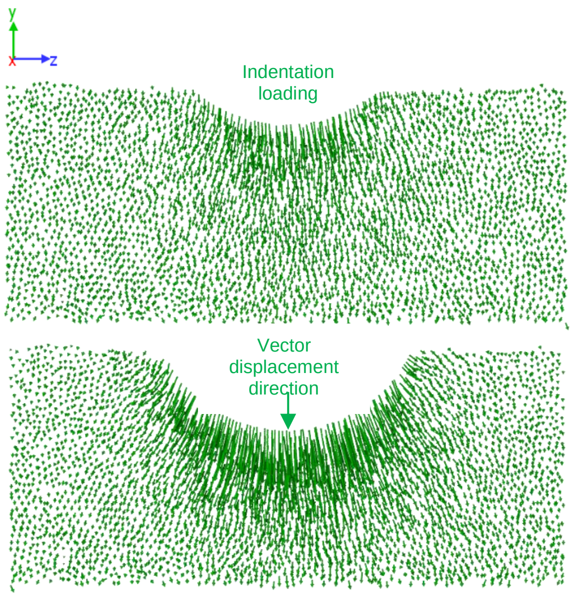


Figure 13 - Vector direction during indentation.

The workpiece atoms were compressed by the loading action of the tool and did not flow along the circumference of the tool. This pressure is characterised by a resistance of the material to tool entry with sharp increase in force readings over depth attained.

Displacement vectors shown in Figure 13 depict the direction atoms moved during compression of the material surface. High compressive forces on atoms was characterised with plastic deformation within the material at the tool entry zone. Despite the expected presence of high compressive force, no structural defects were found within the workpiece, a steady temperature was obtained (Figure 14).

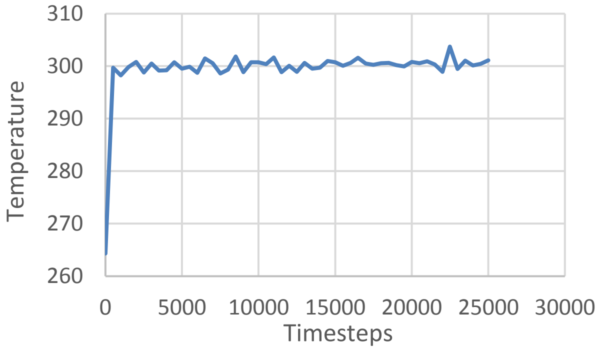


Figure 14 - Temperature reading during indentation.

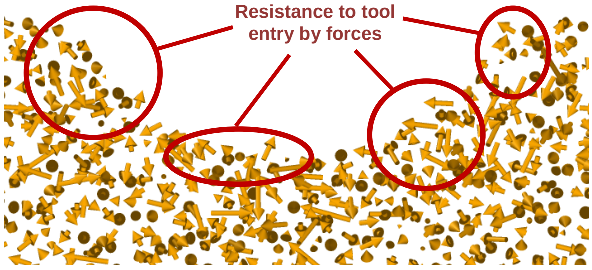


Figure 15 - Depiction of workpiece forces resistance to tool entry.

From Figure 15, a multi-directional force system is shown consisting of attractive forces between atoms and some of the surface forces acting in opposition to tool entry. Movement of atomic positions was caused by this influence of these external forces leading to material deformation. The expansion of interatomic radial separation between atoms led to an increase in negative potential energy and strong attractive forces present between atoms. This can be seen in Figure 16 and 17.

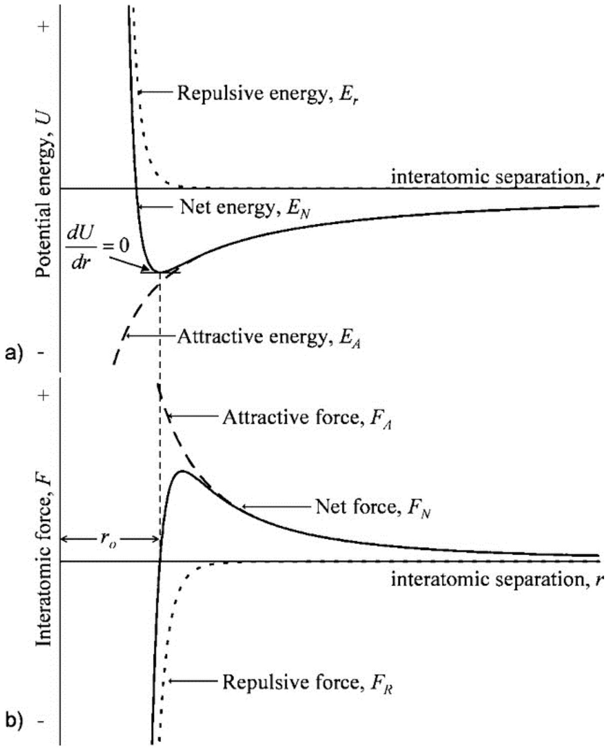


Figure 16 - Interatomic separation (a) vs Potential energy (b) vs interatomic force [24].

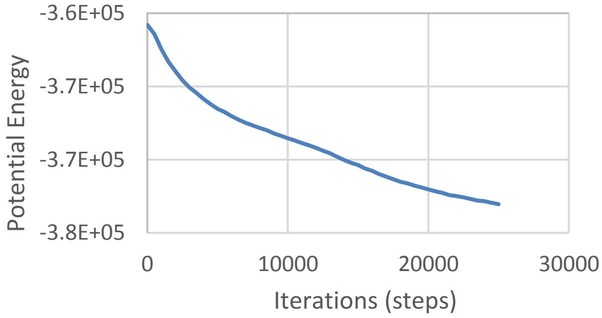
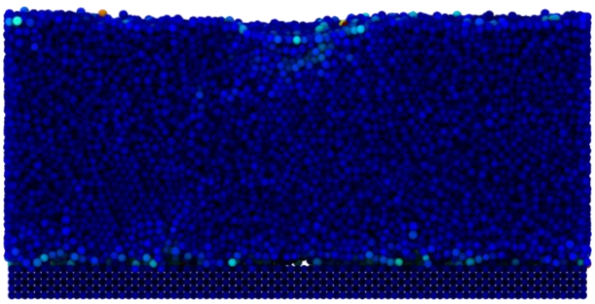


Figure 17 - Potential energy during indentation.

3.4 Shear strain analysis

A preview of the shear strain experienced by the workpiece atoms is as shown in Figure 19.



(a) 7nm of travel

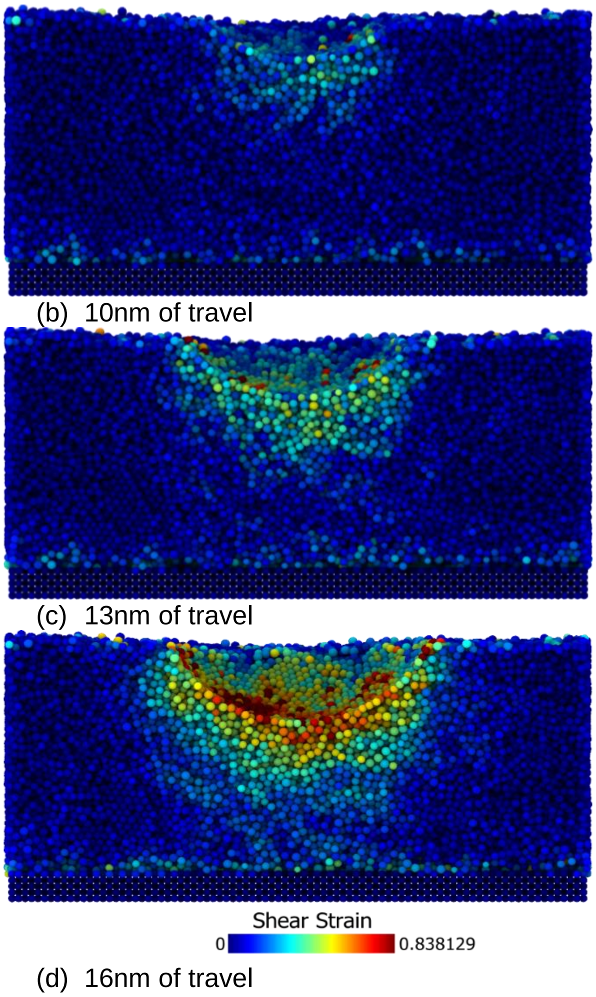


Figure 18 - Simulation of nano-indentation shearing progression from 7nm depth 16nm.

Shear strain progressively increased as the tool penetrated the workpiece. As the indentation depth deepened, the intensity of the shear strain compounded around the indented circumference. This strain effect was concentrated only at the point of entry and expanded within the zone of plastic deformation. The dislocation generation and movement within the structure of the material, resulted in the strengthening of the metal. This is also known as work-hardening. This effect was revealed by the continuous rise in indentation force as shown in Figure 19.

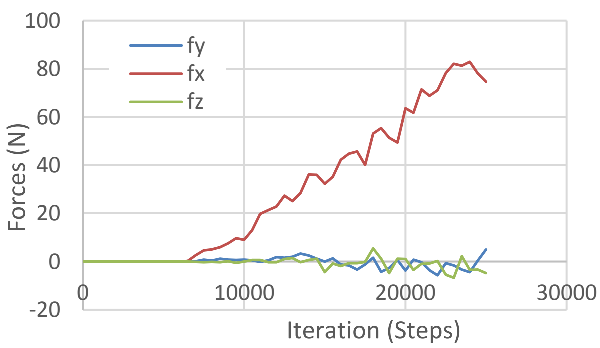


Figure 19 - Force increase on indenter plot

3.5 Workpiece surface elastic recovery

Elastic recovery of the workpiece was also studied during unloading. Elastic deformation experienced on the indented surface recovered with a low stress-strain index. As unloading occurred, push-outs from atoms brisk ejections were responsible for the reduction in surface form. This atoms were displaced outwards of the surface and responsible for reduction in surface roughness during nano-machining.

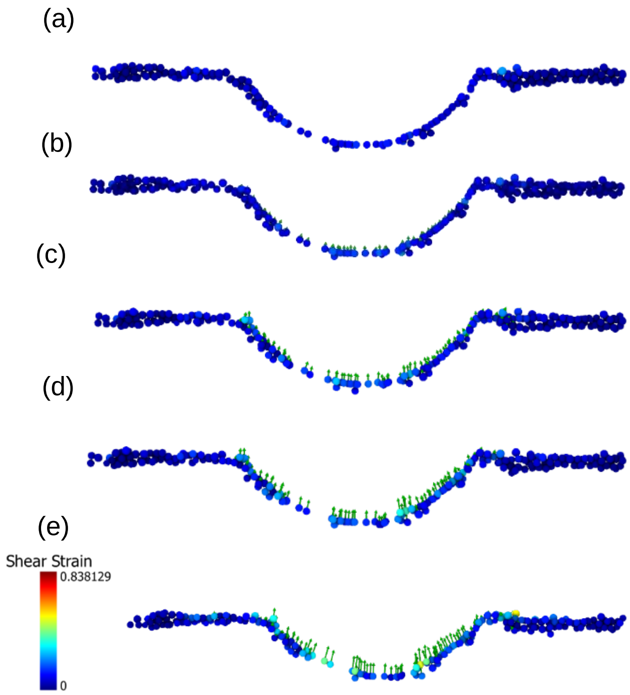


Figure 20 - Simulation of elastic recovery of surface of 3nm.

4 CONCLUSION

This study has presented the simulated deformation study of crystalline $AlMg_1SiCu$ alloys using classical molecular dynamics (MD) simulation. The experimental case study focussed on high speed machining chips of Aluminium alloys 6061-T6 characterised with a fine grain microstructure localized at the primary and/or secondary shear zones. The simulated granular workpiece was designed to mimic conditions and setup experienced in machining aluminum alloys.

From the results, it is seen that the indenter size increase influenced the indentation force needed to attain a particular indentation depth. This was further associated with an increase in intensity in push-outs of atoms during unloading. This condition was linked to the effect of higher compressive forces effect on interatomic spacing during atomic displacement. This increased repulsive forces are experienced within the material.

Deformation formed within the material was both elastic and inelastic. Plastic deformation was also influenced by the effect of temperature. Indentation tests carried out at increased temperature lowered the elastic modulus of the material with an increased plastic indentation depth. This was attributed to thermal softening of the material at such temperatures.

Shear strain also progressively increased as the tool penetrated the workpiece. The intensity of the shear strain compounded around the indented circumference and expanded within the zone of plastic deformation. The dislocation generation and movement within the structure of the material, resulted in work-hardening. This was evident with the increase on indentation force and attractive forces between atoms.

Elastic recovery during unloading also indicated reduced stress-strain on the material. This however explained the experienced reduction in surface integrity after material unloading from the interplay of atoms repositioning and ejections at the surface.

The depiction of the underlying behavior of $AlMg_1SiCu$ alloys explains nano-deformation within the workpiece. The practical case-study assessed the presence of ultrafine grains microstructure introduced during machining at the intersection of the deformation zones close to the surface of the chip and presents its relevance to current machining practices. Temperature results are in agreement with micro-scale studies which indicates a fall in indentation force with increased temperature. Future application will involve an assessment of the influence of loading rate and element composition of the alloy as well as experimental validation.

5 REFERENCES

[1]. Senecaut, Y., et al., *Dynamic Recrystallization Observed at the Tool/Chip Interface in Machining.* Key Engineering Materials, 2015. **651-653**: p. 1223-1228.

[2]. Liu, B., X.R. Zhou, and X.X. Zhang, *Orthogonal machining introduced microstructure modification in AA7150-T651 aluminium alloy.* Materials Characterization, 2017. **123**: p. 91-98.

[3]. *Physical properties, thermal properties*. Available in http://asm. matweb. com and www. aircraftmaterials. com/data/aluminium/6061/html.

[4]. Xu, D.C., et al., *Research on chip formation parameters of aluminum alloy 6061-T6 based on high-speed orthogonal cutting model.* International Journal of Advanced Manufacturing Technology, 2014. **72**(5-8): p. 955-962.

[5]. Hsieh, J.Y., et al., *Temperature dependence in nanoindentation of a metal substrate by a diamondlike tip.* Physical Review B, 2004. **70**(19).

[6]. Liu, C.L., T.H. Fang, and J.F. Lin, *Atomistic simulations of hard and soft films under nanoindentation.* Materials Science and Engineering a-Structural Materials Properties Microstructure and Processing, 2007. **452**: p. 135-141.

[7]. Kenny, S.D., et al., *Molecular dynamics simulations of nanoindentation and nanotribology.* Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences, 2005. **363**(1833): p. 1949-1959.

[8]. Huang, H. and J.W. Yan, *New insights into phase transformations in single crystal silicon by controlled cyclic nanoindentation.* Scripta Materialia, 2015. **102**: p. 35-38.

[9]. Olufayo, O.A. and K. Abou-El-Hossein, *Molecular dynamics modeling of nanoscale machining of silicon.* 14th Cirp Conference on Modeling of Machining Operations (Cirp Cmmo), 2013. **8**: p. 504-509.

[10]. Afify, N.D., G. Mountjoy, and R. Haworth, *Selecting reliable interatomic potentials for classical molecular dynamics simulations of glasses: The case of amorphous SiO2.* Computational Materials Science, 2017. **128**: p. 75-80.

[11]. Oluwajobi, A., *Molecular Dynamics Simulation of Nanoscale Machining*, in *Molecular Dynamics-Studies of Synthetic and Biological Macromolecules*. 2012, InTech.

[12]. Otieno, T. and K. Abou-El-Hossein, *Molecular dynamics analysis of nanomachining of rapidly solidified aluminium.* International Journal of Advanced Manufacturing Technology, 2018. **94**(1-4): p. 121-131.

[13]. Becker, C.A., et al., *Considerations for choosing and using force fields and interatomic potentials in materials science and engineering.* Current Opinion in Solid State & Materials Science, 2013. **17**(6): p. 277-283.

[14]. Jelinek, B., et al., *Modified embedded atom method potential for Al, Si, Mg, Cu, and Fe alloys.* Physical Review B, 2012. **85**(24).

[15]. Mike Baskes; B. Jelinek, S.G., A. Moitra, M. F. Horstemeyer, J., Houze, S. G. Kim, G. J. Wagner, *New MEAM potentials being developed for the Al, Si, Mg, Cu, and Fe alloy system*, in *2011 NIST Workshop on Atomistic Simulations for Industrial Needs* 2011: NIST Gaithersburg.

[16]. Chen WANG, Y.W., Jianzhuang XIAO, *Three-scale stochastic homogenization of elastic recycled aggregate concrete based on nano-indentation digital images.* Front. Struct. Civ. Eng.: p. 0-.

[17]. Buyukozturk, O., et al., *Structural solution using molecular dynamics: Fundamentals and a case study of epoxy-silica interface.* International Journal of Solids and Structures, 2011. **48**(14-15): p. 2131-2140.

[18]. Tham, Y.W., et al., *Microstructure and properties of Al-6061 alloy by Equal Channel Angular Extrusion for 16 passes.* Materials and Manufacturing Processes, 2007. **22**(7-8): p. 819-824.

[19]. Grabco, D., et al., *Elastoplastic response of TCO/Si coated systems to local loading.* Horizons in World Physics, 2012. **277**: p. 20.
[20]. Maljaars, J., F. Soetens, and L. Katgerman, *Constitutive model for aluminum alloys exposed to fire conditions.* Metallurgical and Materials Transactions A, 2008. **39**(4): p. 778-789.
[21]. El-Danaf, E.A., A.A. AlMajid, and M.S. Soliman, *Hot deformation of AA6082-T4 aluminum alloy.* Journal of materials science, 2008. **43**(18): p. 6324.
[22]. Zhang, H., et al., *Hot deformation behavior of the new Al–Mg–Si–Cu aluminum alloy during compression at elevated temperatures.* Materials Characterization, 2007. **58**(2): p. 168-173.
[23]. Summers, P.T., et al., *Overview of aluminum alloy mechanical properties during and after fires.* Fire Science Reviews, 2015. **4**(1): p. 3.
[24]. Niinomi, M., *Metallic biomaterials*. Vol. 11. 2008. 105-10.

6 BIOGRAPHY

Oluwole A. Olufayo obtained his PhD degree in Mechatronics Engineering from the Nelson Mandela University in Port Elizabeth. He worked for some years at Oil and gas installations as a field engineer and project consultant. He is currently a post-doctorate researcher at École de Technologie Supérieure (*ETS*) in Montréal, Canada.

Victor Songmene received a master (M. A. Sc) degree and a PhD degree in manufacturing from École Polytechnique de Montréal, Canada. He has been Senior Project Leader with the Industrial Research and Development Institute (IRDI) from 1995 to 2001. He is since 2001 a professor at the Department of Mechanical Engineering at École de Technologie Supérieure (*ETS*), Montréal, Canada. His research interests are in metal cutting, continuous product/processes improvement and environmentally conscious manufacturing.

Michael Ayomoh holds an MSc. and PhD degree from Industrial Engineering, University of Ibadan and Systems Engineering, University of Lagos respectively. He is currently a Senior Lecturer with the Department of Industrial and Systems Engineering at the University of Pretoria. His specific focus and area of research interest is in Robotics and applications of modelling in Systems Engineering

Simulation-Based Design of an Electrolyte System for Electrochemical Machining with Differentially Switched Currents

T. Petzold[1], M. Hackert-Oschätzchen[1], A. Martin[1], A. Schubert[1]

[1]Professorship Micromanufacturing Technology, Faculty of Mechanical Engineering, Chemnitz University of Technology, 09107 Chemnitz, Germany

Abstract

The manufacturing of components with complex internal features, e.g. for medical applications, aeronautics or automobile industry, is a significant challenge. Those components are often machined in temporarily and locally separated stages of production. Due to these separated stages, the form deviations and positioning errors increase, which lead to additional efforts for the quality assurance. The technology that shall be developed within the project SwitchECM is expected to allow the machining of different complex parts of the workpiece in one single production stage and shall simultaneously allow a high precision. For this purpose, a multi-cathode system will be developed, in which every single cathode can be switched with specific parameters, depending on the requirements of the pre-defined features. In this study a multi-cathode system will be shown which is designed to machine different features of one workpiece. The system consists of two separated cathodes. The features differ in size, position and aspired machining result. Two machining areas are located in a row. Thereby removal and reaction products of the first machining area may influence the second machining area. Based on fluid dynamics simulation the design of electrolyte flow will be analysed, which should be capable to flush the machining areas with required electrolyte.

Keywords

Electrochemical Machining, anodic dissolution, electrolyte flow simulation

7 INTRODUCTION

The electrochemical machining (ECM) process is based on the principle of anodic metal dissolution. The process depends mainly on the chemical properties of the workpiece material and the used electrolyte. The material is dissolved without mechanical contact between the cathode and the workpiece and removed via the electrolyte flow. [1, 2] Thus, there is neither mechanical nor thermal impact on the workpiece.

In the present study a multi-cathode system for ECM of different features in one workpiece applying differentially switched currents will be shown. The determination of the designs for the workpiece and the multi-cathode system were developed in cooperation of SITEC Industrietechnologie GmbH, BENSELER Sachsen GmbH & Co. KG, Leukhardt Schaltanlagen GmbH, IFU Diagnostic System GmbH, Porzellanmanufaktur Reichenbach GmbH and Chemnitz University of Technology. Figure 1 shows the schematic design of a simplified multi-cathode system. Currently, two specific single cathodes are used to machine the features for deburring (feature D) and shaping (feature S). The goal of the multi-cathode system is to combine these iterative steps into one single machining operation by using functional cathodes separated with an electric insulation (green) and controlling the differentially switching of the current of each cathode.

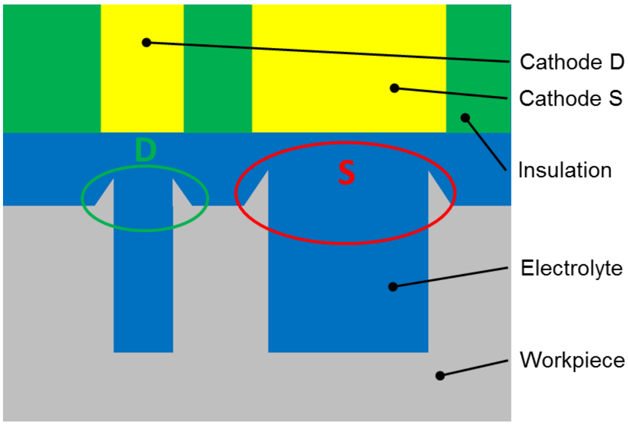


Figure 1 - Schematic design of a simplified multi-cathode system

In order to reduce costs and time-consuming iterative steps, simulations are suitable to assist the design of ECM processes [3].

Since in ECM, the electrolyte flow is a significant process parameter, it is important to design and set up a suitable flushing strategy, in order to continuously supply fresh electrolyte to the machining zone and thus realize constant machining conditions during the ECM process. Therefore fluid dynamics simulations were utilized.

Beside the challenge with the flushing supply, there are further more challenges to be expected. The cathodes will interact during the machining process.

This will influence the current flow and therefore the mass removal. One more challenge will be the shape control of the edge geometries over time.

8 WORKPIECE

The workpiece geometry, shown in figure 2, is manufactured with conventional machining processes. The therefore resulting burrs in the features should be removed with ECM.

Figure 2 - Design of the workpiece geometry

The workpiece consists of the stainless steel 1.4301 and is a cylinder with an internal diameter of 30 mm and a height of 80 mm. The internal features are visible as a single bore, feature D, with a diameter of 6 mm and two slot holes, feature S, with height of 4 mm and width of 8 mm. Both features have a burr with a maximum height of 0.4 mm and the burrs are located at the inside of the features due to the sequence of the machining steps. First the bores in the side of the cylinder are drilled and then the bore with the diameter of 30 mm is drilled.
The internal features represent two separated machining areas. The feature D is the first machining area where a deburring process is required. The two slot holes in feature S are located in a row and are the second machining area. Here a shaping process should be realized. This means after the machining the internal features differ in the edge geometry. Feature D should have an edge rounding of approximately 0.2 mm and the two slot holes should have sharp edges. Due to the different geometry of the two features different process parameters are necessary and therefore two separated cathodes for ECM with differentially switched currents are required.
To design a precise ECM-process it is essential to determine the material removal characteristics of the workpiece by help of an experimental analysis [5]. Such experiments were made to characterize particularly the material removal speed as function of the current density. In opposite to the removal speed, the feed speed is no considered parameter, because there is no feed speed during the processing time. It is considered to use pseudo-DC in the experiments. Pseudo-DC ECM with a voltage pulsation frequency of 200 Hz and pulsed DC ECM with a voltage pulsation frequency of 50 Hz in combination with a pulse duration of 4 ms were selected [6]. The obtained functions are shown in figure 3.

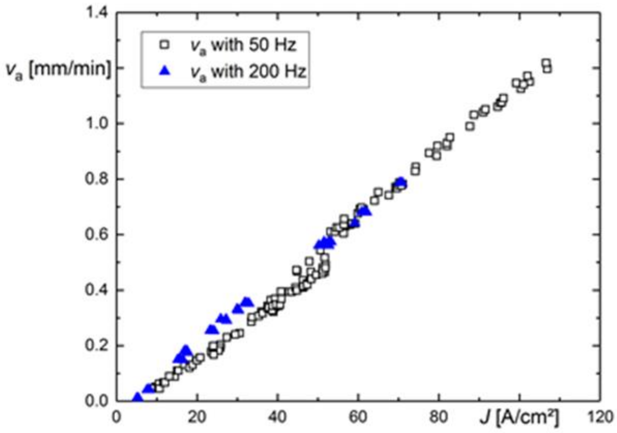


Figure 3 - Material removal speed as function of the current density for pseudo-DC ECM at voltage pulsation frequency 200 Hz and pulsed DC ECM at voltage pulsation frequency 50 Hz [6]

In figure 3 it is visible, that the material removal speed with a voltage pulsation frequency of 200 Hz differs from the material removal speed with a voltage pulsation frequency of 50 Hz at current densities below 55 A/cm². In this range the material removal speed with 50 Hz is up to 14.3 % lower than the removal speed for the experiments with 200 Hz.

9 MULTI-CATHODE SYSTEM DEVICE

The developed device for the selected multi-cathode system is shown in figure 4.

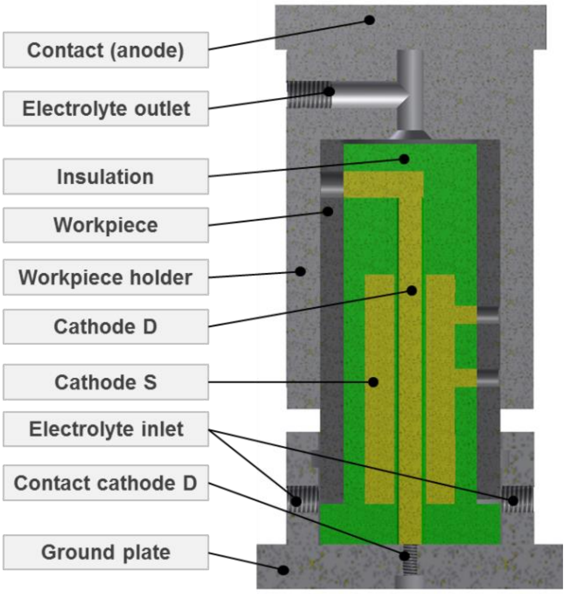


Figure 4 - Developed device for the selected multi-cathode system

The device consists of anode related parts and cathode related parts. The anode related parts are the contact, the workpiece holder, the workpiece and the electrolyte outlet system. The cathode related parts consist of the ground plate, the cathode S, the cathode D, a ceramic composite insulation and the electrolyte inlet system. The contact, the workpiece holder and the ground plate are made of stainless steel 1.4301. The workpiece is contacted through the workpiece holder to the contact. Each cathode has a separate electrical contact through the ground plate and is isolated to the device.

The multi-cathode system is constructed and manufactured by the project partners. This system enables, that the cathode for feature D can be mounted inside the cathode for feature S. The cathodes are coated with a ceramic composite insulation. The outer diameter of the multi-cathode system is 29.6 mm suitable for the shown workpiece geometry with a resulting working gap of 0.2 mm.

The device enables cross flushing. Therefore the electrolyte streams in at the bottom and out at the top of the device.

The multi-cathode system will be centered according to the bore of the workpiece with specific assisting features, which actually are being designed in agreement with the project partners. Concentric alignment limitations of +/- 0.01 mm will be considered in order to avoid collision and to assure pre-defined working distance.

10 SIMULATION-BASED ANALYSIS OF THE ELECTROLYTE FLOW

The fluid dynamics simulation considers the electrolyte flushing as single-phase flow. The simulation was calculated to obtain informations about the pressure distribution and the characteristics of the flow field in the machining areas.

The model considers the electrolyte area of the device shown in figure 4. Therefore a 3D-geometry suitable for simulation was designed in Autodesk Inventor® and transfered to COMSOL Multiphysics®. This electrolyte area includes the complete region of electrolyte from the inlet to the outlet of the device. Because of the high computing effort for fluid dynamics simulation the 3D-model was simplified into a 3D-section model. The resulting fluid domain is shown in figure 5.

It is visible, that a big part of the fluid domain is represented by the electrolyte gap, which is between anode and cathodes and is interrupted by the internal features.

To simulate the electrolyte flow the turbulence k-ω module of COMSOL Multiphysics® was used. The flow was defined to be incompressible. The Raynolds averaged Navier Stokes turbulence model and reference temperature of 293.15 K were chosen. The mass density of the electrolyte was defined with 998.2 kg/m³. The volume rate was selected to be 255 ml/min.

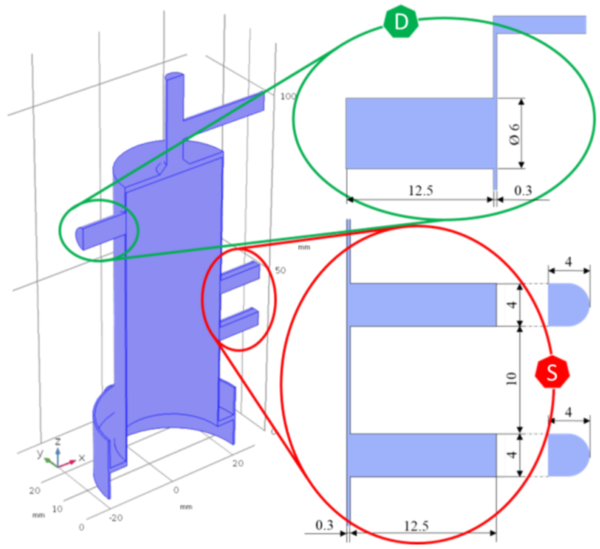


Figure 5 - 3D-sectional fluid domain with detailed view for the internal features S and D [4]

In figure 6 the flow field is shown. Here circulations in the internal features and at the inlet and outlet of the electrolyte can be seen. The most relevant zones are the machining areas and the circulations in the internal features. In figure 7 and 8 detailed views of the electrolyte flow velocity in the internal features are shown.

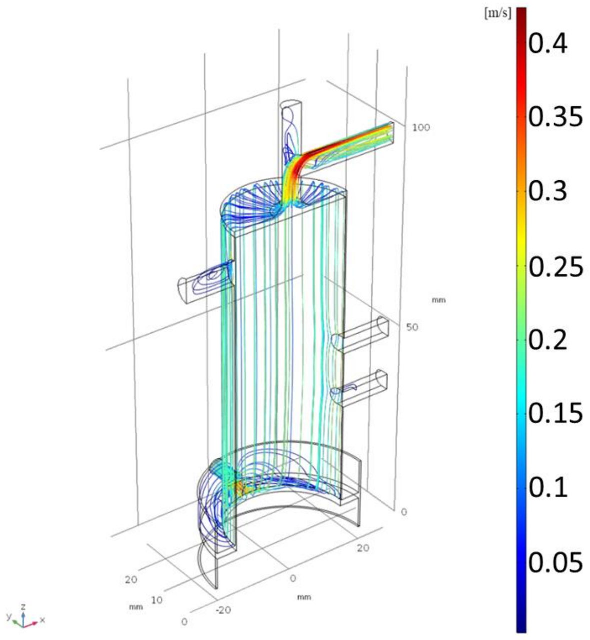


Figure 6 - Flow path of the electrolyte in the 3D-sectional model [4]

Figure 7 shows the internal feature D. Figure 8 shows the internal feature S. In both internal features complex flow characteristics occur and circulations result. It can be derived, that the electrolyte flow is a decisive parameter for designing the device. Especially in the feature S challenges have to be expected, because of the two bores in a row.

In summary the results of the simulation show that the selected flow rate is suitable to flush the complete device. Subsequently it is possible to carry out experiments to determine the dissolution behaviour and adjust the feature geometries.

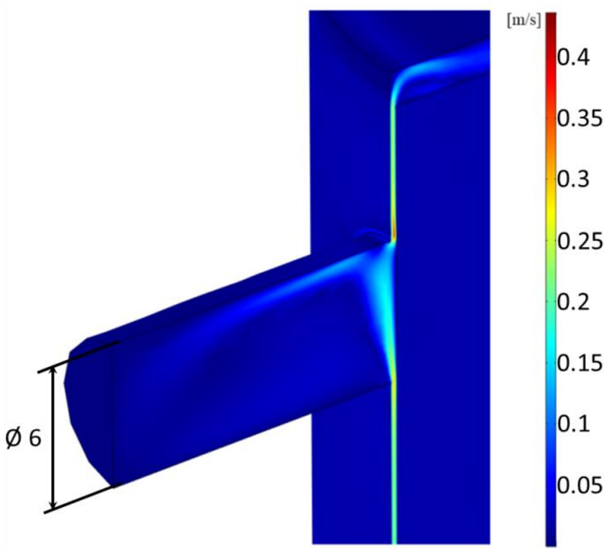


Figure 7 - Electrolyte flow velocity in the internal feature D [4]

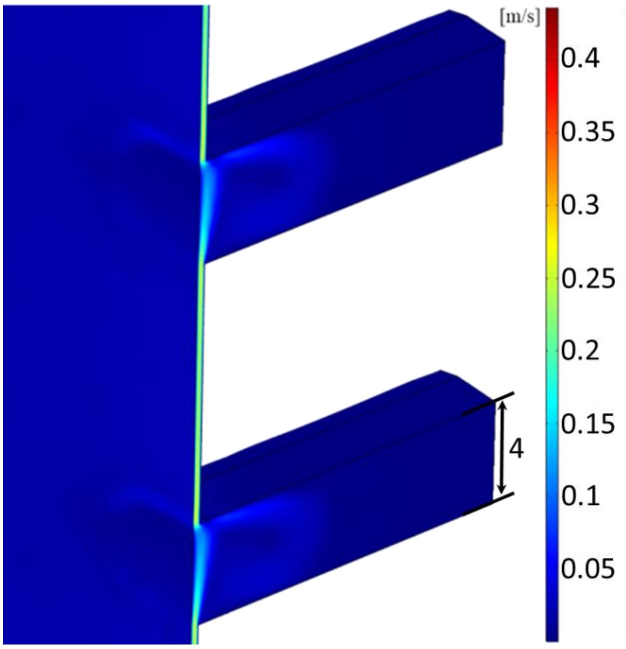


Figure 8 - Electrolyte flow velocity in the internal feature S [4]

11 SUMMARY AND OUTLOOK

In this study a device for a multi-cathode system was shown which was designed to machine different features of one workpiece by help of electrochemical machining with differentially switched currents. Based on fluid dynamics simulation the design of electrolyte flow was analysed. Different flow rates were investigated. It could be shown, that the selected flow rate of 255 ml/min is suitable to flush the complete device according to the simulation results. Furthermore it is expected, that in the feature S challenges can occure due to the circulation in the two bores in a row.

In future steps experiments will be carried out and further simulations for analyses of the mass removal at the features are planed.

12 ACKNOWLEDGMENTS

The project SwitchECM is managed by VDI Technologiezentrum GmbH and sponsored by the German Federal Ministry of Education and Research (BMBF) under grant number 13XP5030F.

Federal Ministry of Education and Research

13 REFERENCES

[1] F. Klocke; W. König: Fertigungsverfahren 3; Abtragen Generieren und Lasermaterialbearbeitung. Springer-Verlag, 2007 ISBN 978-3-540-48954-2

[2] J.F. Wilson: Practice and Theory of Electrochemical Machining. New York: Wiley-Interscience, 1971

[3] M. Hackert-Oschätzchen; S. F. Jahn; A. Schubert: Design of electrochemical machining processes by multiphysics simulation. Proceedings of the European COMSOL Conference, 2011. ISBN 978-0-9839688-0-1

[4] M. Penzel, I. Schaarschmidt, M. Hackert-Oschätzchen, A. Schubert: Numerical Investigation of Electrolyte Flow in a Multi-Cathode System for Electrochemical Machining. Proceedings of the European COMSOL Conference, 2018

[5] G. Meichsner, M. Hackert-Oschätzchen, M. Krönert, J. Edelmann, A. Schubert, M. Putz: Fast Determination of the Material Removal Characteristics in Pulsed Electrochemical Machining, In: Procedia CIRP, 7th HPC 2016 – CIRP Conference on High Performance Cutting, Volume 46, 2016, S. 123-126, DOI:10.1016/j.procir.2016.03.175

[6] A. Martin, T. Petzold, M. Hackert-Oschätzchen, G. Meichsner, A. Schubert: Experimental Derivation of Process Input Parameters for Electrochemical Machining with Differentially Switched Currents, In: Proceedings of the 14th International Symposium on Electrochemical Machining Technology, 2018

14 BIOGRAPHY

Tom Petzold M.Sc. is working on researches in electrochemical machining as scientific employee at the Professorship Micromanufacturing Technology of Chemnitz University of Technology, Germany.

PD Dr.-Ing. habil. Dipl.-Phys. Matthias Hackert-Oschätzchen is chief engineer research of the Professorship Micro-manufacturing Technology of Chemnitz University of Technology, Germany.

Dipl.-Ing. André Martin is leader of the teaching and research group Ablating Technologies and Manufacturing Equipment at the Professorship Micro-manufacturing Technology of Chemnitz University of Technology, Germany.

Prof. Dr.-Ing. Andreas Schubert is head of the Competence Center Micromanufacturing and Surface Technologies (KoMOT) at Fraunhofer Institute for Machine Tools and Forming Technology IWU since 2014. Since 2003, he leads the Professorship Micro-manufacturing Technology of Chemnitz University of Technology, Germany.

Dynamic Modelling and Implementation of an Optimized Control System for an Automated Guided Vehicle

J. Fernandes[1], C. Gilfillan[1], T. van Niekerk[1]

[1]Department of Mechatronics Engineering, Nelson Mandela University, South Africa

Abstract

Automated Guided Vehicles (AGVs) have come a long way since their inception and will play an integral role in the factories of the future as industry strives to meet new levels of efficiency, productivity and flexibility in support of mass customization. Although the technology has been available for many years, existing AGV designs incorporate numerous sub-systems that in many cases are overly complex and yet still lack the level of flexibility and openness warranted by Industry 4.0. In this paper, an AGV is developed that permits reliable and flexible operation within indoor environments. The system is designed to support maximum manoeuvrability, ease of maintenance, re-programmability, remote monitoring and control, maximum machine and operator safety and semi-autonomous operation. The authors also present a cost-effective, model-based trajectory correction system that rapidly corrects the vehicles pose in the event of wheel slippage or unequal wheel loading – a phenomenon especially common in vehicles that lack adequate suspension systems. The corrective action occurs in real-time in the intervals between receiving positional data from the on-board navigation sensor.

Keywords

AGV, Industry 4.0, Navigation, Mecanum wheels, PLC, TIA Portal, NAV350, Trajectory control

1 MOBILE ROBOTS FOR INDUSTRY

1.1 Automated Guided Vehicles

Robotic technology has been used for many years in industry in completing complex or repetitive tasks. However, robots are becoming more autonomous, flexible and cooperative and will eventually form the backbone of smart manufacturing systems [1]. Automated Guided Vehicles (AGVs) belong to a class of robots called mobile ground robots. They were initially introduced as a type of Flexible Manufacturing System (FMS) within the automotive sector [2] to enable efficient collection or delivery of parts to or from warehouses and manufacturing stations. They later also replaced material handling tasks by transporting material directly to the point of application [3].

Figure 1 - AGVs for various applications

As highlighted in Figure 1, AGVs come in many forms and sizes depending on their mode of application and always incorporate a central computer system as well as peripheral sensors responsible for navigation and safety related tasks. Safety remains of utmost importance in protecting the vehicle itself as well as objects or people that may be in its immediate environment. Considering the versatility of AGVs, they have more recently also found application in numerous other fields including in delivery and collection systems in hospitals [4], maritime operations such as in container terminals [5], military applications [6] and exploration.

1.2 Trajectory Control

In plant setups where AGVs are required to operate, a centralized control centre usually manages task allocations and ensures that all operations run efficiently and without collision [7]. Once the control centre has assigned a task to an AGV, it is the responsibility of the AGVs on-board navigation system to ensure that the vehicle maintains a suitable trajectory towards its final destination [8]. To work effectively, the navigation system must be able to accurately recognize and interpret not only its immediate environment but also its own underlying dynamics. It must also be able to provide regular feedback to the AGVs on-board computer regarding its current position. Several approaches for the navigation systems of mobile robots have been proposed and implemented over the last 30 years, including: guide-by-wire systems, magnetic anchoring, line following, global positioning, dead reckoning and laser based navigation [9].

With the more complex navigation systems such as the vision and the laser guided systems, the computational requirements are often quite high and the sensor sample rates are unsatifactory. As a result, real-time trajectory control can be difficult to achieve. To put this in to perspective, it could take as long as 250ms before the navigation system is able to provide an update of positional data to the control system - enough time for the vehicle to veer off its set trajectory.

The Mecanam-wheel AGV is a popular type of AGV that is used in environments where manoeuvrability is paramount, allowing the AGV to travel in any direction as well as to rotate about its axis. By controlling the torque and direction of rotation of each of the four Mecanum wheels, omni-directional locomotion is achieved. However, the major drawback of the Mecanum wheel approach is that it is impossible to quantify how much of the torque is divided into the forces acting parallel and perpendicular to the wheel's axis of rotation. It is therefore assumed that the torque is divided in such a way that the magnitude of the forces is equal and that the net force exerted on the wheel hub is at 45 degrees between the parallel and perpendicular forces as shown in Figure 2.

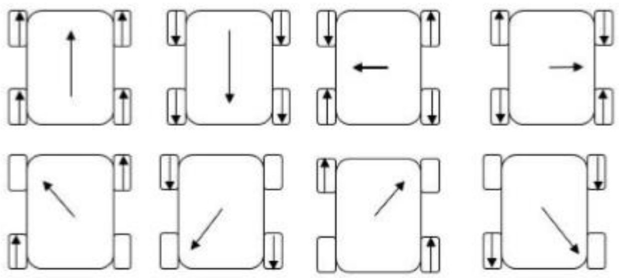


Figure 2 - Movement of omni-directional AGV as related to wheel velocities

For the AGV to be controlled effectively, each wheel must contribute to the resultant force acting on the AGV. This is achieved by controlling each wheel independently. In ideal conditions the AGV would behave predictably as each wheel would contribute equally to the net force thus achieving the desired motion. However, should one or more wheels suddenly loose traction due to an uneven surface, the AGV would veer off its set trajectory. Although a suitable suspension system could help to some degree in overcoming the sudden loss of traction, it is not always the ideal solution as oil spill and other environmental conditions could still result in wheel slippage. Moreover, the inclusion of a suspension system into the wheel assembly of the AGV not only increases the design complexity but also drives up the cost.

In this paper, the authors present a cost-effective, model-based trajectory correction system that rapidly corrects the vehicles pose in the event of wheel slippage or unequal wheel loading – a phenomenon especially common in vehicles that lack suspension systems. The corrective action occurs in real-time in the intervals between receiving positional data from the on-board navigation system.

2 AGVS FOR INDUSTRY 4.0

The world is on the brink of another major technological revolution that is likely to change the role of humans within the process of value creation. The first industrial revolution made use of technology based on water and steam power in order to mechanize production. The second industrial revolution utilized electric power to drive mass production. The third industrial revolution, which has also been called the digital revolution, made use of technology based on semiconductor electronics and information technology to automate production. Recent advances in the manufacturing and process industry have paved the way for the fourth and current industrial revolution or "Industrie 4.0" as termed by the German government. It builds upon the digital attributes of the third revolution and is characterized by the convergence of emerging technologies that together seek to eliminate the barriers between the physical, digital, and biological spheres [10]. Emerging technologies include but are not limited to the Internet of things (IoT), autonomous robots, simulation, big data, robotics, cloud computing, artificial intelligence, embedded systems, mobile internet and augmented reality [1] [11].

In order to survive in a highly competitive global marketplace, industries have to continually find innovative ways to reduce their manufacturing costs, while continuing to maintain production outputs. New and sustainable [12] business models based on Industry 4.0 are thus being explored that meet the following needs [13]:

- Short development periods
- Individualization of demand (i.e. mass customization)
- Flexibility and re-configurability
- Faster decision making procedures by means of decentralization of production management and coordination systems
- Resource efficiency

These new models move away from the traditional approach of mass production towards that of mass customization in which flexible and reconfigurable systems are key in addressing the needs of individual customers [14]. In this approach, the customer essentially becomes the "boss" and participates directly within the process of value creation [15]. To ensure that consumer demands are always met within a mass customization approach, manufacturing systems must be designed to allow for rapid changes in structure and control in order to adjust production capacity and variety in response to market changes.

Beyond the implementation of the trajectory correction system, the AGV discussed in this paper is also developed within an Industry 4.0 context.

This not only permits reliable and flexible operation in any indoor environment but the system furthermore supports maximum data transparency, maximum maneuverability, ease of maintenance, re-programmability, maximum machine and operator safety and semi-autonomous operation.

3 THE AGV PLATFORM

3.1 Overview

The AGV platform consists of six major sub-systems (as shown in Figure 3) including the mechanical system (chassis), the drive system, the power supply, the navigation system, the safety system and the control system. It is crucial that all sub systems are fully functional for the AGV to operate according to specification. In accordance with the requirements of Industry 4.0, the sub-systems must also be totally integrated and provide maximum data transparency at all layers to allow for remote analysis of system data. This could prove useful in a real plant environment where collected data could be used for the purpose of predictive maintenance.

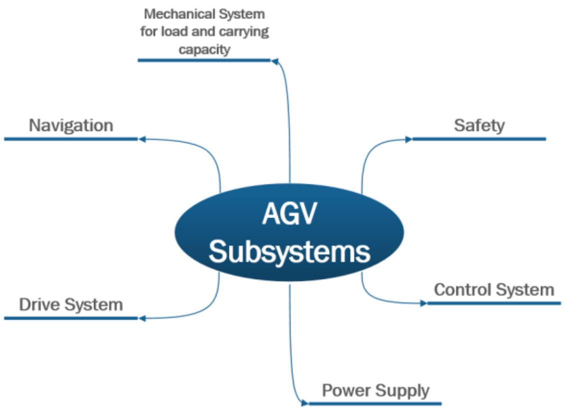


Figure 3 - AGV sub-systems

3.2 Mechanical System (Chassis design)

The chassis of the AGV followed a modular design to allow the AGV to be reconfigured for a variety of applications and to make the replacement of faulty parts as simple and timeous as possible. Figure 4 shows the modular chassis design to which all other components and subassemblies are fixed.

The fully assembled chassis weighs 330 kg and is capable of carrying a payload of up to 400 kg. Hollow steel tubing was used for the frame together with steel plates in certain areas for extra strength and rigidity.

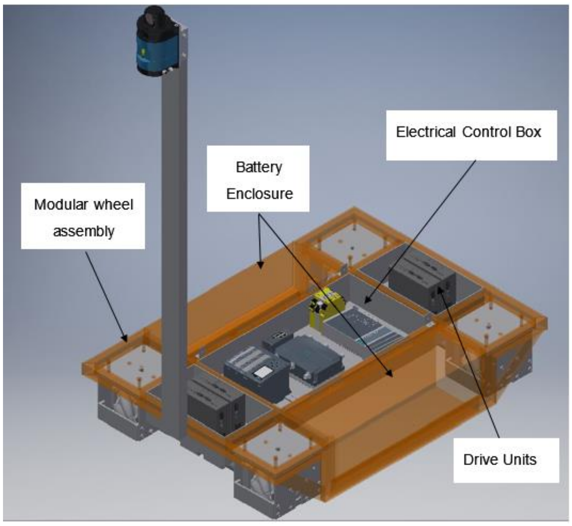


Figure 4 - Modular chassis

3.2.1 Modular Powertrain Design

A modular and compact powertrain design was achieved as highlighted in Figure 4. The dimensions of the shaft and the clutch housing were of particular importance for the assembly of the powertrain. The shaft was designed to reduce the overall length of the powertrain and to reduce any unwanted play which could potentially damage the shaft and lead to the failure of the entire powertrain.

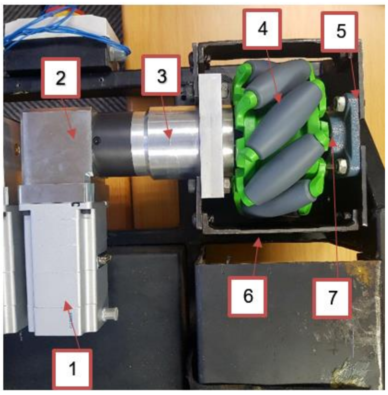


Figure 5 – Modular wheel assembly

The drive system is composed of four identical wheel assemblies each having a 48V Festo stepper motor (Figure 4 – (1)), a gearbox (Figure 4 – (2)), a coupling housing (Figure 4 – (3)), Mecanum wheel (Figure 4 – (4)), a cast-iron adapter (Figure 4 – (5)), a wheel housing (Figure 4 – (6)) and a support bearing (Figure 4 – (7)).

3.3 Electrical System

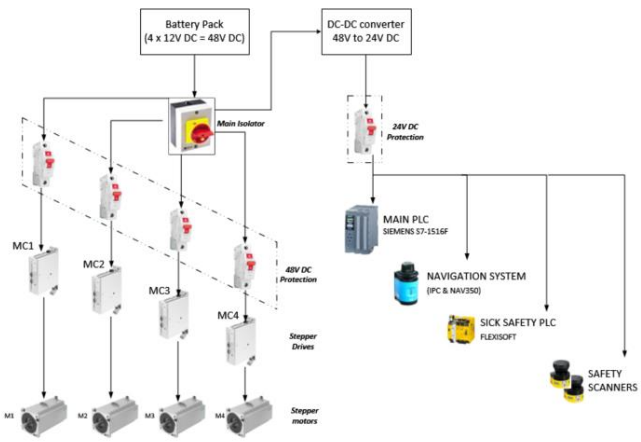


Figure 6 – Electrical distribution system

As depicted in Figure 5, the power distribution system consists of an isolated battery bank with four SMF101 12V 102Ah deep cycle batteries wired in series to provide the required 48V DC. Four 500W Festo stepper drives provide individualized wheel control. The DC-DC converter is a 48S24.6HCM from Calex Electronics and provides a consistent 24V DC supply to the PLC and the safety and navigation systems (see also Figure 6 and 7).

As shown in Figure 6, in order to attain complete data transparency at all layers of the design, each of the sub-systems was linked to the on-board Ethernet network with remote monitoring and control available via the wireless LAN. This allows for critical system information such as wheel velocity, current and torque to easily be retrieved and analysed either on-board or from a remote location.

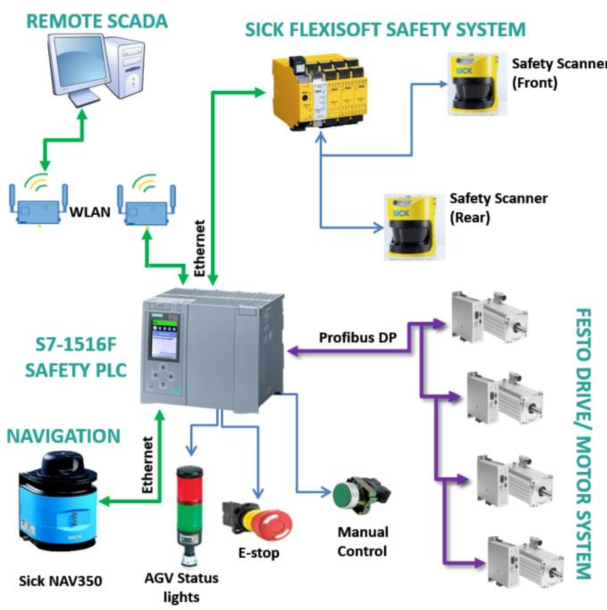


Figure 7 – Integrated electronic system

3.3.1 Modular Electrical Panel Design

The main electrical panel containing the electrical distribution, protection and control circuit was also designed in a modular fashion. This allows for easy access and replacement in the case of a breakdown. The panel can accommodate two sources of power; one from the AGVs on-board 48V DC battery supply and the other from a normal 240V AC wall outlet. This is convenient for technicians or programmers who may need to remove the panel from the AGV in order to carry out routine maintenance or upgrade related tasks. Since the panel integrates with the rest of the AGV through a single connector, faulty panels can be replaced with functional ones thus reducing the likelihood of extended downtime. The panel consists of the main control PLC (Siemens S7-1516F), a safety PLC (SICK Flexisoft), an Industrial PC (IPC) and various other industry standard components as depicted in Figure 7. The circuit layout was designed to take up minimal space yet remain as simple and understandable as possible.

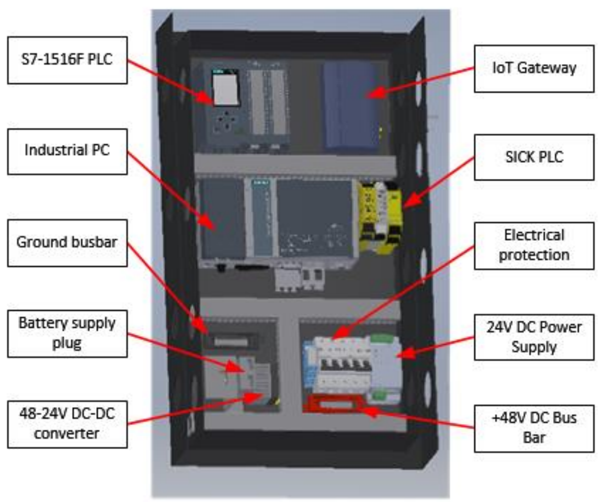


Figure 8 – AGV Electrical control panel

3.3.2 Safety system

In order to prevent harm to people, machines or objects occurring in the AGVs immediate operational environment, a reliable safety system was designed and implemented through a systematic approach thus fulfilling the fundamental requirements of the Machinery Directive 2006/42/EC. By conforming to the relevant harmonized safety standards, quality assurance throughout the AGVs life-cycle can be achieved.

Three main safety devices were included in the design as shown in Figure 6. These were an E-stop and two SICK S300 Mini-Remote laser scanners that continuously monitor the perimeter of the vehicle. Triggering of the e-stop or the detection of an obstruction in the AGVs path would result in immediate disconnection of power to the motors. The S300 laser scanners (mounted at the front and rear end of the vehicle) provide three re-programmable safety zones and therefore three different responses depending on the proximity of the detected obstruction. A SICK Flexisoft safety controller was used to evaluate safety conditions detected by the safety scanners. Detected safety conditions are transmitted over Ethernet to the main PLC where the signal is evaluated together with other relevant safety conditions to transition the AGV in to a safe state in the event of a hazard.

3.3.3 Navigation

Based on the requirements of automotive plants and/ or general manufacturing environments, a set of requirements were formulated to narrow down the choice of sensor for the AGVs navigation system. These requirements are listed in Table 1.

The core aim of the implemented navigation system would be to determine at all times during operation:

- Where the vehicle is located (localization)?
- Where the vehicle would proceed to if no action was taken to change its course?
- What needs to be done in order to safely arrive at a desired destination, and if needed, along a prescribed route.

No.	Navigation sensor requirement
1	Precision of travel: +/- 60 mm
2	Precision of positioning: +/- 5 mm (ground conveyor), +/- 30 mm (forklift)
3	Simple Operation: Route programming by means of teach-in
4	Advanced Operation: Auto-route selection by means of global coordination system (Using beacons)
5	Advanced Operation: Auto-route selection by means of global coordination system (no beacons) – i.e. landmark detection
6	The laser scanner must be for indoor use
9	The beam of the laser scanner must not be harmful to the human eye as the vehicle will be required to operate in the midst of people and other machines
10	The sensor must provide feedback of its absolute position
11	The laser scanner must provide the orientation of the AGV in the absolute coordinate system

Table 1 – Requiements for navigation sensor

Based on these requirements, the SICK NAV350 laser scanner was chosen. The NAV3xx product family are a well-developed range of 2d LiDAR sensors that work by sending out a laser beam which is then reflected back to the sensor by strategically placed beacons with a section of reflective tape. These sensors, though costly, are still much cheaper than their 3d counterparts. In the NAV3xx range, the NAV350 is the top-most model. With its in-built navigation algorithms, it is the best choice for the navigation for AGVs since it provides everything needed to make the navigation task simpler (coordinates and orientation information) including user friendly software.

In contour mode, the NAV350 can also deterine its pose by measuring the contours of its surroundings. In this mode, beacons are not required, however the navigation task becomes more complex. The NAV350 is set up either with the SOPAS ET (Engineering Tool) software from Sick or by sending it direct commands from the PLC [16].

3.4 Kinematic model for Mecannum wheel AGV

Since autonomous operation is the ultimate goal of an Industry 4.0 compliant AGV, it is important for the AGVs on-board computer to able to compute and predict its motion given a set of wheel velocities. To achieve this, a sound mathematical understanding of the AGV is required as summarized in (1-6) [17-19].

The inverse kinematic equations for the AGV are given in (1) and are used in order to determine the individual wheel velocities:

$$\begin{aligned} \omega_1 &= \tfrac{1}{r}\left(v_x - v_y - (l_x + l_y)\omega\right) \\ \omega_2 &= \tfrac{1}{r}\left(v_x + v_y + (l_x + l_y)\omega\right) \\ \omega_3 &= \tfrac{1}{r}\left(v_x + v_y - (l_x + l_y)\omega\right) \\ \omega_4 &= \tfrac{1}{r}\left(v_x - v_y + (l_x + l_y)\omega\right) \end{aligned} \quad (1)$$

The forward kinematic equations are given in (2-3):
Where, the longitudinal velocity in the X direction:

$$v_x(t) = (\omega_1 + \omega_2 + \omega_3 + \omega_4) \cdot \frac{r}{4} \quad (2)$$

And the transversal velocity in the Y direction:

$$v_y(t) = (-\omega_1 + \omega_2 + \omega_3 - \omega_4) \cdot \frac{r}{4} \quad (3)$$

And finally, the angular velocity of the AGV around its Z axis:

$$\omega_z(t) = (-\omega_1 + \omega_2 - \omega_3 + \omega_4) \cdot \frac{r}{4(lx + l_y)} \quad (4)$$

Where:
ω_n is the wheels angular velocity and n = 1,2,3,4.
ω_z is the angular velocity of the entire AGV. $v_x\ and\ v_y$ are the velocities in the global cartesian plain.
$l_x\ and\ l_y$ are the lengths of the AGV in meters and *r* is the radius of the mecannum wheel.

Furthermore, direction and magnitude in relation to the global coordinate axis can be achieved using the following equations:

$$\theta = \arctan(\frac{v_y}{v_x}) \quad (5)$$

$$v_{Rm} = \sqrt{{v_x}^2 + {v_y}^2} \quad (6)$$

4 TRAJECTORY CORRECTION SYSTEM

4.1 Simulink Model

The AGV was modelled and simulated in MATLAB using both the Simulink and Simscape environments (Figure 9). The primary objective of the simulation was to, as accurately as possible, represent the behaviour of the AGV under a variety of conditions that would result in loss of traction. For the purpose of this research, only wheel slippage was considered which may occur as a result of uneven floor surfaces or variations in the coefficient of friction between the floor and wheels. Understanding the behaviour of the vehicle under these conditions then aids in the development of robust control strategies that are able to rapidly correct the vehicles pose.

To simulate slip on one wheel, the wheel's angular velocity was momentarily altered. The altered angular velocities were then fed into the inverse kinematic equations and the resultant motion of the AGV could be observed.

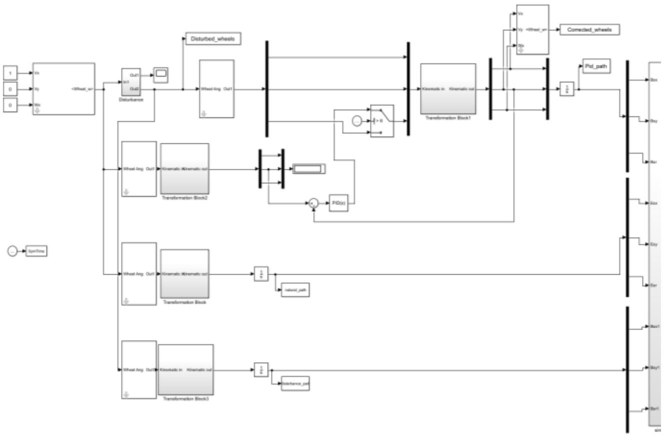


Figure 9 – Simulink Model of AGV

When simulating slip, a few assumptions were made as follows:

- One wheel would experience slip at a time
- The slip experienced will reduce the wheel's force contribution to zero
- Slip is experienced over a 1 second period
- Only velocity in the X direction was considered

4.2 Simulation Results

Figure 10 illustrates the application of a 1 second long disturbance (simulating wheel slip or loss of traction for 1 second) to the Front Left (FL) wheel of the AGV while the Front Right (FR) wheel, the Back Left (BL) and the Back Right (BR) wheel maintained constant set-point velocity. Five seconds prior to the disturbance, all wheels were rotating at equal velocities.

Because of this loss of traction, the AGVs orientation changes and the AGV follows a new trajectory. To remedy this, a control system was designed and implemented based on the plant model (Figure 10). The control system makes use of a single variable from the model as its control parameter. The parameter that was chosen was the orientation variable (θ). The orientation of the AGV is then compared to the expected orientation and an error signal is generated. The control algorithm uses this error and corrects the orientation of the AGV in such a way as to return it to its original trajectory.

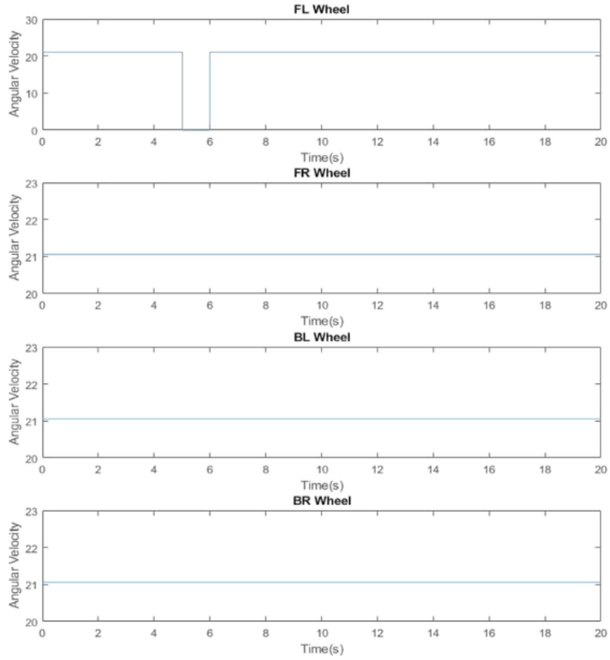


Figure 10 - Individual Wheel angular velocities of uncorrected simulation

Two assumptions were made in the design of the AGV control system:

- The AGV cannot move diagonally as it has physical limitations
- The controller only corrects the AGVs trajectory once traction has been restored

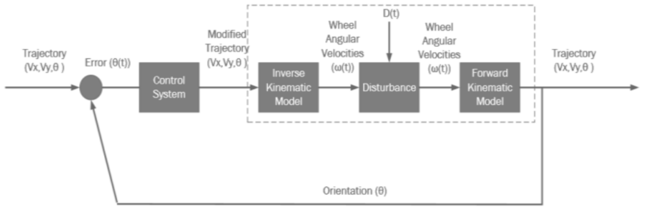


Figure 11 – Block diagram of control system

As shown in Figure 11, to restore the AGV to its original trajectory, it takes the combined effort from all four wheels. The correction only begins after 6 seconds which signifies the moment of restoration of traction to the wheels of the AGV.

Figure 12 depicts the paths that the AGV would follow if it remained undisturbed (i.e. if there was no loss of traction in any of the wheels), after the introduction of a disturbance (i.e. after loss of traction in one wheel for one second) and with the controller implemented to correct the trajectory in the event of a sudden loss of traction in one wheel. The blue dashed line represents the trajectory of the AGV after the simulated loss of traction in one of the wheels. The solid yellow line shows how the correction algorithm steps in and restores the trajectory back to its original state.

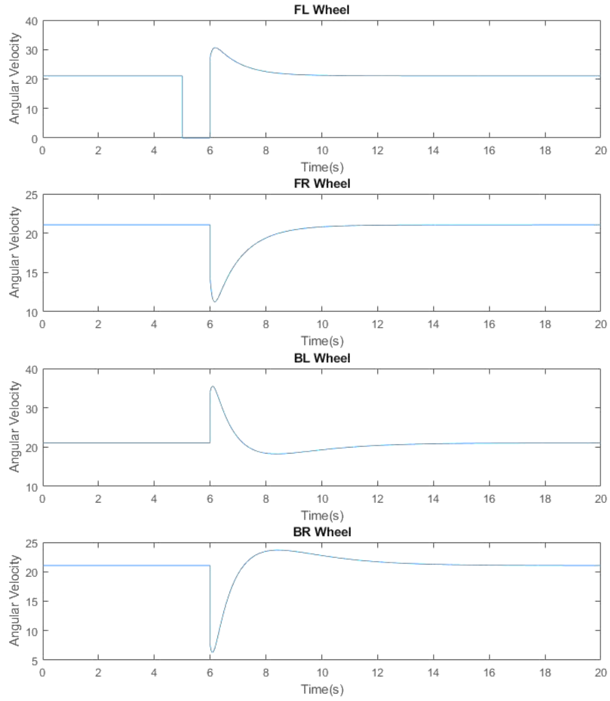


Figure 11 - Individual Wheel angular velocities of corrected simulation

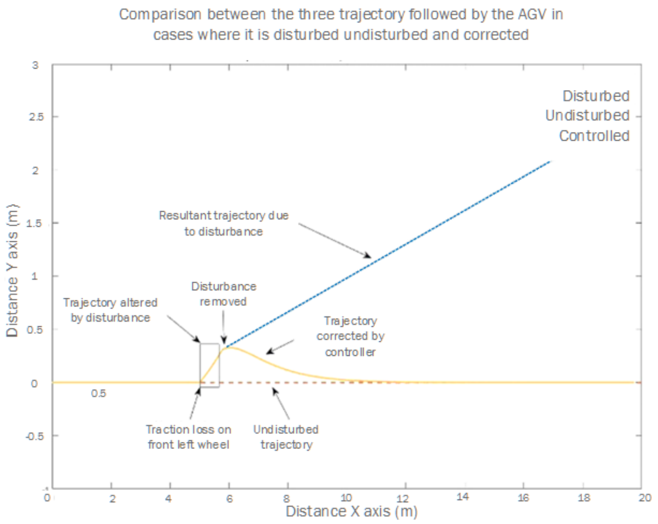


Figure 12 - Trajectory of undisturbed, disturbed and corrected AGV

5 CONCLUSION

In this paper, a materials handling AGV was developed for indoor use. In accordance with the requirements of Industry 4.0, the developed AGV supports maximum manoeuvrability, ease of maintenance, re-programmability, remote monitoring and control, maximum machine and operator safety and semi-autonomous operation. Furthermore, a trajectory correction control system is successfully designed and simulated in Simulink with the aim of correcting the vehicles trajectory in the event of wheel slippage. Simulation results show that the model based trajectory correction system would be an ideal and cost effective solution for real-time correction of the AGVs pose in between successive updates of the absolute positioning data provided by the on-board navigation system.

6 ACKNOWLEDGEMENTS

The authors would like to extend their gratitude towards the CSIR and the AMTC for funding of the research, Mr Sakkie Du Preez from Siemens for his assistance with the Siemens equipment, Mr Keenan Cloverly for his support with the Festo drive-train and Mr Stephen Eltz for his support with the SICK navigation and safety scanners used.

7 REFERENCES

[1] M. Rüßmann, M. Lorenz, P. Gerbert, M. Waldner, J. Justus, P. Engel, *et al.* (2015, Industry 4.0: The Future of Productivity and Growth in Manufacturing Industries.

[2] G. Ullrich, *Automated Guided Vehicle Systems - A Primer with Practical Applications* Berlin: Springer, 2015.

[3] G. Fedorko, S. Honus, and R. Salai, "Comparison of the Traditional and Autonomous AGV Systems," in *MATEC Web of Conferences*, 2017, p. 00013.

[4] M. Takahashi, T. Suzuki, H. Shitamoto, T. Moriguchi, and K. Yoshida, "Developing a mobile robot for transport applications in the hospital domain," *Robotics and Autonomous Systems,* vol. 58, pp. 889-899, 2010.

[5] J. Xin, R. R. Negenborn, and G. Lodewijks, "Trajectory planning for AGVs in automated container terminals using avoidance constraints: a case study," *IFAC Proceedings Volumes,* vol. 47, pp. 9828-9833, 2014.

[6] S. J. Julier and J. K. Uhlmann, "A counter example to the theory of simultaneous localization and map building," in *Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on*, 2001, pp. 4238-4243.

[7] V. Digani, L. Sabattini, C. Secchi, and C. Fantuzzi, "Towards decentralized coordination of multi robot systems in industrial environments: A hierarchical traffic control strategy," in *Intelligent Computer Communication and Processing (ICCP), 2013 IEEE International Conference on*, 2013, pp. 209-215.

[8] L. Sabattini, V. Digani, C. Secchi, G. Cotena, D. Ronzoni, M. Foppoli, *et al.*, "Technological roadmap to boost the introduction of agvs in industrial applications," in *Intelligent Computer Communication and Processing (ICCP), 2013 IEEE International Conference on*, 2013, pp. 203-208.

[9] P. Moore, S. Ujvari, J. Pu, J.-O. Lundgren, and C. Xie, "Intelligent semi-autonomous vehicles in materials handling," *Mechatronics,* vol. 9, pp. 881-892, 1999.

[10] K. Schwab. (2015, The Fourth Industrial Revolution. Available: http://www.vassp.org.au/webpages/Documents2016/PDevents/The%20Fourth%20Industrial%20Revolution%20by%20Klaus%20Schwab.pdf

[11] Y. Wang, H.-S. Ma, J.-H. Yang, and K.-S. Wang, "Industry 4.0: a way from mass customization to mass personalization production," *Advances in Manufacturing,* vol. 5, pp. 311-320, 2017.

[12] T. Stock and G. Seliger, "Opportunities of sustainable manufacturing in industry 4.0," *Procedia Cirp,* vol. 40, pp. 536-541, 2016.

[13] H. Lasi, P. Fettke, H. Kemper, and T. Feld, "Industry 4.0," *Business & Information Systems Engineering* vol. 6, 2014.

[14] M. M. Tseng and S. J. Hu, "Mass customization," in *CIRP Encyclopedia of Production Engineering*, ed: Springer, 2014, pp. 836-843.

[15] F. Shrouf, J. Ordieres, and G. Miragliotta, "Smart factories in Industry 4.0: A review of the concept and of energy management approached in production based on the Internet of Things paradigm," in *Industrial Engineering and Engineering Management (IEEM), 2014 IEEE International Conference on*, 2014, pp. 697-701.

[16] Sick. (2015, 4 January). *Nav350 Laser Positioning Sensor*. Available: file:///C:/Users/Gareth/Downloads/NAV350_Laser_Positioning_Sensor__2015-01-04__06-56-27.pdf

[17] N. Tlale and M. de Villiers, "Kinematics and dynamics modelling of a mecanum wheeled mobile platform," in *2008 15th International Conference on Mechatronics and Machine Vision in Practice*, 2008, pp. 657-662.

[18] P. Muir and C. Neuman, "Kinematic modeling for feedback control of an omnidirectional wheeled mobile robot," in *Robotics and Automation. Proceedings. 1987 IEEE International Conference on*, 1987, pp. 1772-1778.

[19] B. Q. Hamid Taheri, Nurallah Ghaeminezhad "Kinematic Model of a Four Mecanum Wheeled Mobile Robot " *Journal of Computer Applications* vol. 113, pp. 6-9, 2015.

8 BIOGRAPHY

John Fernandes is a Lecturer within the Department of Marine Engineering at the Nelson Mandela University. His research interests include mobile robotics and machine safety.

Cullen Gilfillan is studying towards a Master's degree in Mechatronics Engineering at the Nelson Mandela University. He is also involved within the School of Engineering as a professional Siemens SITRAIN training consultant.

Theo van Niekerk's research interests include Mobile Industrial Robotics, Automation of Manufacturing Systems, Process Control and Instrumentation. He lectures in the Department of Mechatronics.

Knowledge Discovery-Based Process Engineering – A Machine Learning Model Approach

G. Gudergan[1], D. Krechting[1], J. Reinartz[1]
[1]FIR Institute for Industrial Management at RWTH Aachen University

Abstract
The acquisition, processing and analysis of internal and external data is one of the key competitive factors for corporate innovation and competitive advantage. Many firms invest a significant amount of resources to take advantage of advanced analytics methods. Machine learning methods are used to identify patterns in structured and unstructured data and increase predictive capabilities. The related methods are of particular interest when previously undiscovered and unknown structures are discovered in comprehensive data sets in order to more accurately predict the outcome of manufacturing or production processes based on a multitude of parameter settings. So far, this knowledge is often part of the individual or collective knowledge of experts and expert teams, but rarely explicit and therefore not replicable for future applications. On the one hand, it is demonstrated in this paper how different machine learning algorithms have been applied to better predict the output quality in the process industry. On the other hand, it is explained how the application of machine learning methods could contribute to making previously not accessible process knowledge explicit. In order to increase the prognostic accuracy of the model different methods were combined, later on compared and evaluated within an industrial case. In this paper a comprehensive approach to knowledge-based process engineering is being presented.

Keywords
Machine Learning, Business Analytics, Process Industry, Knowledge Extraction, Data Analytics

1 INTRODUCTION AND OBJECTIVE

Collecting, storing and analyzing of data is nothing unknown for most companies today. Nevertheless, the sustainable optimization of future business performance on the basis of knowledge gained from given data records remains a challenge. This is especially true for small and medium-sized enterprises. In order to maintain their competitive advantage, companies must develop more specialized products due to growing customer requirements. Combined with the loss of expert knowledge, they consume more and more resources. However, technological advances in data collection, storage and processing and the spread of powerful analytical software and algorithms have made it easier for businesses to take advantage of digitisation, at least in theory [1]. In practice, when implementing advanced analytics, organizations face many organizational and procedural challenges in addition to technological challenges. The analysis of internal and external business data is generally seen as one of the key factors for business innovation and competitive advantage [2] [3]. Nevertheless, companies often find it difficult to identify the concrete added value for them. As data volumes continue to grow and IT solutions improve, they are seeking a viable way to take the first steps toward a learning organization. They strive towards breaking down knowledge barriers and improving their quality of decision-making. While the amount, quality and usability of data is beneficial to achieving this goal, resources, know-how and IT infrastructure requirements are an obstacle.

The aim of this paper is to demonstrate how different machine learning algorithms can be applied to better predict the output quality in the process industry. The objective is further to explain how the application of machine learning methods could contribute to making previously not accessible process knowledge explicit. Considering these objectives, the paper contributes to a better understanding of how machine learning can be applied and successfully implemented in order to achieve an important strategic goal of a company, namely to discover structures in existing knowledge to better meet future customer needs.

2 STUDY BACKGROUND AND PERSPECTIVE

The development of methods to support the exploitation of the full potential of machine learning is of particular interest against the background of a more general discussion on the application of machine learning in knowledge-intensive processes and organizations. A major debate is taking place, driven by the view machine learning and artificial intelligence (AI) have the potential to seriously damage established businesses and eventually lead to a significant loss of traditional jobs [4] [5]. At the same time, other experts recognize the limits of AI now and for at least the next 10 years [6]. Other experts disagree about the more negative consequences of AI [7]. Some recent studies from a variety of organizations, including the OECD, the World Economic Forum, Forrester and McKinsey, expect machine learning and AI will only result in a limited net loss of jobs taking new jobs created by machine learning and AI into account. There is obviously no uniform perspective on this. Since this is relevant against the background of the specific objective of this research, it is of great importance to investigate in detail the effects of machine learning on existing and future processes. The question of how human and machine-based knowledge are linked should also not be neglected. The perspective developed by Lichtenthaler [8], which is outlined below, is convincing here.

So far, the main focus of machine learning and AI has been on the efficiency gains achieved through automation based on intelligent algorithms. They have the ability to replace at least some human work. The potential for the development of new services and products through the combination of human and AI has so far received only limited attention [9]. This strategic perspective considering humans and AI to be complementary rather than competing, is used for this research and can be described in terms of a matrix mapping different types of relationships between machine learning and human intelligence (Figure 1). The relationship types are as follows.

Standard—At this level, AI takes the lead. Standard procedures requiring low levels of either human or machine intelligence are taken over by relatively low-level automation devices, e.g., the automation of production processes.

Substitute—Machine learning and AI provide a substitute for human intelligence. This type of development has attracted the most attention in recent discussions. One example is the increasing automation of jobs with limited complexity and easily definable qualification levels in the insurance industry. The focus lies on increasing efficiency.

Superiority—In this section of the matrix, human intelligence dominates not only now but also in the future. Unique human skills, such as creativity, empathy, judgment, storytelling, and motivation as speeches [10] [11], are not (yet) easily replicable via AI. It will be very difficult for AI to replicate human abilities fostering creativity and inventiveness.

Synthesis—In this domain, machine learning and AI are used to enhance human intelligence. For example, automated business intelligence systems discover and structure data and information previously being unstructured. The algorithm works in close collaboration with human researchers to update and realign its goals, i.e. to deliver effective business intelligence insights [12].

In the domain of synthesis, human and AI work together, creating an integrated intelligence. This study explicitly follows the perspective of synthesizing machine learning with human intelligence.

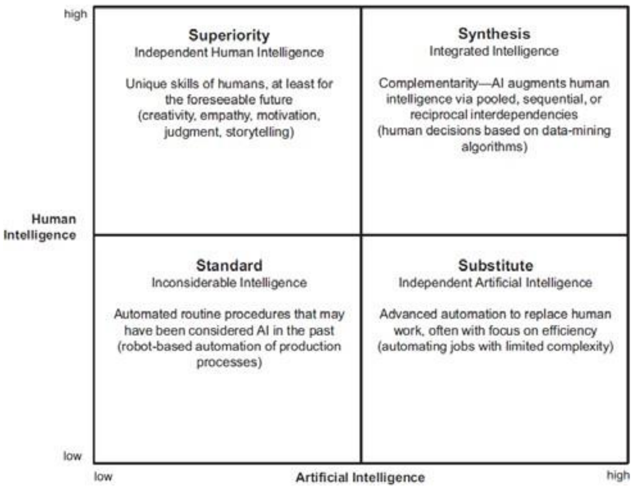


Figure 1 - Machine learning and human relationships

3 METHODOLOGY OF KNOWLEDGE DISCOVERY

3.1 KNOWLEDGE DISCOVERY

The pragmatic goal and objective of this paper is to draw a practical process to discover knowledge from structured or unstructured data. This process of discovering knowledge is often referred to as data mining. However, as explained below, data mining is only one step in the iterative process of knowledge extraction. [13] [14] defined the process of Knowledge Discovery in Databases (KDD) as “the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data” [14]. Their approach is considered the first attempt in order to establish standards for the process of knowledge discovery from complex and heterogeneous databases, therefore serving as the basis for many other methodologies. Fayyad et al. point out, their definition of knowledge has nothing to do with the philosophical or popular view. Instead, they consider knowledge as “purely user oriented and domain specific and determined by whatever functions […] the user chooses” [14]. According to Fayyad et al., extracting knowledge from databases is an interactive and iterative process. In order to obtain the desired knowledge, the user has to go through five central stages. The process starts with the *Selection* of the targeted data, being relevant for the user’s goal of the KDD process. The second stage – *Pre-Processing* – focuses on cleaning the targeted data in order to obtain a consistent set. In addition, the existing database can be enriched with additional attributes if required by the respective task. The third stage – the *Transformation* stage – consists of the transformation of the data using dimensionality reduction or transformation methods, i.e. the discretization of numerical values [15]. In the fourth stage – *Data Mining* – the main search for patterns of interest takes place. In section 3.2 the CRISP-DM will be presented as the preferred methodology to structure the Data Mining endeavor. In order to gain valuable knowledge, the fifth stage – *Interpretation/Evaluation* – has to be completed. Revealed patterns are interpreted and their usability is evaluated.

3.2 CRISP-DM

The project procedure is based on the well-known Cross Industry Standard for Data Mining (CRISP-DM). Although being conceived in 1996, CRISP-DM is still very relevant in research and practice when it comes to implementing data-mining projects in companies of various industries [16]. One of the central factors for the unbroken success is the high practical relevance, as CRISP-DM provides a comprehensive blueprint for project implementation for beginners as well as for data mining experts [17]. As depicted in Figure 1, CRISP-DM breaks down the life cycle of data-mining projects into six major phases [18].

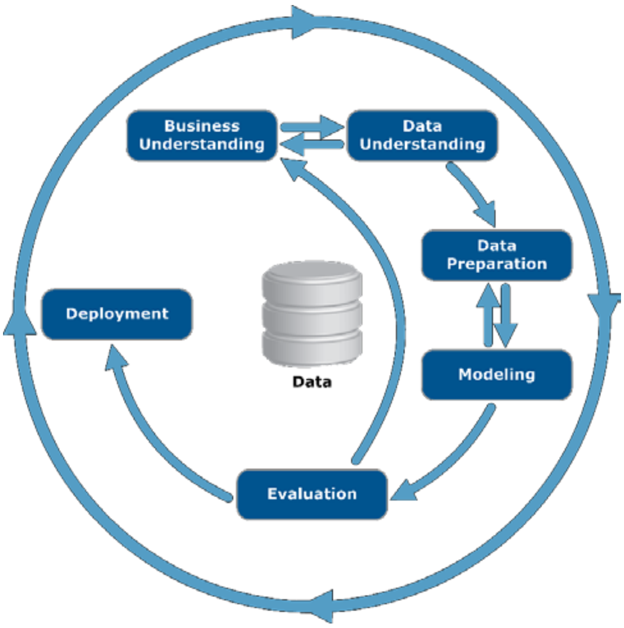


Figure 2 - CRISP-DM reference model

The six phases are not to be seen as a single, sequential procedure, but more as an iterative and repetitive process. According to [19], iteration is the rule rather than the exception. The phases can be described as follows:

(1) Business Understanding: Getting a clear understanding of the business problem and defining goals and requirements for the project.

(2) Data Understanding: Initial review of available data and assessment of overall data quality.

(3) Data Preparation: Preparation of data set for the modelling process.

(4) Modeling: Application of different data mining methods identifying patterns and optimizing parameters.

(5) Evaluation: Review of discovered patterns and selection of the model meeting the given task best.

(6) Deployment: Integration of the model into the customer's business processes.

3.3 CRISP-DM in the process industry

As a reference, literature for existing use cases extracting information confirming the usability and practicability of the CRISP-DM in the process industry was reviewed. A key objective of one case study was to develop an analytics platform architecture for process industry platforms. The principle was to split the platform between the plant operations platform, providing all relevant production data, and the cloud spanning data lab, processing and storing the data. This enabled a centralized approach to data analysis in the data lab with the ability to share knowledge between plants and projects [20]. In another use case from the petrochemical process industry, the CRISP-DM model achieved a robust result for the classification of a time series data set. By combining symbolic aggregate approximation in pre-processing with a decision tree classifier in modeling, they succeeded in developing a cost-effective method for feature selection and evaluation. The acquisition of important dependencies enabled them to predict the anomalous event and thus to react to it. For future research, they identified the visualization and understanding of CRISP-DM as needing improvement [21].

4 APPROACH / USE CASE

4.1 Challenge

In our use case, a mid-size German company from the process industry initiated the project to, in the future, facilitate customer-oriented compounds up to a batch size of one and partly automatize the development process. Representing the existing knowledge acquired by years of experimenting for stable compositions, a database of these chemical compounds provided the basis for our analysis. In a workshop, the expected benefits were elaborated to capture and visualize the company's knowledge, accelerate the compound development process and increase the documentation and development efficiency. The first step involved a pilot to prove the accuracy of Data Mining techniques, in which we designed a model out of the following algorithms:

4.2 Designing the model

4.2.1 Used algorithms

Naive Bayes

The Naive Bayes classifier is a method based on the Bayes theorem determining group affiliations of an attribute based on probabilities. With an ideal dataset, the Bayes classifier achieves a similar performance to methods of decision trees and certain neural networks. Apart from classification problems, the Naive Bayes method allows theoretical justification and accuracy verification of other classification problems. In use cases where attributes are highly interdependent, this can lead to inaccuracies within the results [22]. In return, the Naive Bayes classifier enables modelling with uncertain parameter probabilities through an iterative process only requiring final verification and quantification [23]. This classification method owes its great popularity above all to its high speed and accuracy when applied to large databases [22].

Generalized Linear Model

The Generalized Linear Model (GLM) is the generic term for many standard statistical models like linear regression, multiple linear regression and analysis of variance. In a nutshell, the GLM consists of three components, a random component, a systematic component and a function connecting these two [24]. The random components derive from observations of the same distribution as the statistical components. These statistical components depend on the chosen method like forming a linear estimator for multiple linear regression models [24]. In contrast to the regression analysis, the random component of the GLM does not require to be a Gaussian variable, which generalizes the classical regression analysis. On one hand, their field of application overlaps with the regression analysis. On the other hand, it extends to cases of binomial and other non-Gaussian variables.

C 4.5 algorithm

Decision trees give the possibility to predict outcomes by representing decision paths leading to a certain statement. The expression *decision tree* can be taken literally, visualizing a tree, where leaf nodes represent the test of an attribute and branches the possible values of an attribute [25]. In this study, the C4.5 algorithm for classification was chosen. Voted most influential algorithm by the IEEE International Conference on Data Mining in 2008, the C4.5 algorithm is a widely used statistical classifier. Generating a decision tree from training data, the C4.5 algorithm classifies its data relying on the concept of information entropy [26]. It provides accurate output for nominal and numeric data sets. In this case, the C4.5 algorithm sets a threshold maximizing the information gain per node. In comparison to other decision tree algorithms, C4.5 creates many branches requiring further pruning to improve accuracy [27]. The more efficient C5.0, which creator Ross Quinlan only markets commercially, succeeded C4.5 algorithm [26].

Deep Learning

Presented by Frank Rosenblatt in 1958, the deep learning approach is based on several artificial neurons, which are linked by weighted connections and a threshold. In single-layered networks, input neurons are fully connected to output nodes and the weighted connections need to be learned [28].

Combining multiple layers of networks by nonlinear functions enables the model to recognize complex patterns [28]. Due to diverse hyperparameters, the deep learning training process is far from trivial and requires some experience. But it is a useful tool to improve weak points of existing machine learning methods. With its ability to map hierarchically features, it is designated for image and speech processing [28].

Random Forest

A Random Forest is a supervised learning algorithm using the ensemble learning technique performing a bundling of a large number of decisions. The advantage of the Random Forest algorithm predicting classes from an existing database originate from its [29]

- high classification accuracy,
- skill to model complex interactions between the variables,
- possibility to handle missing data.

Furthermore, the Random Forest is an algorithm easy to use, only having a few hyperparameters to deal with [30].

The algorithm relies on bagging as the ensemble method, in which a weighted result is calculated from the individual forecasts in the form of a mean value, increasing the overall performance of the result [31].

Gradient Boosted Tree

Gradient Boosting is an ensemble learning method merging classes with high prediction errors in order to form a class with a low prediction error. Improving the overall accuracy, it iteratively adds or mixes components until the optimum between prediction bias and prediction variance is reached [32]. The Gradient Boosted Tree is the result of a decision tree classifier enhanced by gradient boosting methods. Furthermore, it offers an accuracy improvement for a variety of problems like cost-sensitive misclassifications, in which boosting optimizes the costs for error instead of the prediction errors.

Although boosting is not restricted to decision trees, it tackles its weaknesses regarding poor prediction accuracy and hard interpretability for large decision trees [33]. In comparison, algorithms like the C4.5 already weight instances making the gradient boosting method less efficient than for most algorithms.

4.2.2 Evaluation

As suggested by the CRISP-DM, the phases *Business* and *Data Understanding* were followed by the process of *Data Preparation*. In this phase, the given data set was prepared for the modeling process by transforming, cleansing and labeling all data. The initial data types were polynomials and real numbers as well as integers and binary values. Labeling the data correctly and removing gaps in the data set allowed proceeding to the *Modeling* phase.

The model`s task was predicting two interdependent values. The values consisted of two different data types. The first value, a polynomial data type, contained information about a specific ingredient for the production process. The second value, in the form of an integer, depicted the corresponding quantity of the ingredient. The prediction accuracy for the second value directly depended on the prediction quality of the first one. Considering the fact, the predictive process being iterative and taking place multiple times based on the previously predicted values, the risk of a strongly decreasing prediction accuracy was obvious. In order to prevent the automatic deterioration of the forecasts, it was important using the most suitable algorithm for the prediction of both values at all times. Therefore, preselected possible algorithms were run simultaneously in the model. The prediction results were directly compared to the training data set in order to evaluate the individual prediction quality of the algorithms.

The relevant algorithms for the polynomial value were three different types of Decision Trees, a Naive Bayes, a Generalized Linear Model, a Random Forest and several Gradient Boosted Tree algorithms. For each algorithm, the results were individually compared to the training data set using a confusion matrix. Thus the individual accuracy of an algorithm for each calculation step was determined. An exemplary result can be seen in Figure 3. In this case, the Gradient Boosted Tree proved to be most accurate.

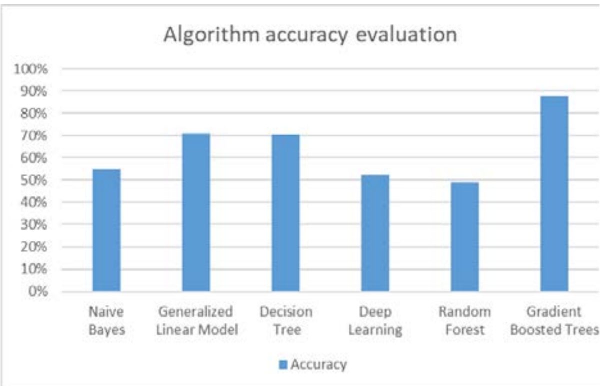


Figure 3 - Evaluation of algorithm accuracy

The second value having to be predicted by the model was a concrete value. Therefore, Decision Tree, Naive Bayes, Random Forest and Gradient Boosted Tree algorithms were relevant options for the calculation model. Again, each of the algorithms was tested and the results were evaluated by the Root Mean Squared Error, as shown in Figure 4. The evaluation was based on a distance measurement.

It should be noted, the prediction was considered individual and not dependent within the first prediction. A combination of Gradient Boosted Trees and Decision Tree was chosen for the prediction. This procedure ensures, as soon as new data sets are read in, the model optimizes itself and selects the correct algorithms for a new or extended data set.

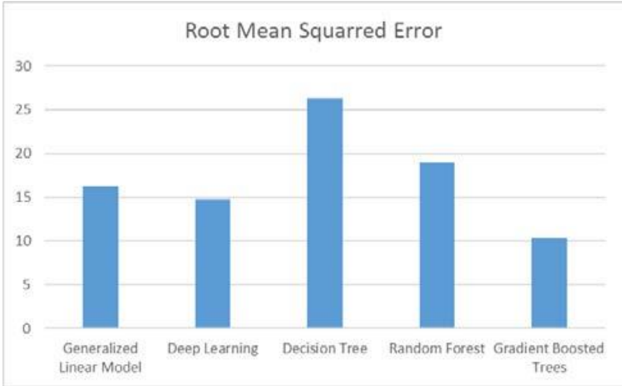


Figure 4 - Evaluation of algorithm accuracy

5 FINDINGS AND CONCLUSIONS

The development of methods to support the exploitation of the full potential of machine learning is of great interest. Hence, these methods will be of strategic importance for the development of new services and products through the combination of human and artificial intelligence. This strategic perspective, regarding human and artificial intelligence as complementary rather than competitive has been taken into account in this research. In a case study, it could be demonstrated, how different machine learning algorithms were used to gain insights from structured and unstructured data. The system was developed in a systematic approach based on existing knowledge discovery methods. Main findings are the simple application of machine learning on given data sets not being sufficient on its own. In fact, a large proportion of teamwork is needed in order to develop and maintain a system producing valuable results. The study furthermore reveals, despite high expectations being related to the application of machine learning and AI, parameter adjustments and process controls are still depending on human knowledge. The coverage can not be easily achieved by algorithms. The view of machine learning substituting human intelligence will not transform into reality in the near future. However, it could be demonstrated, synthesizing human and machine knowledge is a promising prospect to successfully apply machine learning. Hence, existing frameworks and tools can lead to a satisfactory machine learning based system in an efficient and time-saving process.

6 REFERENCES

[1] Stein, Brian, and Alan Morrison. "The enterprise data lake: Better integration and deeper analytics." PwC Technology Forecast: Rethinking integration 1 (2014): 1-9.

[2] Bhaduri, Saumitra N., David Fogarty, and Bhaduri. Advanced Business Analytics. Springer, 2016.

[3] Davenport, Thomas H., and Jeanne G. Harris. Competing on analytics: The new science of winning. Harvard Business Press, 2007.

[4] Agrawal, Ajay, Joshua S. Gans, and Avi Goldfarb. What to expect from artificial intelligence. MIT Sloan Management Review, 2017.

[5] Davenport, Thomas H., and Rajeev Ronanki. "Artificial Intelligence for the Real World." Harvard Business Review 96.1 (2018): 108-116.

[6] Kolbjørnsrud, Vegard, Richard Amico, and Robert J. Thomas. "How artificial intelligence will redefine management." Harvard Business Review 2 (2016).

[7] Plastino, E., and Purdy, M. 2018. Game changing value from artificial intelligence: Eight strategies. Strategy & Leadership 46(1): 16–22.

[8] Lichtenthaler, Ulrich. "Substitute or Synthesis: The Interplay between Human and Artificial Intelligence." Research-Technology Management 61.5 (2018): 12-14.

[9] Agrawal, Ajay K., Joshua S. Gans, and Avi Goldfarb. Human Judgment and AI Pricing. No. w24284. National Bureau of Economic Research, 2018.

[10] Plastino, Eduardo, and Mark Purdy. "Game changing value from Artificial Intelligence: eight strategies." Strategy & Leadership46.1 (2018): 16-22.

[11] Wilson, H. James, Paul Daugherty, and Nicola Bianzino. "The jobs that artificial intelligence will create." MIT Sloan Management Review 58.4 (2017): 14.

[12] Plastino, Eduardo, and Mark Purdy. "Game changing value from Artificial Intelligence: eight strategies." Strategy & Leadership46.1 (2018): 16-22.

[13] Azevedo, Ana Isabel Rojão Lourenço, and Manuel Filipe Santos. "KDD, SEMMA and CRISP-DM: a parallel overview." IADS-DM(2008).

[14] Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth. "From data mining to knowledge discovery in databases." AI magazine 17.3 (1996): 37.

[15] Azevedo, Ana Isabel Rojão Lourenço, and Manuel Filipe Santos. "KDD, SEMMA and CRISP-DM: a parallel overview." IADS-DM(2008).

[16] Piatetsky, Gregory. "CRISP-DM, still the top methodology for analytics, data mining, or data science projects." KDD News (2014).

[17] Beilecke, F., R. Messner, and R. Weskamp. "Wissenschaft inszenieren: Perspektiven des wissenschaftlichen Lernens für die gymnasiale Oberstufe." (2014).

[18] Chapman, Pete, et al. "CRISP-DM 1.0 Step-by-step data mining guide." (2000).

[19] Provost, Foster, and Tom Fawcett. "Data science and its relationship to big data and data-driven decision making." Big data 1.1 (2013): 51-59.

[20] Sarnovsky, Martin, Peter Bednar, and Miroslav Smatana. "Big Data Processing and Analytics Platform Architecture for Process Industry Factories." Big Data and Cognitive Computing 2.1 (2018): 3.

[21] Atzmueller, Martin, et al. "Explanation-aware feature selection using symbolic time series abstraction: Approaches and experiences in a petro-chemical production context." Industrial Informatics (INDIN), 2017 IEEE 15th International Conference on. IEEE, 2017.

[22] Bapat, Varsha, et al. "WSN application for crop protection to divert animal intrusions in the agricultural land." Computers and Electronics in Agriculture 133 (2017): 88-96.

[23] Schutte, Nicola S., et al. "Development and validation of a measure of emotional intelligence." Personality and individual differences 25.2 (1998): 167-177.

[24] Kirch, Wilhelm, ed. Encyclopedia of Public Health: Volume 1: A-H Volume 2: I-Z. Springer Science & Business Media, 2008.

[25] Michalski, Ryszard S., Jaime G. Carbonell, and Tom M. Mitchell, eds. Machine learning: An artificial intelligence approach. Springer Science & Business Media, 2013.

[26] Wu, Xindong, et al. "Top 10 algorithms in data mining." Knowledge and information systems 14.1 (2008): 1-37.

[27] Witten, Ian H., et al. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2016.

[28] Schulz, Hannes, and Sven Behnke. "Deep learning." KI-Künstliche Intelligenz 26.4 (2012): 357-363.

[29] Cutler, D. Richard, et al. "Random forests for classification in ecology." Ecology 88.11 (2007): 2783-2792.

[30] Couronné, Raphael, Philipp Probst, and Anne-Laure Boulesteix. "Random forest versus logistic regression: a large-scale benchmark experiment." BMC bioinformatics 19.1 (2018): 270.

[31] Breiman, Leo. "Bagging predictors." Machine learning 24.2 (1996): 123-140.

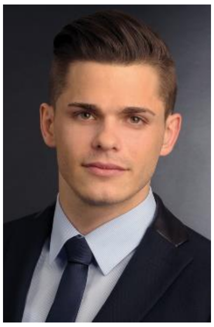
[32] Friedman, Jerome H. "Stochastic gradient boosting." Computational Statistics & Data Analysis 38.4 (2002): 367-378.

[33] De'Ath, Glenn. "Boosted trees for ecological modeling and prediction." Ecology 88.1 (2007): 243-251

7 BIOGRAPHY

Dr.-Ing. Gerhard Gudergan has a master in mechanical engineering and a PhD from RWTH Aachen University. He is head of the Business Transformation department at FIR at RWTH Aachen University. Since 2009, he is the head of business unit research at FIR. Previously, he established and ran the service management department and is the founding head of the service science innovation lab at RWTH

Denis Krechting has a master in industrial engineering. He has been a consultant and project manager at FIR since 2014. Within the scope of his work in the area of Business Transformation, he has accompanied projects at an OEM in the automotive sector with a focus on digitization strategies of companies and the application fields of business analytics.



Jonathan Reinartz holds a Master's degree in Industrial Engineering from RWTH Aachen University. Since 2016, he is project manager at the FIR of the RWTH Aachen in the research group Strategic Business Analytics in the field of Business Transformation. His current research focuses on data analysis and the development of business models for smart commercial buildings.

Development of a Universally Applicable Machine Learning Method for Optical Fault Detection in the Industrial Manufacturing Sector

C. Linder[1,2], J. Klöber-Koch[1], J. Hertle[2], S. Braunreuther[1,3], G. Reinhart[1]
[1]Fraunhofer Research Institution for Casting, Composite and Processing Technology IGCV, Augsburg, Germany
[2]Munich University of Applied Sciences, Munich, Germany
[3]Augsburg University of Applied Sciences, Augsburg, Germany

Abstract
In cyber-physical production systems, the generated amount of data is continuously increasing. At the same time, many companies do not possess the expertise to extract knowledge from the generated data. Therefore, machine learning solutions, realized as easily adaptable services, can support companies to benefit from machine learning methods with justifiable effort. The integrated quality control for transparent products, where the focus lies on the detection of faulty products, is an example for a process, where machine learning methods can support companies to increase their process efficiency. Besides production parameter based models, optical fault detection is used to facilitate an evaluation of the quality of the products. For that matter, an approach is proposed where a convolutional neural network (CNN) is used in combination with an automatic architecture discovery method. CNNs need no prior knowledge for feature extraction, which is why they can be used for such an approach that is supposed to be applicable for different use cases. Nevertheless, finding an optimal architecture for a specific dataset is difficult and requires expert knowledge. In order to overcome this challenge, a reinforcement learning based approach for finding an architecture automatically is used. The problem of data availability in manufacturing systems and the requirement of CNNs for a big amount of data is addressed by using data augmentation techniques. The developed approach is validated on real data from a production system.

Keywords
Manufacturing, Quality Control, Convolutional Neural Network

1 INTRODUCTION

Cyber-physical systems have become a widely used technology in the manufacturing environment [1]. This technology is a key enabler for the fourth industrial revolution [1]. Thus, more sensors and subsequently more sensor data are present in production systems. This data can be used to optimize industrial processes which can lead to cost savings [2, 3, 4]. Nevertheless, many companies do not possess the expertise to utilize the available data and extract knowledge from it [5]. A solution to this problem is the provision of generic cloud-based services which analyse the available data [2, 6]. Manufacturing companies could utilize these services, which are based on machine learning techniques, to extract knowledge from the available data for different use cases on the shop floor.

One suitable use case for such a service is the integrated quality control. Hereby, the focus is the detection of faulty products. Currently, the quality control of such parts is usually done manually [7, 8]. An automated approach, implemented as a service, has multiple benefits compared to a manual inspection [8]. The human inspection, e.g., is not always consistent in evaluating the quality of the products [8]. Furthermore, some defects are not well recognizable with the human eye [9]. In some production environments, unsafe environments prevent human inspection as there exist dangers for the employees [8]. Additionally, the need for line stoppage and exact positioning for manual inspection approaches can be reduced by using an automated inspection approach [8].

Therefore, an approach for an automated visual detection of faults on the surface of products is developed and realized as a service. The approach has to be universally applicable to ensure that different companies can benefit from it. Subsequently, the service needs to be independent of the product and its material. Furthermore, the approach should not be restricted to one fault type but rather be able to detect different types of faults on the products' surface. Another requirement for the approach is its usability without needing expertise in machine learning techniques.

The outline of this paper is organized as following: The next section summarizes relevant approaches for automated optical fault detection. Based on these approaches, a universally applicable machine learning method for optical fault detection is developed that is descript in section 3. Section 4 illustrates the application and results of the proposed approach for two different use cases. The paper ends with the conclusion and the outlook for possible further research activities.

2 RELATED WORK

In the following sections, automated optical fault detection approaches are examined. Automated optical fault detection for textured materials is usually based on texture descriptors to extract relevant features of fault images [10, 11]. The extracted features serve as input for fault classification methods. In recent years, approaches based on convolutional neural networks (CNNs) have also gained popularity for optical fault detection due to their success in similar image classification tasks [12, 13]. Therefore, the next sections focus on approaches based on texture descriptors and CNNs.

2.1 Texture descriptors

Texture descriptor methods can be divided into four categories: structural, model-based, statistical and transform [13]. Jiang et al. [14], e.g., use transform methods for surface analysis and Yazdchi et al. [15] use a model-based method for defect detection on steel surfaces.

Although approaches based on texture descriptors achieve satisfactory results if adjusted correctly, they still are problem dependent and usually have to be modified to new or different datasets [16]. Therefore, prior knowledge is needed to extract optimal features from the dataset and the quality of the results strongly depends on the extracted features [17].

2.2 Convolutional neural networks

CNNs are a type of deep neural network that is composed of multiple layers which generally consist of convolutional, pooling and fully connected layers [18]. In the first layers of the network, a combination of convolutional and pooling layers automatically extracts features. The extracted features then are combined in the last layers of the network by fully connected layers to conduct the classification [12]. Therefore, CNNs combine the steps feature extraction and classification in one process [12]. Hence, no prior knowledge is needed for feature extraction [12]. Figure 1 shows an example for a possible CNN structure, which is the result of the aforementioned classification and extraction steps.

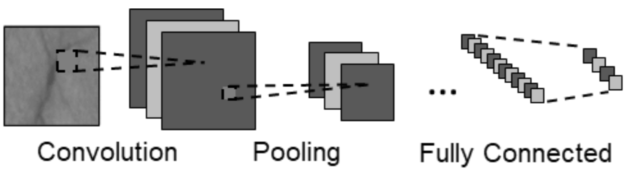


Figure 1 – Example of a CNN based on [19].

Additionally, the quality of the results generated by CNNs is comparable to the results of texture descriptor approaches. Masci et al. [10] and Makantasis et al. [11], e.g., compare different texture descriptor approaches with CNN approaches. In both cases, the CNN outperforms the texture descriptor approaches. CNNs already have been applied in different optical fault detection problems. Park et al. [12] use a CNN to detect faults on seven different materials. They compare the results generated by CNNs to texture descriptor approaches. In five out of seven cases, the CNN outperforms the texture descriptor approaches.

Nevertheless, CNNs have also drawbacks, which are addressed in the following sections.

2.2.1 Data augmentation

One drawback for the application of CNNs in the production environment is the requirement of a big amount of data [7]. In manufacturing systems, the amount of available surface defect data often is relatively small because of the costly aspect collecting labelled fault images [20]. However, this problem can be addressed by using data augmentation techniques. Hereby, the training data gets augmented by applying domain-specific transformations, e.g., random cropping [21]. Thus, the size of the training dataset can be increased. Chen et al. [22] for example use data augmentation techniques to increase the variety in the data and reduce overfitting for fault detection on material surfaces.

2.2.2 Design of neural network architectures

Besides the high amount of data required by CNNs, it also is difficult to design a neural network architecture for a given problem [23]. The reason is the requirement of having to set multiple so called hyperparameters, e.g., number, type or ordering of layers and layer dependent parameters, like the field size of a convolutional layer [24]. This step is extremely important and has to be done thoroughly as the architecture design strongly influences the performance of the CNN [19, 25].

The design of a neural network architecture can be developed with human expertise by crafting the architectures manually. This can be achieved by testing different architectures or applying modifications to existing architectures [24]. Affonso et al. [7] and Weimer et al. [16], e.g., preselected different architecture designs and then validated the architectures based on their performance.

Nevertheless, the before mentioned approaches require human expertise for finding an optimal architecture of the CNN for a specific problem. Other approaches try to automate this process and limit the required amount of human participation during architecture design. Therefore, Real et al. [26] propose a neuro-evolution based method, which does not require human interaction. Starting from just a single-layer model, they evolve deep architectures through mutation. Baker et al. [24] and Zoph & Le [23] propose reinforcement learning based approaches, by consecutively selecting layers at a time. Hereby, the validation accuracy serves as the reward for selecting a specific architecture.

However, the drawback of these automated approaches is the often poor performance compared to handcrafted architectures. In case the performance of the automated approaches is comparable to the manually designed architectures, the computation time to find the architecture often is very long.

3 APPROACH FOR A UNIVERSALLY APPLICABLE MACHINE LEARNING METHOD FOR OPTICAL FAULT DETECTION

In this approach, a CNN is used for feature extraction and fault classification because of its capability to automatically extract features without needing human intervention. The design of the neural network architecture is realized by an automatic architecture search method. Thus, no human intervention is required for this step as well. The requirement of CNNs for a big amount of data and the sparse availability of data in manufacturing environments is addressed by using data augmentation techniques.

3.1 Overview

The approach is divided into the following steps:

1. preprocessing and data augmentation
2. architecture search
3. final model

In the first step, the available data is preprocessed and data augmentation techniques are applied. Afterwards, an automated architecture search method is used to find a suitable neural network architecture for the processed data. In the last step, the found architecture is taken to create a final model which can be used for the application phase.

3.2 Preprocessing and data augmentation

In the first step, different data augmentation techniques are applied to overcome the sparse availability of data in the manufacturing environment. Thus, the risk of overfitting is reduced and the variety of images increased [22]. The data augmentation is realized by a rotation of the available images by 90, 180 and 270 degrees. Afterwards, six random crops are extracted from the original image to further increase the variety in the dataset. The data augmentation techniques, which are only applied to the training set, increase the amount of training data by the factor 10. Afterwards, the images are resized to a certain resolution to ensure that all images have the same size, which is another requirement of CNNs. The pixel-values of each image are normalized to ensures that, these values are in the range between 0 and 1 in order to save computation time.

3.3 Architecture search

A reinforcement learning based approach is applied to find an architecture for the processed dataset automatically. Therefore, the approach from Baker et al. [24], which uses Q-learning, is used because of its good trade-off between runtime and the number of parameters getting optimized.

In this approach, a learning agent is trained that chooses sequentially neural network layers via the ε-greedy strategy [27]. This layer selection process is modelled as a Markov Decision Process. Furthermore, the experience replay technique is used to accelerate the training procedure [28]. The experience replay stores the architecture and validation accuracy that serves as the reward for selecting a specific architecture.

The approach from Baker et al. [24] has to be adapted to the requirements of the focussed problem. The computation time, e.g., still is too high to be viable for industrial applications. Therefore, a reduced version is used by limiting the state and action space, reducing the maximum number of layers and modifying the hyperparameters. Subsequently, an architecture will be created consisting only of convolution (C), max pooling (P), fully connected (FC), dropout (D) and softmax (SM) layers.

The next sections describe the modified state and action space of the Markov Decision Process and the following, state specific actions. Furthermore, the selected hyperparameters of the Q-learning procedure will be described.

3.3.1 State space

The states are represented by the four parameters *layer type*, *layer depth*, *representation size* and *representation depth*. Each parameter is relevant for consequently choosing a layer at a time. The *layer type* contains information about the type of the current layer, which is important for choosing a valid action. *Layer depth* contains information about the number of convolutional and fully connected layers in the network and its depth. *Representation size* and *representation depth* represent the dimension of the signal. R*epresentation depth* is equal to the number of filters in the previous layer. *Representation size* offers information about the current signal size, which is reduced after, e.g., pooling operations. The *representation size* is defined by three discrete buckets as it was done in [24] to limit the state space. Bucket 1 (B1) contains all signal size values greater than 7, bucket 2 (B2) consists of values between 4 and 7 and bucket 3 (B3) consists of all values that are smaller than 4. Table 1 depicts all possible values that the parameters can attain.

State	Values
Layer type	ϵ {C, P, FC}
Layer depth	ϵ {1, 2, 3, 4, 5, 6}
Representation size	ϵ {B1, B2, B3}
Representation depth	ϵ {-, 32, 64, 128, 256}

-: original image depth

Table 1 – State space.

The maximum value for *layer depth* was set to 6 to limit the size of the architecture. Furthermore, a convolutional layer initializes *layer type*. *Layer depth* is initialized with “1”, *representation depth* with the original image depth and *representation size* with the corresponding bucket of the original image size.

3.3.2 Action space

The possible actions depend on the current state and therefore, on the different state parameters. Table 2 illustrates the possible actions for each layer type. An example for a action in a convolutional layer is the selection of a value for *number of filters* and *field size*. The *stride* parameter of a convolutional layer is always set to “1” to restrict the action space. The other state parameters restrict the possible actions further. If the *layer type* is a convolutional layer and the *layer depth* is smaller than “2”, *field size* cannot be set to “0”. Thus, very small architectures, which result mostly into a low accuracy, are avoided and the action space is restricted. In this context, a *field size* value of “0” means that no more convolution or pooling layers follow the current layer. Furthermore, the *field size* and *stride* parameter of a pooling operation depend on the *representation size* to prevent the occurrence of invalid operations. If *representation size* is equal to B2 only the operations {(3,2), (2,2), (1,1)} are allowed and if it is equal to B3 only the operation {(1,1)} is allowed where (1,1) means that no pooling is applied. Furthermore, the value of “0” for *number of neurons* means that the fully connected layer is skipped and no dropout is applied.

3.3.3 State transition

The transition to a new state depends on the current state and action. Table 3 shows the transition of the layer type. The *layer type* transitions into a fully connected layer, e.g., if the current *layer type* is a convolutional layer and the action-value for *field size* is “0”.

Layer type	Actions	Values
C	Number of filters	ϵ {32, 64, 128, 256}
	Field size	ϵ {0, 1, 3, 5}
P	(Field size, Stride)	ϵ {(5,3), (3,2), (2,2), (1,1)}
FC	Number of neurons	ϵ {0, 128, 256, 512}
	Dropout rate	ϵ {0, 0.4}

Table 2 - Action space.

Layer type	Conditions	Result (= new layer type)
C	If field size = 0	FC
	Else	P
P	If layer depth < 5	C
	Else	FC
FC	If dropout rate = 0	SM
	Else	D + SM

Table 3 – Layer type transition.

The parameter *layer depth* increases by 1 after each pooling layer, *representation size* changes only after specific pooling operations and *representation depth* changes after a convolutional layer according to the taken action-value of *number of filters*. Hence, the state transition is realized according to these rules and, continuously, an architecture is created. In the end, the architecture consists of a maximum of five consecutive convolution and pooling layers. These layers can be followed, depending on the taken action-values, by a fully connected and dropout layer. A softmax layer completes the architecture.

3.3.4 Selected hyperparameters

The selection of hyperparameters is oriented on the selection of parameters from the approach of Baker et al. [24]. The runtime is minimized by modifying some of these parameters. For the Q-learning procedure, different hyperparameters have to be selected. Q-learning depends on the Q-learning rate and discount factor. The Q-learning rate is set to a start-value of “0.1” and decreases every 200 Q-value iterations by a factor of “0.5”. The discount factor is set to “1”. The ε-greedy strategy depends on the ε-value and its decrease over time. Hereby, the value for ε remains at “1” for the first 800 iterations and afterwards decreases consistently to “0” for the following 800 iterations. Furthermore, the Q-value update starts after the first 800 iterations and is done in reversed transition order. Only architectures that reduce the image size to 32 or below are trained and considered for the Q-value update process due to hardware limitations. The experience replay depends on the number of architectures sampled and used for the Q-value update process. In the developed approach, 100 architectures are sampled and used for the Q-value update process. The learning scheme of the CNN requires the definition of multiple parameters. The Adam optimizer [29] is used with a learning rate of “0.0001”, “$\beta_1 = 0.9$”, “$\beta_2 = 0.999$” and “$\varepsilon = 10^{-8}$”. Furthermore, the number of training epochs is set to “20” and the batch size is set to “4”. The value “4” for the batch size is chosen due to limitations in the hardware. Weight initialization is realized via the Xavier initialization [30].

3.4 Final model

The last step of the developed approach is the creation of the final model that is used for the application. Therefore, the architecture that achieved the best validation accuracy of the architecture search is selected and used to create the final model. It is possible that the best validation accuracy can be achieved by multiple architectures. In such cases, the last found architecture of those generating the best results is selected for the model. Afterwards, the prediction performance of the selected architecture is improved by retraining the architecture for a longer period of time. Then, the model is ready to be applied in the industrial scenario.

4 APPLICATION AND RESULTS

In order to validate the developed approach, two different use cases with real data have been selected. One use case is the detection of faults on the surface of battery separators [31]. Therefore, images of different battery separators were taken on a test facility and manually labelled. The resulting dataset has 5 classes that are depicted in Figure 2. The five classes are: no fault (a), contamination (b), pinhole (c), inhomogeneity (d) and particles (e). For each class, 50 images are available and the dataset is then randomly split into a training and validation set so that the class distribution is preserved. 60 % of the data is being used for the training and 40 % for the validation set. After this separation, the training set consists of 150 and the validation set of 100 images. The amount of training samples increases to 1500 by applying data augmentation techniques. Furthermore, the images' resolution is automatically resized to 96 x 96 pixels and the pixel-values of each image are normalized.

Another use case is the detection of faults on transparent products from an injection moulding machine. 33 products are available and from each product, pictures were taken on a test facility. Due to the size of the component, twelve pictures have to be taken of each product. Afterwards, each picture is manually labelled into two classes. The first class consists of images without faults and the second class, accordingly, of images with faults. Typical fault types for this use case are scratches or streaks at the surface of the transparent component. Figure 3 illustrates examples where (a) depicts an image with a scratch in the marked area. This scratch is magnified in (b) whereas (c) shows an example without any fault on the surface. In the end, the dataset consists of 78 images with fault and 318 images without fault. Only 100 images without fault are used for the application to avoid class imbalance. This dataset is then randomly divided into 60 % training and 40 % validation data so that the class distribution is preserved.

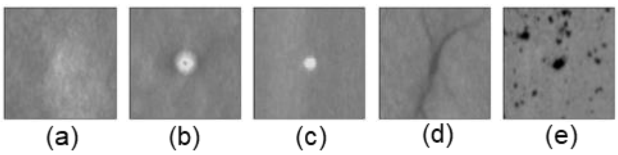


Figure 2 – Fault types of the battery separator dataset.

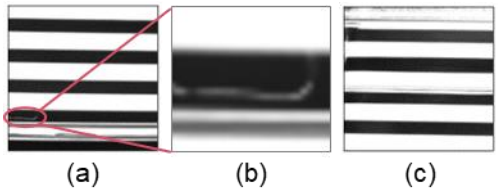


Figure 3 – Examples from the injection moulding machine dataset.

After applying data augmentation methods, the training set consists of 1070 images and the validation set of 71 images. Again, the images' resolution is automatically resized to 96 x 96 pixels and the pixel-values of each image are normalized.

In the next step, the automatic architecture search is applied to both datasets. Figure 4 depicts the prediction accuracy of the found architectures for the battery separator (a) and the injection moulding machine (b) dataset. The rolling mean of prediction accuracy with window "200" is used. In both cases, the mean accuracy is increasing after the exploration phase with the number of iterations and therefore, better-performing architectures are found with proceeding search time. The runtime for the automatic architecture search was 41 hours for the battery separator dataset and 24 hours for the injection moulding machine dataset using an NVIDIA GeForce GTX 1080 Ti and TensorFlow [32]. The difference in runtime mainly results from the difference from the amount of data of the two use cases.

In the last step of the approach, the architecture with the best prediction accuracy is selected to create the final model. Table 4 depicts the results for both use cases. Hereby, C(n, f) is a convolutional layer with n filters and field size f, P(f, s) a pooling layer with field size f and stride s, FC(n) a fully connected layer with n neurons, D(p) a dropout layer with a dropout rate of p and SM(n) a softmax layer with n neurons. For both use cases the architecture was trained 10 times. The hereby-achieved mean prediction accuracy on the validation set is shown in Table 4. The architecture for the battery separator dataset was trained for 80 epochs and in case of the injection moulding machine dataset for 20 epochs. After 20 epochs overfitting occurred in case of the injection moulding dataset, which is why the number of epochs has been restricted to 20.

The low accuracy result on the injection moulding machine dataset originates from the small amount of data with only 78 defect images and 100 images without defect. Furthermore, the detection of the faults in transparent parts is more difficult than the detection of the faults on the battery separator. The reason is that the faults are often very small and hard to detect on a transparent surface. Nevertheless, a valid architecture is found that is suitable for the dataset.

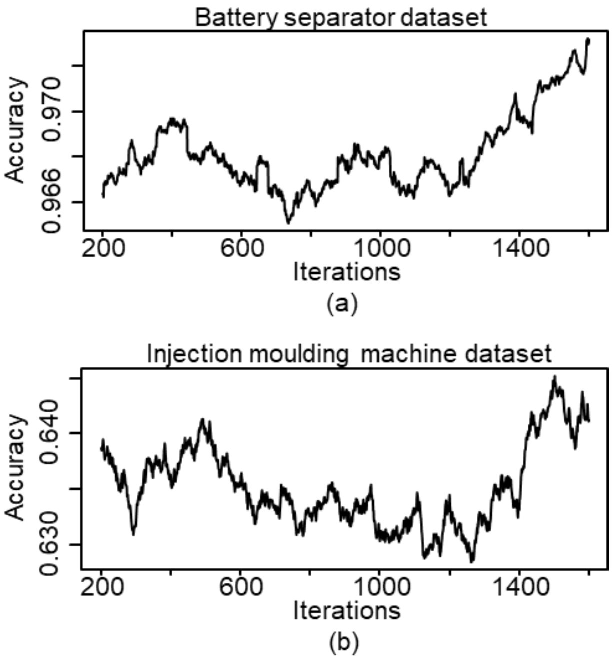


Figure 4 – Results from the architecture search.

5 CONCLUSION AND OUTLOOK

This paper proposes an approach for a universally applicable machine learning method for optical fault detection, which is validated on two different use cases. The proposed approach consists of a CNN that is combined with an automatic architecture search method and data augmentation techniques.

As a result, the proposed approach does not depend on any product, material or fault type because the feature extraction is done automatically by a CNN. Furthermore, the application of the proposed approach depends only on labelled images. Every other step of the approach can be done automatically. Therefore, the proposed approach is usable without needing expert knowledge in machine learning. Furthermore, a low runtime could be achieved for both investigated use cases by using a reduced version of an existing architecture search method. For both use cases, suitable architectures are found.

Dataset	Model architecture	Accuracy (%)
Battery separator	[C(128,5), C(128,5), P(3,2), C(256,3), P(5,3), C(64,3), P(3,2) C(256,1), P(5,3) FC(128), D(0.4), SM(5)]	99.00
Injection moulding machine	[C(64,5), C(64,1), P(3,2), C(32,3), P(2,2), C(64,5), P(5,3), C(32,5), P(5,3), FC(128), D(0.4), SM(2)]	68.59

Table 4 – Best architectures found.

The low amount of available data is addressed by using data augmentation techniques. This helps to reduce overfitting. Nevertheless, this problem could not completely be solved as overfitting occurred in case of the injection moulding machine dataset. In case of the battery separator dataset, a high accuracy could be achieved.

The approach has to be applied to further use cases to be fully validated. Especially the configuration of parameters for the architecture search has to be checked if it is suitable for other datasets as well. Future research should also investigate the aspect of class imbalance and parallelization.

6 ACKNOWLEDGMENTS

The OpenServ4P research and development project (www.openserv4p.de) is funded by the German Federal Ministry for Economic Affairs and Energy (BMWi) within the “Smart Service World” framework concept and is managed by the German Aerospace Center (DLR).

7 REFERENCES

[1] Monostori L., Kádár B., Bauernhansl T., Kondoh S., Kumara S., Reinhart G., …, 2016, Cyber-physical systems in manufacturing, CIRP Annals, 65(2):621-641.

[2] Smart Service Welt Working Group/acatech (Eds.), March 2015, Smart Service Welt – Recommendations for the Strategic Initiative Web-based Services for Businesses, Final Report, Berlin.

[3] Berger C., Hees A., Braunreuther S., Reinhart G., 2016, Characterization of Cyber-Physical Sensor Systems, Procedia CIRP, 41:638-643.

[4] Klöber-Koch J., Braunreuther S., Reinhart G., 2018, Approach For Risk Identification And Assessment In A Manufacturing System, Procedia CIRP, 72:683-688.

[5] Lee J., Kao H.-A., Yang S., 2014, Service Innovation and Smart Analytics for Industry 4.0 and Big Data Environment, Procedia CIRP, 16:3-8.

[6] Schreiber M., Klöber-Koch J., Richter C., Reinhart G., 2018, Integrated Production and Maintenance Planning for Cyber-physical Production Systems. Procedia CIRP, 72:934-939.

[7] Affonso, C., Rossi, A.L.D., Vieira, F.H.A., de Leon Ferreira, A.C.P., 2017, Deep learning for biological image classification, Expert Systems with Applications, 85:114-122.

[8] Newman T.S., Jain A.K., A Survey of Automated Visual Inspection, 1995, Computer Vision and Image Understanding, 61(2):231-262.

[9] Nayar, S.K., Weiss, L.E., Simon, D.A., Sanderson, A.C., 1988, Structured highlight inspection of specular surfaces using extended Gaussian images, In Optics, Illumination, and Image Sensing for

Machine Vision III, 1988 Robotics Conferences, Boston, MA, 235.

[10] Masci, J., Meier, U., Ciresan, D., Schmidhuber, J. and Fricout, G., June 2012, Steel defect classification with max-pooling convolutional neural networks, In The 2012 International Joint Conference on Neural Networks (IJCNN), IEEE, 1-6.

[11] Makantasis, K., Protopapadakis, E., Doulamis, A., Doulamis, N., Loupos, C., 2015, Deep convolutional neural networks for efficient vision based tunnel inspection, 2015 IEEE International Conference on Intelligent Computer Communication and Processing (ICCP), IEEE, 335-342.

[12] Park, J.-K., Kwon, B.-K., Park, J.-H., Kang, D. J., 2016, Machine learning-based imaging system for surface defect inspection, International Journal of Precision Engineering and Manufacturing-Green Technology, 3(3):303-310.

[13] Materka, A., Strzelecki, M., 1998, Texture analysis methods–a review, Technical university of lodz, institute of electronics, COST B11 report, Brussels, 9-11.

[14] Jiang X., Scott P., Whitehouse D., 2008, Wavelets and their applications for surface metrology, CIRP Annals, 57(1):555-558.

[15] Yazdchi, M., Yazdi, M., Mahyari, A.G., March 2009, Steel surface defect detection using texture segmentation based on multifractal dimension, In International Conference on Digital Image Processing, IEEE, 346-350.

[16] Weimer D., Scholz-Reiter B., Shpitalni M., 2016, Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection, CIRP Annals, 65(1):417-420.

[17] Bengio Y., 2009, Learning Deep Architectures for AI, Foundations and Trends® in Machine Learning, 2(1):1-127.

[18] LeCun Y., Bengio Y., Hinton G., 2015, Deep learning, Nature, 521(7553):436.

[19] Albelwi, S., Mahmood, A., 2017, A framework for designing the architectures of deep convolutional neural networks, Entropy, 19(6):242.

[20] Ren R., Hung T., Tan K.C., 2018, A Generic Deep-Learning-Based Approach for Automated Surface Inspection, IEEE transactions on cybernetics, 48(3):929-940.

[21] Zhang, C., Bengio, S., Hardt, M., Recht, B., Vinyals, O., 2016, Understanding deep learning requires rethinking generalization, arXiv preprint arXiv:1611.03530.

[22] Chen, W., Gao, Y., Gao, L., Li, X., 2018, A New Ensemble Approach based on Deep Convolutional Neural Networks for Steel Surface Defect classification, Procedia CIRP, 72:1069-1072.

[23] Zoph, B., Le, Q.V., 2016, Neural architecture search with reinforcement learning, arXiv preprint arXiv:1611.01578.

[24] Baker, B., Gupta, O., Naik, N., Raskar, R., 2016, Designing neural network architectures using reinforcement learning, arXiv preprint arXiv:1611.02167.

[25] Negrinho, R., Gordon, G., 2017, Deeparchitect: Automatically designing and training deep architectures, arXiv preprint arXiv:1704.08792.

[26] Real, E., Moore, S., Selle, A., Saxena, S., Suematsu, Y.L., Tan, J., …, 2017, Large-scale evolution of image classifiers, arXiv preprint arXiv:1703.01041.

[27] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., …, 2015, Human-level control through deep reinforcement learning, Nature, 518(7540):529.

[28] Lin, L.-J., 1993. Reinforcement learning for robots using neural networks, Technical report, DTIC Document.

[29] Kingma, D.P., Ba, J., 2014, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980.

[30] Glorot, X., Bengio, Y., 2010, Understanding the difficulty of training deep feedforward neural networks, In Proceedings of the thirteenth international conference on artificial intelligence and statistics, 249-256.

[31] Huber J., Tammer C., Schneider D., Seidel C., Reinhart G., 2017, Non-destructive Quality Testing of Battery Separators, Procedia CIRP, 62:423-428.

[32] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., …, 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, https://www.tensorflow.org/, Accessed 4 September 2018.

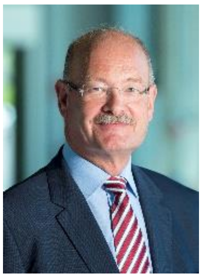
8 BIOGRAPHY

Christian Linder, B. Sc., graduated in Scientific Computing at the Munich University of Applied Sciences and is currently pursuing his M.Sc. in Applied Research in Engineering Sciences focusing on machine learning in production environments. The Master’s program is done in cooperation with Fraunhofer IGCV.

Jan Klöber-Koch, M.Sc., studied mechanical engineering at the Technical University of Munich. Since 2015 he is scientific assistant in the department production management of the Fraunhofer IGCV.

Prof. Dr. Jochen Hertle (professor at University of Applied Research Munich, Germany) focusses on data science, software development and entrepreneurship for IT start-ups. Prior to his university career, he gained experience in the telecommunications sector through various positions in development, business development & strategy and marketing.

Prof. Dr.-Ing. Stefan Braunreuther is head of the department for production management at the Fraunhofer IGCV in Augsburg. Since 2017 he is Professor at the Augsburg University of Applied Sciences, where he lectures regarding topics like factory operations and production management.



Prof. Dr.-Ing. Gunther Reinhart holds the chair for Industrial Management and Assembly Technologies at the *iwb* of the Technical University of Munich. Furthermore, he is Executive Director of the Fraunhofer IGCV in Augsburg. The *iwb* and the Fraunhofer IGCV conduct research into assembly technology, automation, robotics as wells as production management and logistics.

A Systematic and Narrative Literature Review of Maturity Models, Business Readiness and Requirments with the Focus on Artificial Intelligence Implementation

M.A. Nortje[1], S.S. Grobbelaar[1]

[1]Department of Industrial Engineering,
Stellenbosch University, South Africa

Abstract
The fourth industrial revolution is characterized by the combination of advanced digital technologies such as Artificial intelligence and physical technologies, which affect most aspects of life today. There are however challenges that businesses face in terms of adopting AI into their business. Workplace adoption barriers include the integration of machine and human capabilities. Literature identified that as long as people and machines are integrated in a process, the problem of coordination and communication of work between machines and people arise. The challenges that businesses face is thus the successful and effective implementation of AI into their business operations. This article investigates maturity models, business readiness and requirements with the focus on AI implementation, thus initializing the process of a solution towards the challenges stated. The paper is an initial effort in better understanding concepts and components for solving these challenges. The paper presents findings from a systematic and narrative literature review through which key processes, principles and requirements are explored.

Keywords
Artificial intelligence, implementation, readiness

1 INTRODUCTION

1.1 Background

The field of artificial intelligence (AI) transcends to just understand intelligence, but to develop it [7]. There are different definitions for AI. The definitions are based on two aspects. The first aspect is performance, which is divided into human performance and ideal performance. The second aspect is human-orientated tasks, which is divided into thought reasoning/processing and behaviours. Four definitions of artificial intelligence from the literature are shown in the table below.

	Human performance	Ideal performance
Thought reasoning and processes	Human thinking Automating activities with characteristics of human thinking, such as learning, problem solving and decision-making [3].	Rational thinking Study of calculations that allow perception, reason and action [5].
Behaviours	Human actions Creating machines that perform tasks requiring human-level intelligence [4].	Rational actions AI focuses on intelligent behaviour in artifacts [6].

Table 1 - Definitions of artificial intelligence [7]

Artificial intelligence is used in a variety of subfields that encompasses both general and specific aspects. General aspects include learning and perception, whereas specific aspects are proving mathematical theorems, diagnosing diseases, writing poetry and playing chess [7].

The impact of artificial intelligence could contribute an estimated $15.7 trillion to the global economy by 2030. It is estimated that $6.6 trillion of the total will most likely be due to increases in productivity [8].

The drivers of the economic impacts can be divided into three parts:

- Business process automation productivity gain [9].
- Productivity increases due to workforce and AI augmentation [9].
- Increases in user demands due to increased-quality products and services [9].

Businesses and companies wishing to capitalize on this opportunity will face some challenges with regards to AI. In a survey conducted by a business, which included 1500 senior executives that are “cognitive-aware”, it was determined that the difficulty to integrate cognitive projects into existing systems and projects was the highest perceived challenge among the senior executives [10]. The different types of challenges and their weighting can be seen in figure 1.

Businesses and companies thus are required to evolve towards a culture of software-management and engineering excellence. Capability maturity models provide a guide on methods towards gaining control of development and maintenance processes, as well as how to evolve towards an excellent software-engineering and management culture [13].

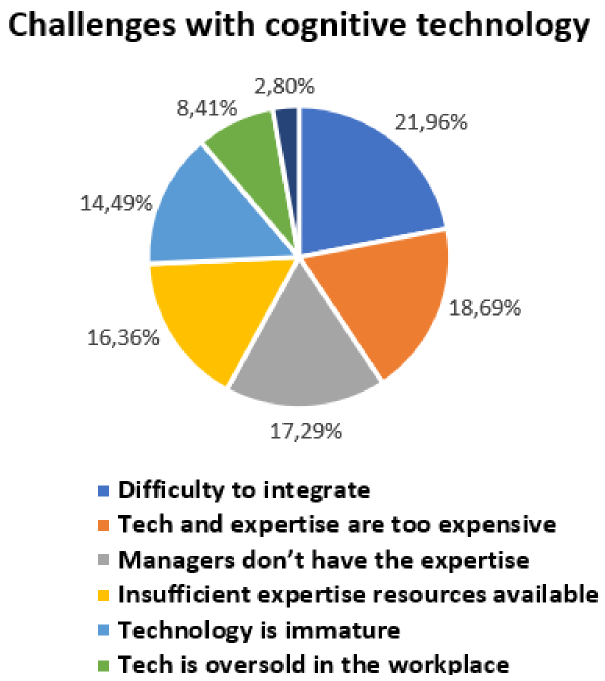


Figure 1 - Challenges associated with cognitive technology [10]

1.2 Problem statement

The integration of cognitive projects into existing systems and processes is seen as the biggest adoption barrier. The focus is towards the initial phases of AI implementation. This phase encompasses maturity models, business readiness and other factors that contribute to AI implementation. This can be seen as the centre of the Venn diagram that indicates the intersection between AI, Maturity models and implementation in Figure 2 below.

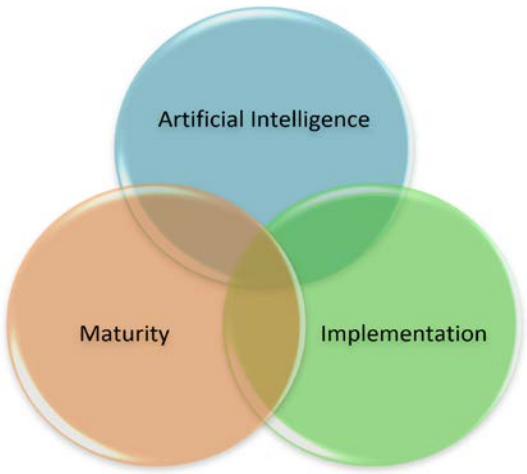


Figure 2 - Inter-relationship between key concepts.

The problem questions that arise are:

- What are the key building concepts and advantages of maturity models?
- What are the factors involved in implementing AI into business operations?
- What dimensions/categories are considered for determining AI readiness?

2 METHODOLOGY

A systematic literature review and narrative literature review was conducted. The steps can be seen in the figure below. The narrative literature review, reviews selective material, but the criteria for selection is not always apparent [14]. The focus for the narrative review is to supplement some concepts found in the systematic literature review. Conducting a systematic literature review is a valid method for the purposes of summarizing existing evidence of a technology and identifying the gaps in current research and literature around a specific topic [1].

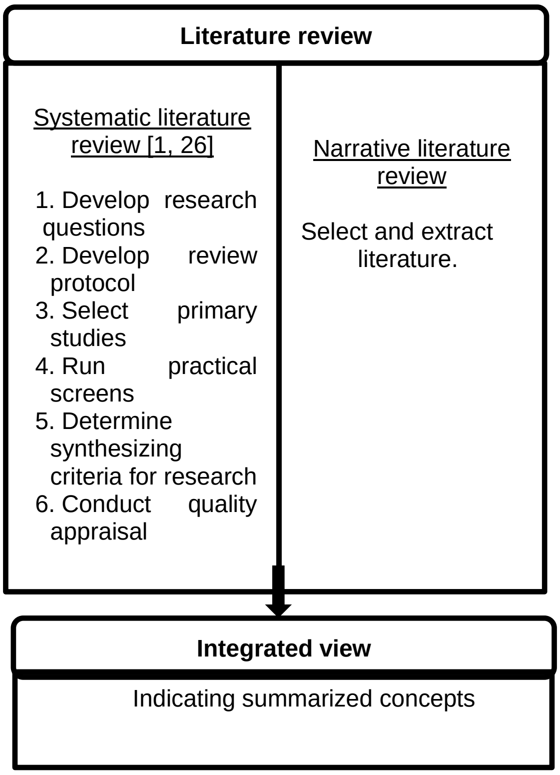


Figure 3 - Methodology of study

3 SYSTEMATIC LITERATURE REVIEW

3.1 Search Terms

The primary studies were searched in the Scopus database [2]. An iterative search was done by adding additional terms in the database search. The data base search was filtered to search for abstracts, titles and keywords. The search terms can be seen below.

Scopus database search (conducted on 05/06/2018)	
Search terms	**Number of studies found**
Artificial + Intelligence + Maturity	351
Artificial + Intelligence + Maturity + Implementation	55
Artificial + Intelligence + Maturity + Implementation + Model	35

Table 2 - Literature search results

It can be seen in the table above that that search term iteration reduced the number of primary studies to 35 studies. The author names, paper title, publication year, affiliations, abstract and methodology were retrieved from these primary documents. These documents were assessed according to the selection criteria.

3.2 selection criteria

The selection criteria for inclusion or exclusion of the literature is shown below.

Reference of criteria	**Criteria**	**Description of criteria**
CAT1	Availability	The full document must be freely available online.
CAT2	Language	English literature only.
CAT3	Types of literature	Conference reviews, lecture notes and lecture presentations are excluded.
CAT4	Applicability of literature	Ensure that the literature has relevant input towards the proposed study from evaluating the abstract.
CAT5	Academic robustness of the paper	The literature is evaluated in terms of validity of the methodology used, Number of citations, use cases, interviews and length of the paper.

Table 3 - Selection criteria

The selection criteria are applied in the sequence of CAT1, CAT2, CAT3, CAT4 AND CAT5. After the primary studies were validated against the selection criteria, 9 studies remained.

3.3 Quality assessment

The quality assessment criteria against which the papers were evaluated is indicated below in Table 4.

Quality assessment categories	**Description**
Completeness of document	Sufficient sections are included in the study, such as abstract, methodology and validation of research.
Methodology	Robust/satisfactory methodology which should be appropriate for the stated research question.
Aim/Goals	Clear and thorough statement of the research.

Table 4 - Quality Assessment criteria

The 9 studies' quality is assessed against quality assessment criteria. All the studies passed the selection criteria.

3.4 Data Extract*ion*

The papers were individually read and the information in regards with the research questions and quality criteria are extracted.

3.5 Dissemination

3.5.1 Citations

The number of citations for each of the papers that was obtained is shown in the figure below.

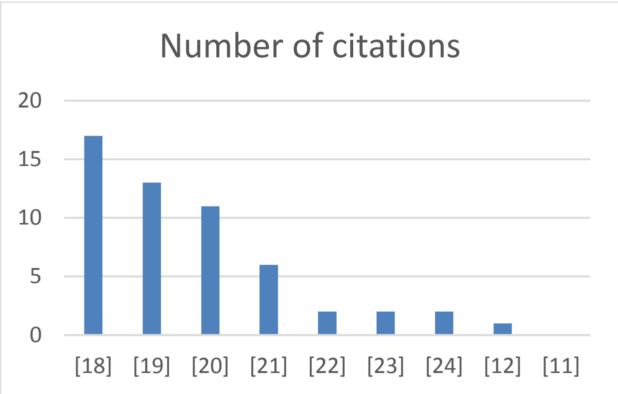


Figure 4 - Included studies' number of citations

From a researchers' stand point, this a valuable analysis in terms of determining the highly cited papers and from there determine the relevant authors. This increases the researcher's awareness of authors; whose work is applicable towards their own studies.

3.5.2 Digitalization

Digitization is the process of transitioning from analogue to digital forms. Digitalization is using digital technologies to improve value-adding opportunities and business models [11]. Digitization is a prerequisite for successful digitalization towards implementing software business processes [11]. AI technologies and digitalization projects require drivers such as data collection and data usage in the business, as well as resources, budgets and support from management [11].

There are two approaches towards digitalization. The approaches are top down and bottom up [11]. Questions involved with these approaches are:

- Top down [11]: How does digitalization change the business model? What are the benefits of digitalization for the business?

- Bottom up [11]: How can digitalization optimize current business processes? Which data sources are available? Which data analytics are required?

3.5.3 Output and advantages of maturity models

Capability maturity model integration describes discrete levels of process improvement, which can be applied towards processes and organizations [12].

3.5.4 Maturity model levels

The capability maturity model integration presented in the systematic literature review has five maturity levels and twenty-two key process areas. The maturity levels are initial, managed, defined, quantitatively managed and optimizing [12]. The maturity levels advance in ascending order. For example, if a business wants to achieve a level 3 CMMI, then all the key processes in level 2, as well level 3 must be satisfied [12].

This section provides a maturity model (capability maturity model integration) that focuses on quality improvements. There must thus be further studies conducted on the different application of maturity models, with the aim of integrating the literature into an effective maturity model, which assists in AI implementation.

4 NARRATIVE LITERATURE REVIEW

4.1 AI readiness

In terms of readiness for AI. Some businesses' ideas and concepts were obtained with regards to artificial intelligence readiness for this section.

Organizations wishing to improve or implement AI into their business can be divided into three groups namely, organizations that are new to the concept of AI, organizations that are ready to scale up AI and organizations that are widely implementing AI [16].

Organizations are at different stages in their AI journeys. To ensure continuous development and success for the next stages of their AI journeys, organizations need to ensure that the correct components are established in areas, such as technology and infrastructure, processes and models and resources and skills [16]. AI readiness is divided into three sequential types of readiness. These different types of readiness and their situation dependant corresponding components are shown below.

Phases of AI readiness	**Characteristics (where applicable)**
Foundational	Cloud resources
	Infrastructure platform
	Data sources
	Software packages
Operational	Skills and expertise
	Agile delivery

	Cyber security
	Governance, compliance and risk
	Operational management
Transformational	Business acceptance
	Business case clarity
	Business opportunity
	Strategic leadership

Table 5 - AI readiness phases and components [16]

A business elaborated on maturity and readiness assessment for artificial intelligence. The maturity model identified four dimensions into which maturity model elements are grouped in. The dimensions are [25]:

- Strategy
- Data
- People
- Legal

These can provide some insight into company dimension that must be recognized and evaluated when the readiness for AI is considered.

4.2 Maturity models

The maturity model (capability maturity model) is designed as a detailed referencing model towards improving processes of software developing companies [15].

The successful assessment of a capability maturity model should include the identification of implementation challenges regarding the model. Some challenges and factors associated with the implementation of capability maturity models identified are [15]:

- Change management.
- Ownership of the process.
- Awareness.
- Decision making.
- Evaluation of the team.
- Knowledge management.
- Meeting guidelines.

These challenges were identified for capability maturity models. Further research into different types of maturity model implementation challenges could provide some important information when developing a more AI focused maturity model for either implementation or readiness.

As AI forms a part of the development and growth of industry 4.0, company dimensions used to group maturity items, which in turn is used to determine business readiness for industry 4.0 are identified [17]. This can be useful to start understanding what company elements are involved in determining a business' readiness for artificial intelligence technologies. These company dimensions are:

- Strategy
- Leadership
- Customers
- Products
- Operations
- Culture
- People
- Governance
- Technology

5 SUMMARIZED AND COMBINED VIEW OF FINDINGS

The concepts from the accepted and chosen literature studies are identified, named, deconstructed and categorized. The concepts are:

- Non-readiness factors for implementing AI.
- Output and advantages of maturity models.
- Levels of maturity models.
- Capability maturity model implementation challenges.
- Maturity dimensions and categories for determining readiness for AI.

5.1 Non-readiness section factors for implementing AI

The digitalization section indicated some important prerequisites for AI implementation, such as digitization before digitalization. Businesses are required to focus on collecting the correct digitized data for the AI technologies. Digitalization projects must be supported by data collection, data usage, management and resources.

5.2 Generic maturity model factors

The output and advantages of maturity models, levels of maturity models and maturity model implementation challenges are summarised in the table below.

output and advantages of maturity models	**levels of maturity models [12]**	**Capability maturity model implementation challenges and factors [15]**
A referencing model towards improving processes of software developing companies [15]. (capability maturity model)	Initial ↓ Managed ↓ Defined ↓	Change management Ownership of the process Awareness

Describe discrete levels of process improvement [12]. (capability maturity model integration)	quantitatively managed ↓ optimizing	Decision making Evaluation of the team Knowledge management Meeting guidelines

Table 6 - Generic maturity model factors

5.3 Maturity dimensions for determining readiness for AI

This section is focused on indicating, identified maturity business dimensions used in maturity models, as well readiness components to assist in the development of a more focused and tailored maturity model for AI readiness in the future. The elements identified are from academic sources, as well as business sources. The different elements in relationship to the origin of it will be compared in the table below.

Maturity model element dimensions			
From academic sources		**From business sources**	
Maturity dimensions for industry 4.0 readiness [17]	Digitalization requirements [11]	AI readiness and maturity assessment maturity dimensions [25]	The AI readiness model components [16]
Strategy		Strategy	
Leadership			
Customers			
Products			
Operations			Operations
Culture			
People		People	
Governance	Management		Governance
Technology			
	Data	Data	Data
	Resources		Resources
			Security
		Legal	

Table 7 - Readiness dimensions/categories

It can be seen in the table above that there are some shared readiness elements between the different parts of literature. This provides some insight into which maturity dimension with regards to readiness can be considered when developing a more accurate maturity model for assessing a business' readiness for AI technologies in the future.

6 CONCLUSION

From the obtained literature it is concluded that before digitalization of processes through technologies, such as AI, the digitization phase must be completed. It is important to drive projects and initiatives towards gathering and maintaining data digitally. This is important due to it being a prerequisite for digitalization. The approaches for digitalization in terms of the questions provide some insights on the concepts that are considered when looking at the different approaches, such as the top down approach. It is evident that the focus is more on the business model and overall benefits for the organization, whereas the bottom-up approach focuses on business processes, data sources and data analytics.

Concepts, advantages and challenges of generic types of maturity models were identified. These were identified while searching for tailored maturity models, which focused on implementation and readiness for AI. There was a lack of fully applicable academic papers found on this subject in the systematic literature review. This indicates that further studies on this subject could yield large contributions towards a better understanding and more effective ways to implement AI into the business.

Business dimension used in maturity models for determining readiness, as well as readiness components were identified from academic papers and business sources. This provided some insight into what businesses with AI affiliations are focusing on when determining readiness in comparison to what the academic sources indicate. There are some shared dimensions between the different sources, which indicates further research into this might yield more accurate and effective maturity models for determining business readiness for AI.

It is important to note that the successful implementation of AI technologies in addition to the technological structures required, is the support and sponsorship of senior management and attention to change management in the business.

This paper serves as the initial effort in realizing what concepts are important in relation to artificial intelligence implementation, readiness and use of maturity models to assist in assessing these areas. Further studies on this subject could contribute greatly to a more effective and efficient way to implement a large driving technology of the 4th industrial revolution into businesses.

7 LIMITS OF THE STUDY

There are limitations towards the systematic and narrative literature review conducted. These limitations were experienced in various stages during the project. The limitations are:

- The Scopus search terms were increased to reduce the amount of papers to a more manageable amount.
- The inclusion and exclusion criteria further limited the amount of studies especially the language and online availability restrictions.
- The data extraction and interpretation of the data was conducted by one researcher, which increases the possibility of errors.
- There were not enough similar papers on the subject area found in the systematic literature review to perform a meta-analysis.

8 REFERENCES

[1] - Keele, S. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering*. [online] EBSE Technical Report. Available at: https://www.elsevier.com/__data/promis_misc/525444systematicreviewsguide.pdf [Accessed 2 Jun. 2018].

[2] - Elsevier.com. (2018). *Scopus | The largest database of peer-reviewed literature | Elsevier*. [online] Available at: https://www.elsevier.com/solutions/scopus [Accessed 3 Jun. 2018].

[3] - Haugeland, J. (Ed.). (1985). Artificial Intelligence: The Very Idea. MIT Press.

[4] - Kurzweil, R. (1990). The Age of Intelligent Machines. MIT Press.

[5] - Winston, P. H. (1992). Artificial Intelligence (Third edition). Addison-Wesley

[6] - Nilsson, N. J. (1998). Artificial Intelligence: A New Synthesis. Morgan Kaufmann

[7] - Russell, S. and Norvig, P. (n.d.). *Artificial Intelligence A Modern Approach*. 3rd ed. Pearson Higher Education.

[8] - PwC. (2018). *PwC's Global Artificial Intelligence Study: Sizing the prize*. [online] Available at: https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html [Accessed 15 Jun. 2018].

[9] - PwC (n.d.). *Sizing the prize What's the real value of AI for your business and how can you capitalise?*. [online] Available at: https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf [Accessed 2 Jun. 2018].

[10] - Davenport, T., Loucks, J. and Schatsky, D. (2017). *Cognitive technology survey: Early adoption insight | Deloitte US*. [online] Deloitte United States. Available at: https://www2.deloitte.com/us/en/pages/deloitte-analytics/articles/cognitive-technology-adoption-survey.html [Accessed 3 Jun. 2018].

[11] - Heberle, A., Löwe, W., Gustafsson, A. and Vorrei, Ö. (2017). Digitalization Canvas – Towards Identifying Digitalization Use Cases and Projects. *Journal of Universal Computer Science*, 23(11 (2017), 1070-1097).

[12] - Zhou, W. and Li, Y. (2012). Research on Quality Measuring of CMMI Cyclic Implementation in Software Process. *Journal of Software*, 7(8).

[13] - Paulk, M., Curtis, B., Chrissis, M. and Weber, c. (1993). *Capability maturity model, version 1.1*.

[14] - Cronin, P., Ryan, F. and Coughlan, M. (2008). Undertaking a literature review: a step-by-step approach. *British Journal of Nursing*, 17(1), pp.38-43.

[15] - Kumta, G. and Shah, M. (2002). CAPABILITY MATURITY MODEL A HUMAN PERSPECTIVE. *Delhi Business Review*, [online] 3(1). Available at: http://www.dbr.shtr.org/v_3n1/dbrv3n1e.pdf [Accessed 10 Aug. 2018].

[16] - The AI Readiness Model. (n.d.). *White Paper Data Center*, [online] intel. Available at: https://www.intel.com/content/dam/www/public/us/en/documents/white-papers/ai-readiness-model-whitepaper.pdf [Accessed 18 Aug. 2018].

[17] – Schumacher, A., Erol, S. and Sihn, W. (2016). A maturity model for assessing Industry 4.0 readiness and maturity of manufacturing enterprises. In: *Changeable, Agile, Reconfigurable & Virtual Production*. [online] Available at: https://www.sciencedirect.com/science/article/pii/S2212827116307909 [Accessed 10 Aug. 2018].

[18] - Prieto, A., Prieto, B., Ortigosa, E., Ros, E., Pelayo, F., Ortega, J. and Rojas, I. (2016). Neural networks: An overview of early research, current frameworks and new challenges. *Neurocomputing*, 214, pp.242-268.

[19] - Solar, M., Meijueiro, L. and Daniels, F. (2013). A guide to implement open data in public agencies. In: *12th International Conference on Electronic Government*.

[20] - Mehdiyev, N., Krumeich, J., Enke, D., Werth, D. and Loos, P. (2015). Determination of Rule Patterns in Complex Event Processing Using Machine Learning Techniques. In: *Complex Adaptive Systems*.

[21] - Kurze, C. and Gluchowski, P. (2010). Computer-aided warehouse engineering (CAWE): Leveraging MDA and ADM for the development of data warehouses. In: *Americas conference on information systems*.

[22] - Zhao, X., Hwang, B. and Low, S. (2016). An enterprise risk management knowledge-based decision support system for construction firms. *Engineering, Construction and Architectural Management*, 23(3), pp.369-384.

[23] - Srivastava, V., Sharfuddin, A. and Datta, S. (2012). Managing quality in outsourcing of high-end services: a conceptual model. *Total Quality Management & Business Excellence*, 23(11-12), pp.1315-1327.

[24] - XU, H. (2009). A development process of KMS based on systems engineering methodology. In: *12th International Conference on Computer Vision*.

[25] - Faktion.com. (n.d.). *AI Maturity and Readiness Assessment – Faktion*. [online] Available at: https://www.faktion.com/service/ai-maturity-and-readiness-assessment/ [Accessed 9 Dec. 2018].

9 BIOGRAPHY

Morne Nortje holds a B.Eng degree in industrial engineering from the university of Stellenbosch. Currently he is enrolled for a M.Eng degree in engineering management at the university of Stellenbosch, South Africa.

Prof Sara Grobbelaar is employed as an Associate Professor at the Department of Industrial Engineering at Stellenbosch University. Her research areas include innovation systems, innovation ecosystems, innovation platforms, technology management and inclusive development.

[26] - Okoli, C. and Schabram, K. (2010). A Guide to Conducting a Systematic Literature Review of Information Systems Research. *SSRN Electronic Journal*.

A Real-Time Scheduling System in a Sensorised Job Shop

S. Snyman[1], J. Bekker[2]
[1,2]Department of Industrial Engineering
University of Stellenbosch, South Africa

Abstract
Scheduling is a challenge that continues to trouble the operational phase of the manufacturing life-cycle and can be attributed to the complex, dynamic and stochastic nature of a manufacturing system. Scheduling problems are categorised in literature as *NP-hard* problems, where job shop scheduling forms part of the most difficult class of scheduling problems. This can be attributed to the combinatorial solutions that grow exponentially with problem size. The work described in this paper forms part of a research project with the objective to develop a prototype of a real-time simulation scheduling system, which is to serve as a decision support tool for real-time rescheduling of machine steps in a job shop. The purpose of this paper is to describe the development, testing and validation of the proposed system prototype. The prototype incorporates a cloud-based information system for the storage of data as well as a cloud-based simulation scheduler that generates schedules. It also includes sensors that keep track of the movement of jobs through the job shop. The development of the sensors will be described with reference to the hardware used as well as the functionality that is incorporated. The implementation of metaheuristics was also tested and compared to dispatching rules.

Keywords
Job shop, Real-time scheduling, Cloud-based simulation, Sensorised

1 INTRODUCTION

Manufacturing is defined as the transformation of materials into items of greater value by one or more processing and/or assembly operations [1]. The transformation is accomplished by combining different manufacturing components, such as machinery, tools, power, and manual labour. This description suggests that the process can be considered a complex engineering endeavour, where the coordination of people, material, equipment, and information to accomplish a manufacturing goal demands considerable time and effort.

When considering the manufacturing system life cycle, it can be divided into the design, planning, implementation, operation and termination phases [2]. Careful planning can be used to successfully overcome coordination challenges faced in the design and implementation phases; however, such challenges continue to persist during the operational phase. Scheduling is such a challenge, which can be attributed to the complex, dynamic, and stochastic environments exhibited by the manufacturing systems. The technological advances in cloud-based computing and the improved capabilities of sensor networks, have offered the opportunity of designing a real-time simulation scheduling system that can overcome the scheduling challenges, as well as the opportunity for the creation of software architectures to support these real-time scheduling systems.

This paper forms part of a research project with the objective to develop a prototype real-time simulation scheduler of a sensorised job shop, which is to serve as a decision support tool so that unexpected disturbances ([3], [4], [5]) in the shop can be overcome by the generation of new schedules with real-time data from the shop. This paper will focus on providing the background and defining the job shop scheduling problem, as well as describing the development of the proposed system. The functionality and testing of the system will also be discussed.

We subsequently provide a summary of the literature in Section 2, which is followed by an architecture of the proposed scheduler in Section 3. The development and implementation of the proposed solution will be described in Section 4, while the testing of the system will be discussed in Sections 5. Finally, the conclusion of the research project will be discussed in Section 6.

2 LITERATURE

This section will provide the required literature to understand the domain of job shop scheduling problems, as well as the methods that can be used to solve such problems. Previous architectures and systems that were developed will also be addressed.

2.1 The job shop scheduling problem

In this section, the job shop scheduling problem will be defined and discussed. The problem is commonly referred to in literature as an *NP*-hard problem. According to [6], the environment of the problem can be defined as:

- There is a set J of n jobs $\{J_1, J_2, \dots, J_n\}$ that needs to be processed on a set M of m machines $\{M_1, M_2, \dots, M_m\}$.
- Each job i consists of a finite and predetermined sequence of j operations O_i = $\{o_{i,1}, o_{i,2}, \dots, o_{i,j_i}\}$, where the sequence of operations is fixed.
- An operation may only be assigned to an available machine forming part of the set M.
- A machine can only process one operation at a time, and no pre-emption is allowed.
- Each operation $o_{i,j}$ has a fixed processing time $p_{i,j}$.
- The aim is to find a schedule for processing these n jobs on the m machines.

Sadeh [7] provides a good example of a job shop scheduling problem where there are four jobs and five machines. The different operations of the jobs are illustrated by the nodes and are labelled with the name of the operation, *e.g.* $o_{1,2}$ which is the second operation of the first job. The processing time is shown as an integer, while the allocated machine is illustrated with M_k, where k is the machine number.

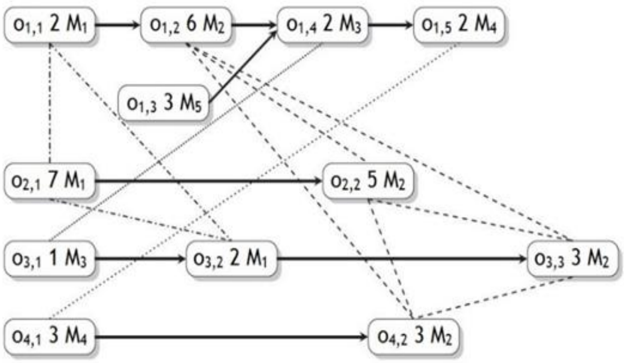


Figure 1 - An example of a job shop scheduling problem [7].

Error! Reference source not found. illustrates this example, where the arrows are the precedence constraints and the dashed lines are the capacity constraints.

An example of a precedence constraint is the precedence between $o_{4,1}$ and $o_{4,2}$. Which ensures that $o_{4,1}$ is processed before $o_{4,2}$. An example of the capacity constraint is between $o_{4,1}$ and $o_{1,5}$, where both operations need to be processed on M_4. Now that the job shop scheduling problem has formally been defined, the dispatching rules and performance indicators will be discussed in the next section.

2.2 Dispatching rules and performance indicators

Dispatching rules are used to guide the production sequence of jobs in manufacturing environments and ensure that operators know what the sequences are and which job to process next. Some of the common dispatching rules that can be found in literature for the job shop manufacturing environment, are listed in **Table 1**.

Dispatching rule	**Description**
Shortest processing time first	The job with the shortest processing time is processed first.
First-come-first-serve	The job that arrives first is processed first.
Most-important-job-first	The job with the highest priority or importance is processed first.
Earliest due date	The job with the earliest due date is processed first.
Critical ratio	The job with the smallest critical ratio is processed first. The critical ratio can be calculated by dividing the time until the due date, with the remaining processing time.
Minimum slack time	The job with the minimum slack time per remaining operation is processed first.

Table 1 - Common dispatching rules ([8], [9]).

Although these dispatching rules can be used to create a schedule, they will all produce different schedules. Therefore, performance indicators must be identified and used to assist the manager with selecting the correct schedule to meet the required needs. Common job shop performance indicators found in literature are provided in **Table 2**.

Performance indicator	**Description**
Average flow time	The flow time starts when a job enters the system and end when the jobs exits. The flow time is then averaged over a number of jobs.
Average queue time	The queue time is the total flow time of a job minus the processing time of the job. The queue time is then averaged over a number of jobs.
Average job lateness	Lateness is the difference between the completion date and the due date. Lateness can therefore be negative if the jobs is finished before the due date. The lateness is averaged over a number of jobs.
Average job tardiness	Tardiness is the amount of time a job finishes beyond the due date. If a job finishes early, the tardiness is zero. Tardiness is averaged over a number of jobs.
Makespan	Makespan is the total elapsed time to complete a number of jobs.

Table 2 - Common performance indicators ([8], [9]).

Dispatching rules are not the only methods that are used in literature to generate machine schedules. There are several different metaheuristics that have been used to solve the job shop scheduling problems. In this project, the authors incorporated three metaheuristics, which include a traditional and hybrid simulated annealing (SA) algorithm. The hybrid algorithm makes use of a *2-opt move* (as proposed by [10]) to search the solution space. The third metaheuristic consist of a population of 50 solutions where a 2-opt move is applied to each solution in every iteration, where the best solution in the population is retained after every iteration. All three these metaheuristics make use of an integer encoding, as presented in [11]. The metaheuristics are an important component of the scheduler, but are not the primary focus of this paper, therefore they are not discussed further.

The identified dispatching rules and metaheuristics will be incorporated into the simulation scheduler, while the scheduler will conduct tests to determine which rule performed best for each performance indicator.

2.3 Developed scheduling systems found in literature

Lu and Sheng [12] proposed some key technologies and a new implementation framework for a monitoring and management system of a discrete manufacturing process that is based on Internet of Things (IoT) technology. The framework is made up of the perception layer, network layer and the application layer.

The perception layer consists of the physical resources, *i.e.* the RFID and sensor technology, as well as the configuration of various types of sensors with a wireless network. In the network layer, the signal is extracted, debugged, analysed, packaged and stored. Finally, in the application layer, manufacturing resources are monitored, data mining is conducted, processing quality is diagnosed, dynamic scheduling is performed, and the material distribution service is optimised.

Zhang and Tao [13] also propose and develop an IoT manufacturing system for rescheduling production plans in a typical manufacturing environment. The proposed system incorporates RFID sensors, for capturing real-time data from the shop floor. It also uses simulation modelling together with a genetic algorithm to perform the rescheduling of production plans. Finally, the system has a cloud computing based manufacturing resource configuration method that is responsible for pairing tasks and machines. This implementation is however of a typical manufacturing system, while the focus of this study is to develop such a system for the job shop manufacturing domain specifically.

Even though both these systems are very sophisticated, they assume the processing times of tasks are deterministic. In the real-world manufacturing environment, systems are often subject to random events which may disturb their working process. Therefore, considering a system in a stochastic context is more realistic than in a deterministic one [14].

We will therefore design and develop a prototype cloud-based real-time simulation scheduling system for a sensorised job shop with stochastic processing times. The next section will provide an overview of the proposed system.

3 PROPOSED SYSTEM OVERVIEW

The detail of the system design is described in [15] and [16]. **Figure 1** illustrates an overview of the proposed system as well as how each component in the system will interact with each other. The overview illustrates that there is a shop floor where different machines are installed. At each machine, there is a sensor which can be used by operators to log operation status changes. The sensors can log one of three possible states (*i.e. Waiting*, *Processing* or *Completed*). There is also a RFID card linked to each job present in the shop, which can be swiped by the machine operator when a status change needs to occur. The state is then transmitted to a gateway which in turn transmits the state to the cloud-based information system. The simulation scheduler (*i.e. Tecnomatix Plant Simulation*), also located in the cloud, consumes data from the information system and generates new schedules. There is also a web interface located in the cloud, which can also use and log data on the information system. Finally, there is a mobile device that can access both the scheduler and the web interface.

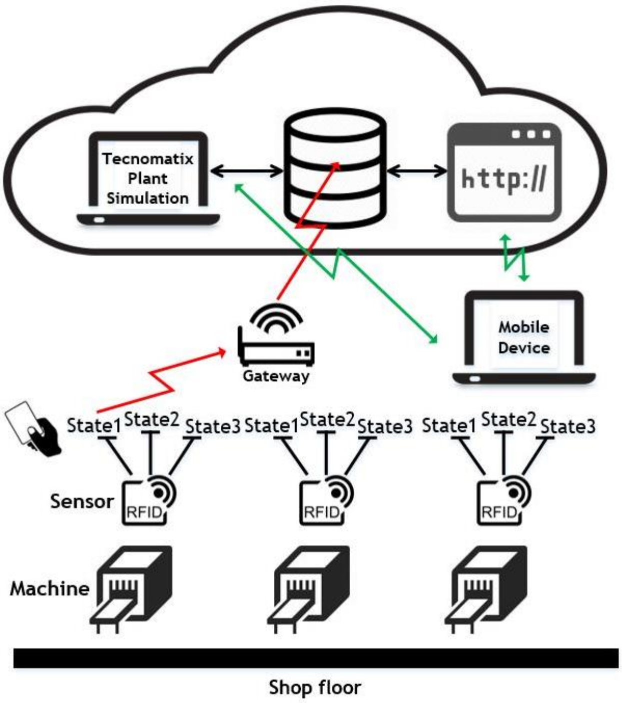


Figure 1 - Proposed system overview

The next section will describe the development of the proposed system.

4 SYSTEM DEVELOPMENT

This section will describe the development of the proposed system. The description will incorporate the sensors, as well as the development of the information system and scheduler.

4.1 Sensor integration

According to [13], RFID is a key technology of IoT, and has been widely studied to capture real-time data in a manufacturing environment. Liu & Miao [17], describe how RFID technology can be implemented to solve problems in the discrete manufacturing industry. Lu and Sheng [12] also used RFID sensors as part of their proposed management system to acquire multi-source heterogeneous information.

We therefore decided to integrate RFID sensors in the proposed system, that will be used to track the movement of job through the shop in real-time. The sensors will also communicate the real-time data through a gateway, to the information system in the cloud.

4.2 Information system and web interface design

The information system provides a database where all data from the job shop is recorded and stored. It is necessary to track the status of the shop floor, as it is required by the digital twin (the simulation model) when rescheduling of existing jobs is required. The information system, therefore, documents information regarding machines, operators, jobs and operations in the shop which can either be used by the scheduler or displayed on the web interface.

Due to the system requirement which stipulates that the information system must be cloud-based, a cloud server was created. *Google Cloud Platform* (GCP) was chosen as a cloud-computing service provider, as it provided sufficient functionality required by the authors. *MySQL* is used by GCP as the information system development tool, which is why a MySQL server was created.

The web interface is what the user of the system will use to either manually log data changes, view scheduler results or choose new schedules. The web interface was created through *Microsoft Visual Studio* and the web pages that were created for the interface include:

- logging a broken machine,
- logging a new job that entered the shop,
- changing an operator status,
- displaying scheduler results,
- manual capturing of sensor data,
- addition or withdrawal of a machine, and
- selecting a new schedule.

4.3 Simulation scheduler

The simulation scheduler is used to generate new schedules according to the dispatching rules provided earlier. New schedules are created each time the shop floor is disturbed by events such as a machine failure, a new order arriving, an operator being absent, *etc*. The scheduler also has the requirement to be cloud-based, which is why a cloud server was created to run the scheduler. Thereafter, the model was created with the use of *Tecnomatix Plant Simulation* which is a discrete event simulation software package. The elements that were incorporated into the scheduler, include:

1. MySQL data import, which is used to import data from the cloud-based information system. This will ensure that the scheduler uses the current state of the shop.
2. Dispatching rules, which are used by the scheduler to generate machine schedules. The simulation scheduler did not only use dispatching rules, but metaheuristics were also incorporated. These metaheuristics include a 2-opt move [18], simulated annealing [19] and hybrid simulated annealing algorithm.
3. Machine schedules, which are the data tables the machines refer to for selecting the next operation to start processing.
4. Machine behaviour, which ensures that each machine adheres to the sequences in the generated schedules.
5. Experiment inputs, which informs the scheduler as to which dispatching rule must be used for the current experiment.
6. Experiment outputs, which are the results of the performance indicators for a specific dispatching rule. These values will be compared to the corresponding values of the other dispatching rules, after which the best performing dispatching rule can then be identified for each performance indicator.

The metaheuristics that were selected for implementation in the simulation scheduler, are all local search metaheuristics. The reason why these were chosen, was to avoid the duplication or omission of positions in the encoded solution caused by crossover operations found in metaheuristics such as the genetic algorithm [20].

The different software packages that were used in the development of the proposed system, *e.g. Tecnomatix Plant Simulation*, can be substituted with other software packages, such as *Arena* or *Simio*. The purpose of the study was not to perform a technology analysis, but rather to illustrate that, among other things, a real-time scheduling system can be developed using typical modern simulation software packages.

5 SYSTEM TESTING

For the system testing, the authors conducted three experiments, which can be defined as:

- Test scenario 1: 50 jobs are entered into the system, each with different expected processing times.
- Test scenario 2: 100 jobs are entered into the system, each with different expected processing times.
- Test scenario 3: 200 jobs are entered into the system, each with different expected processing times.

The expected processing times in each test scenario can vary from short to long because they are sampled from triangular distributions. These distributions are used because they mimic processing time estimations by planners. The scheduling problem was solved for a single-objective instance, where the makespan of the schedule must be minimised. The experiment descriptions are provided in **Table 3**. The output of the three test scenarios are provided in **Table 4**, **Table 5** and **Table 6**, in "days:hours:minutes:seconds".

Experiment	Description
1	Shortest processing time
2	First-come-first-serve
3	Most-important-job-first
4	Earliest due date
5	Critical ratio
6	Minimum slack time
7	2-opt move
8	Traditional SA
9	Hybrid SA

Table 3 - Experiment descriptions

Experiment number	Makespan	Standard deviation	Minimum	Maximum
1	25:15:32:55	1:10:06:46	21:19:08:12	28:19:26:28
2	28:12:30:45	22:44:33	26:16:09:44	31:16:43:44
3	28:11:14:58	17:50:35	26:20:13:05	30:20:47:12
4	27:22:49:36	1:00:49:44	26:02:34:00	30:03:55:04
5	27:09:37:36	1:18:46:12	24:01:58:15	33:02:24:36
6	32:13:13:36	1:06:18:51	29:02:50:01	35:03:47:50
7	11:13:01:36	21:50:59	10:21:11:39	15:16:18:56
8	14:10:24:22	17:39:08	12:16:21:21	16:16:41:04
9	13:10:35:27	1:01:03:25	11:16:12:57	17:17:55:25

Table 4 – Output of Test scenario 1

Experiment number	Makespan	Standard deviation	Minimum	Maximum
1	54:10:59:19	3:11:05:07	46:20:21:22	61:20:06:46
2	52:18:41:17	1:09:38:21	48:17:32:17	55:18:18:01
3	51:06:38:52	1:09:23:55	47:16:50:27	55:22:05:46
4	55:13:06:56	1:21:24:14	51:16:50:00	61:17:34:31
5	54:14:24:05	2:07:59:01	49:22:06:24	60:21:47:42
6	58:09:04:28	1:13:52:24	55:17:03:36	61:17:40:55
7	19:05:33:11	21:19:05	18:09:16:23	22:08:49:47
8	22:24:22:54	1:05:03:03	20:12:06:37	25:13:18:24
9	21:07:15:28	1:18:19:18	17:18:44:09	25:08:52:19

Table 5 - Output of Test scenario 2

Experiment number	Makespan	Standard deviation	Minimum	Maximum
1	107:15:37:42	4:17:11:00	97:22:14:55	119:21:41:41
2	107:11:52:27	2:00:19:48	101:11:59:45	112:23:11:05
3	108:07:43:37	1:22:03:46	101:12:16:20	112:12:11:35
4	111:22:02:34	3:00:36:44	104:20:00:31	119:20:36:53
5	107:17:32:56	4:07:22:45	97:21:41:09	122:22:38:01
6	104:19:56:54	1:17:33:41	99:21:41:41	108:23:57:28
7	34:16:46:25	1:07:34:02	33:13:03:34	37:13:05:49
8	38:16:40:41	1:00:43:07	36:18:25:06	42:12:24:51
9	40:03:08:35	2:02:44:20	36:19:13:55	46:15:00:42

Table 6 - Output of Test scenario 3

From the output of each test scenario, as shown in **Table 4**, **Table 5** and **Table 6**, we can observe that the metaheuristics (experiment 7 – 9) outperformed the common dispatching rules (experiment 1 – 6) in all the test scenarios. In each test scenario, the 2-opt move algorithm achieved the shortest makespan. Due to the metaheuristics outperforming the dispatching rules, we will only consider them in the follow-up analysis. The authors performed 10 pseudo-independent replications of the test scenarios, after which a paired *t*-test was done on all three combinations (*i.e.* 2-opt move *vs* traditional SA, 2-opt move *vs* hybrid SA, traditional SA *vs* hybrid SA). The results of the paired *t*-tests are provided in **Table 7**.

Test	2-opt vs Traditional SA	2-opt vs Hybrid SA	Traditional SA vs Hybrid SA
1	$-2.6 \leq \mu_d \leq -1.6$	$-3.0 \leq \mu_d \leq -1.8$	$-0.8 \leq \mu_d \leq 0.2$
2	$-3.7 \leq \mu_d \leq -2.2$	$-3.2 \leq \mu_d \leq -2.0$	$-0.2 \leq \mu_d \leq 0.8$
3	$-5.2 \leq \mu_d \leq -2.5$	$-6.8 \leq \mu_d \leq -3.5$	$-2.1 \leq \mu_d \leq -0.5$

Table 7 - Paired t-test results

The paired *t*-tests confirmed that the 2-opt move algorithm performed best, while the two SA algorithms performed similarly. Although the 2-opt move algorithm performed best, it took in excess of two hours to finish, while both SA algorithms took less than 10 minutes to complete. Therefore, the 2-opt move algorithm should be used if the user of the system has sufficient time; otherwise, one of the SA algorithms should be used to generate a new schedule.

6 CONCLUSION

This paper discussed aspects of a cloud-based simulation scheduler for job shops. The functionality and testing of the system was discussed, and the results of the testing illustrated that metaheuristics outperform the common dispatching rules. This paper showed that the authors successfully developed an operational real-time simulation scheduling system; therefore, the authors made a contribution to the research of scheduling in job shop manufacturing environments, as well as industrial engineering as a whole. The next logical step is to implement the system in a real-world environment, which will be addressed in the future work of the study.

7 REFERENCES

[1] Groover, M.P., 2013, Principles of modern manufacturing, 5th Edition, US: John Wiley and Sons.

[2] Mitsuishi, M., Ueda, K. & Kimura, F., 2008, Manufacturing systems and technologies for the new frontier: The 41st CIRP Conference on Manufacturing Systems May 26-28, 2008, Tokyo, Japan. London: Springer.

[3] Leusin, M.E., Frazzon, E.M., Maldonado, M.U., Kück, M. & Freitag, M., 2018, Solving the Job-Shop Scheduling Problem in the Industry 4.0 Era, Technologies, 6(4).

[4] Ghoniem, A.S., 2002, Static and dynamic job-shop scheduling using rolling horizon approaches and the Shifting Bottleneck Procedure, Virginia Polytechnic Institute and State University.

[5] Rodammer, F.A. & White, K.P., 1988, A recent survey of production scheduling, IEEE Transactions on Systems, Man, and Cybernetics, 18(6), pp. 841-851.

[6] Kuhpfahl, J., 2015, Job Shop Scheduling with Consideration of Due Dates: Potentials of Local Search Based Solution Techniques. Springer.

[7] Sadeh, N., 1991. Look-ahead techniques for micro-opportunistic job shop scheduling, Carnegie Mellon University, Pennsylvania.

[8] Wisner, J. D., 2016, Operations Management: A Supply Chain Process Approach. SAGE Publications.

[9] Dominic, P. D. D., Kaliyamoorthy, S. & Kumar, M. S., 2004, Efficient dispatching rules for dynamic job shop scheduling, International Journal of Advanced Manufacturing Technology, 24(1), pp. 70–75.

[10] Croes, G. A., 1958, A Method for Solving Traveling-Salesman Problems, Operations Research, 6, pp. 791-812.

[11] Liu, T., Chen, Y. and Chou. J., 2015, Solving Distributed and Flexible Job-Shop Scheduling Problems for a Real-World Fastener Manufacturer. IEEE Access, 2, 1598-1606.

[12] Lu, Q. & Sheng, B., 2015, Design and Implementation of Monitoring and Management System for Discrete Manufacturing Process Based on IOT Technology, International Conference on Civil, Materials and Environmental Sciences (CMES 2015), Atlantis Press.

[13] Zhang, Y. & Tao, F., 2016, Optimization of Manufacturing Systems Using the Internet of Things, Academic Press.

[14] Lei, D.M & Xiong, H.J., 2008, Job shop scheduling with stochastic processing time through genetic algorithm. International Conference on Machine Learning and Cybernetics, IEEE.

[15] Snyman, S. & Bekker, J., 2017, Real-time scheduling in a sensorised factory using cloud-based simulation with mobile device access, The South African Journal of Industrial Engineering.

[16] Snyman, S., Bekker, J. & Botha, J.P., 2018, A real-time scheduling system for a sensorised job shop using cloud-based simulation with mobile device access, Proceedings of the 29th SAIIE Annual Conference.

[17] Liu, F. & Miao, Z., 2006, The Application of RFID Technology in Production Control in the Discrete Manufacturing Industry, Proceedings of the IEEE International Conference on Video and Signal Based Surveillance, IEEE.

[18] Croes, G., 1958, A Method for Solving Traveling-Salesman Problems, Operations Research, 6(6), pp 791-812.

[19] Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P., 1983, Optimization by Simulated Annealing, Science, 220(4598), pp. 671-680.

[20] Whitley, D., Starkweather, T. & Fuquay, D., 1989, Scheduling Problems and Traveling Salesmen: The Genetic Edge Recombination Operator. In Proceedings of the 3rd International Conference on Genetic Algorithms, pp. 133-140, Morgan Kaufmann Pubishers.

8 BIOGRAPHY

James Bekker obtained his PhD degree in Industrial Engineering from the University of Stellenbosch. He is currently the head of the Unit for System Modelling and Analysis group in the Industrial Engineering department.

Stephan Snyman holds a degree in Industrial Engineering from the University of Stellenbosch. He is currently pursuing his MEng in Industrial Engineering at Stellenbosch University. His interests in the field of Industrial Engineering include information systems and simulation modelling.

Technology-Based Data Collection in Manual Assembly using an Assembly Recorder

G. Schuh[1], J.-P. Prote[1], M. Molitor[1], N. Kettler[1]

[1]Laboratory for Machine Tools and Production Engineering (WZL) RWTH Aachen University, Aachen, Germany

Abstract

In order to ensure competitiveness in the long term, products must be produced with the help of high-quality digital infrastructures and a wide variety of variants with smaller batch sizes. The Aachen approach of the Internet of Production plays an important role in the optimization of competitive factors. The approach pursues the goal of providing information in relative real time at any place at any time. As part of this strategy, the Digital Shadow is a key element in achieving high process quality. The Digital Shadow is a virtual, yet aggregated image of the real production and requires comprehensive but especially complete data. In many companies, this database is often incomplete and of insufficient quality or requires high manual effort to generate within the production. Hence there is a need to collect data automatically to ensure the necessary data quality. In order to meet these challenges and to reduce administrative efforts, automated data recording on the shop floor is required to ensure high data quality. This is where the method of automatic identification, Auto-ID for short, comes into play. Auto-ID collects production-specific data automatically in the required quality and forms the data basis for the digital shadow from a production process point of view in order to be able to reproduce production and assembly processes digitally. This paper presents a new approach based on an “assembly recorder” to record the activities of the employees on the shop floor with little effort and to make these available for the assembly planning and later series production.

Keywords

Manual Assembly, Internet of Production, Digital Shadow

1 INTRODUCTION

Nowadays, manufacturing companies are facing a volatile and dynamic market situation characterized by shorter product life cycles and highly customized products [1]. As a result, the number of variants increases, which leads to an increased complexity of products. Furthermore, the amount of related development and manufacturing activities rises as well [2] [3].

To counter these challenges, highly iterative development processes offer a possible solution. This approach aims at the agile development of physical products. The entire development process is divided into several shorter development cycles, which are represented by a deterministic-normative or an empirical-adaptive phase [4] [5]. As a result, a high number of solutions represented by testable prototypes is suggested after each development cycle [6].

The high number of prototypes within a short time after each development cycle increases the workload for the assembly planning for each prototype. Furthermore, the continuous integration of the customer feedback leads to a high amount of adaptations to be regarded in assembly planning [7]. In conclusion, the assembly planning process needs to be highly iterative as well. One opportunity to reduce assembly planning activities is the technology-based data recording of employee activities during the assembly process on the shop floor. Based on the recorded data an assembly plan can be created and used for further development cycles.

The Aachen approach of the Internet of Production with the key element of the Digital Shadow is capable to meet these challenges. The Internet of Production describes an infrastructure for data refinement, aggregation and visualization across the entire product life cycle. The acquisition, processing and utilization of data is provided from the development through the production until the user cycle. For these tasks the infrastructure of the Internet of Production is divided into four layers. The lowest layer deals with the acquisition of raw data within the present application software, e.g. ERP systems. Above this the so-called Middleware+ is located, which takes the data and prepares it for further processing. Within the third layer data analysis takes place and the Digital Shadow is located. The subsequent aggregation and enrichment of the data with contextual knowledge generates smart data. The Digital Shadow is defined as a precise, but aggregated image of all production, development and corresponding processes [8]. The highest layer includes contextual applications for decision making [9].

1.1 Challenges of manual assembly

The high number of variants and fast development of new products result in a high amount of assembly tasks to plan in a shorter time due to the highly iterative product development [10]. In conclusion, the challenge is to provide the necessary information for the assembly planning to manage this effort in terms of time and quality [11]. Especially in a production with small lot sizes, the assembly process is not specified, nor are assembly instructions available [12]. In addition, the existing employee knowledge on the shop floor is mostly unexploited due to insufficient technology-based support in the planning of manual assembly activities. Moreover, the potential of available digital planning knowledge is just as little exploited as the possibilities of digital data collection [13].

1.2 Derivation of requirements

To ensure the availability of the latest data, it is required to collect data in real-time. In order to make this possible a high degree of automation is necessary. In addition, automated data collection reduces the proportion of non-value-added activities. Furthermore, the collected data must be available throughout the entire enterprise so that the data is available everywhere. In this regard, there are already approaches that focus on central data storage to prevent the usage of history data [10]. For technology-based support of the assembly planning the used components and resources need to be identified on the shop floor to take advantage of the employees' knowledge. To guarantee the correct chronological order timestamps needs to be recorded. In this regard, the amount of available technologies for data collection is sophisticated, so it is necessary to objectively select the technologies. Comprehensive evaluation criteria are necessary for the selection, which enable a technology selection according to the given boundary conditions. To collect the necessary data and make better use of it, the application of Auto-ID technologies represents a possible solution to automatically and digitally collect the data in the manual assembly process in order to create a planning basis [14] [15].

1.3 Research aim and approach

This paper presents an approach which provides the technology-based data collection within the manual assembly using an assembly recorder which enables data acquisition for the Digital Shadow too. This leads to the following research questions:

1. Which technologies enable the data acquisition in manual assembly?
2. How does a concept for data acquisition in manual assembly look like?
3. How can the concept of an assembly recorder be visualized?

2 STATE OF THE ART

The following chapter presents current approaches to the technology-based analysis of manual assembly processes with the intent to record data for further validation or processing.

To challenge the high number of adaptions between the engineering bill of materials (E-Bom) and the manufacturing bill of materials (M-Bom) within the highly iterative development process SCHUH ET AL. present an approach for a generative, technology-based development of a M-Bom out of the E-Bom. A concept is presented which continuously records the assembly process with the support of Auto-ID technologies like barcode systems or RFID. Parts to be mounted are identified automatically. Since an identification down to the c-part level is not economical, a main-building-component is defined in a rough planning. All other components are virtually linked to the main-building-component to form a main-building-group. Through the data acquisition, information such as the identification number of the component, the assembly time or assembly station can be recorded. Due to the automatic recording of the components, a manual update of the M-Bom is no longer necessary [16].

KÄRCHER ET AL. are focused on the technology-based analysis of manual assembly activities with the help of different types of sensors given the fact that current methods for analyzing manual assembly activities are associated with high efforts. In summary, the deficit of current systems is that many methods analyzing the manual assembly process require an attachment to the employee's body, which is problematic in terms of freedom of movement and privacy. Furthermore, the methods are associated with high financial and temporal effort and have been examined under laboratory conditions only. The approach presented by KÄRCHER ET AL. uses different types of sensors to analyze manual assembly and to manage the mentioned deficits. For example, RFID tags are used to locate tools and components. After the installation of the sensors on the objects to be identified, it will be possible to record local position, time and other process data [12].

A concept for video-based externalization of knowledge is presented by SCHUH ET AL. The so-called authoring system is intended to support the learning process of the employee. This is done with an app that supports the independent structured recording of video-based instructions. The use of video in comparison to conventional documentation increases the quality of the recorded information. With the help of this authoring system, it is possible to build a knowledge base from which other employees will benefit and that helps to improve the process. Furthermore, the workplace-specific provision of tutorials allows to expand the range of application of employees so that they can be used flexibly [17]. Furthermore, the video-based recording of the activities on the shop floor enables derivation of the assembly planning.

WINKES and AURICH presented an approach with the application of virtual reality (VR), which provides a concept to avoid failures within the assembly planning and to evaluate assembly steps before the physical implementation. Current VR concepts focus on the preparation, illustration and ergonomics analysis of assembly processes. The core of this concept is the VR workshop. On the one hand hard- and software suitable for VR are necessary and on the other hand team members from the planning and the manufacturing department are required for the workshop. During the workshop every assembly step is reviewed and evaluated by the team members with the help of VR. For the evaluation criteria such as distance or obstacles are considered. After the evaluation and adjustments if necessary, data and results are reloaded to the database [18].

None of the presented approaches fully complies with the requirements presented in chapter 1.2 of a data-based recording in manual assembly for generation of the digital shadow to automate the assembly planning.

3 CONCEPTION OF A MODEL

The following chapter presents the model of an assembly recorder. In this regard, the suitable framework for Auto-ID technologies is presented. In the next step the concept for data collection in manual assembly will be presented. In the final step, the concept of the assembly recorder is presented, which enables a visualization of the technologically-based activities of the employees on the shop floor, in order to automatically derive the assembly plan from it.

3.1 Framework for Auto-ID technologies

Auto-ID technologies are defined as automatic identification procedures to provide information about people, goods or products [19]. There are frameworks for the classification of Auto-ID technologies, which differentiate according to optical as well as electronic identification [20] [21]. As an extension, HELMUS ET AL. present a framework that differentiates between biometric systems, script or symbol-based procedures and electronic procedures. Biometric systems continue to differ in acoustic and optical methods. The optical character recognition (OCR), and the barcode systems are assigned to script or symbol-based methods. Finally, magnetic card and RFID systems are to be assigned to the electronic procedures [22].

In order to select the best fitting technology, different criteria must be considered to meet the given conditions. The first criterion is the reading distance between data carrier and reader. In addition, investment and operating costs need to be considered. Furthermore, the influence of environmental conditions like dirt or humidity is regarded [19]. The type of the identification feature also needs to be considered. In this regard, a differentiation is made between natural or artificial features. As only artificial features allow the identification of components and other objects in manual assembly, the identification of natural features is not necessary. Biometric systems are only able to identify natural features, in conclusion they do not fit data collection in manual assembly [19] [23].

Regarding the first criterion, the reading distance by OCR and magnetic card systems stand out as they need direct contact between reader and data carrier, while the barcode offers a usable reading distance of 50 cm. RFID is even capable to overcome a distance up to 5 m. In terms of the environmental influence the OCR or barcode systems are fragile and can fail completely. On the other hand, RFID is quite uninfluenceable by dirt, but the application can be compromised by electromagnetic waves or metallic objects. The costs caused by OCR or barcode systems are low by comparison to the medium high costs caused by RFID [19]. However, RFID is the only technology capable of capturing multiple data carriers simultaneously [23].

As an overall result, biometric systems are not suitable for data capturing in the manual assembly as they are only capable of identifying natural features. Furthermore, card systems and OCR are also not suitable because of the need for a direct contact between reader and data carrier which increases the effort for the identification of objects to an unacceptable level. Besides advantages and disadvantages barcode and RFID are the most suitable technologies for data collection in the manual assembly.

3.2 Concept for data acquisition in manual assembly

The essential information is based on the reference planning process for assembly planning by EVERSHEIM. As shown in figure 1, this planning process differentiates between assembly, test and cost planning [24].

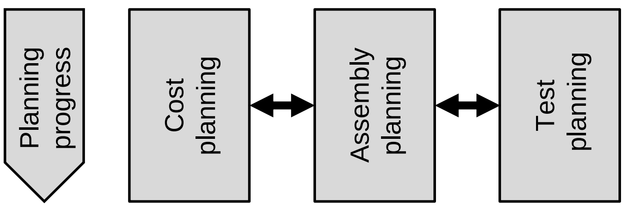


Figure 1 - Planning process by EVERSHEIM [24]

From this planning process the necessary information has been derived into data requirements which has to be identified in the manual assembly process. So-called Auto-ID events are introduced as the basis for automated data collection. In this regard, several approaches exist that use similar principles for automatic data collection in the supply chain, but with a more logistic oriented focus [25]. Auto-ID events are defined as technical or organizational activities in the assembly process, which automate data collection that has previously been collected manually, time-delayed or not been collected at all. This concept distinguishes different types. Based on the planning process by EVERSHEIM and in order to ensure a complete digital image for the assembly recorder, a distinction is made between assembly, workstation, tool and quality events. What these events have in common is that they are arranged at a certain point in the value stream and aim to identify certain components, elements or objects. In addition, different types of Auto-ID events are necessary to create a complete digital image of all activities taking place in a manual assembly process.

3.3 Visualization of the assembly recorder

The concept of the assembly recorder enables the recording of the employee's activities on the shop floor. The visualization of the assembly recorder takes place in three dimensions: assembly plan, quality inspection and costing. For preparation purposes, the assembly recorder needs to be parameterized to record the assembly task. The required bill of material, available test equipment and other related information are stored within the master data.

The assembly recorder tracks the actual-times of the set-up, assembly, testing and inspection processes and matches them to the right event. To add additional information the employee uses the designated interface which is described below. The quality check is based on the prior test planning and takes place parallel to the assembly processes and the cost controlling, which is based on the actual-time and the master cost rates.

The graphical user interface (GUI) of the assembly recorder is divided into three main areas which is shown in figure 2. The first area displays the assembly and workplan which shows data such as the part-ID or the workstation. The real-time tracker located in the bottom left corner displays the time needed for the tasks performed and is utilized as a tool for comprehension and controlling the employee's inputs. The cost-controlling area is located on the bottom-right, where the current costs are displayed.

The assembly recorder allows the digital visualization flow of information and enables the documentation of the assembly activities. As a result, a data base is created for the assembly planning of prototypes or future products to be produced in serial production.

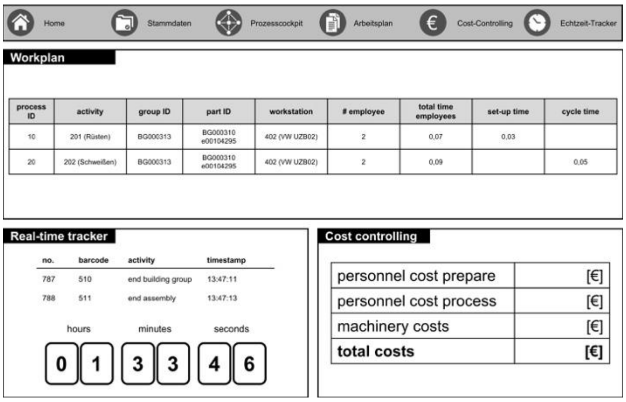


Figure 2 - Basic GUI of the assembly recorder

4 FIRST CASE STUDY AND BASIC VALIDATION

In the environment of the Aachen Demonstration Factory, a first case study has been carried out to investigate the performance of the assembly recorder under real conditions. The Aachen Demonstration Factory offers the possibility to examine such use cases in a real production environment. To validate the assembly recorder, the production of an electric-driven kart was selected. The production of this kart is characterized as a single and small batch production. Due to a simple and inexpensive implementation a barcode system has been chosen for the validation. The used parts have been equipped with barcodes printed on labels. A wireless barcode scanner was used to capture and transmit the barcodes.

Based on the technology-based recording of the assembly activities on the shop floor, the assembly plan was derived automatically. The effort for quality control was reduced by recording the quality results during the assembly. Moreover, the assembly costs became transparent through the simultaneous cost-controlling. In total, the measured time saving was approximately 29 hours compared to assembly and data collection without the presented assembly recorder.

The concept of technology-based data acquisition in manual assembly with an assembly recorder demonstrated that during the assembly of the prototype all physical assembly activities of the employees could be recorded on the shop floor and a complete assembly plan could be derived. This enables a parallel design of the series process.

5 CONCLUSION

Shorter product life cycles and highly customized products lead to a high number of new products in a short time. Highly iterative product development is designed to handle these challenges. The increasing number of prototypes in a shorter time leads to a disproportionate increase in the effort in assembly planning. As a consequence, the assembly needs to be planned in a shorter time as well. In order to develop a technology-based approach for manual assembly, existing approaches were identified taking into account derived requirements. Since none of the approaches fully met the derived requirements, the concept of technology-based data acquisition in manual assembly based on the assembly recorder was introduced. First of all, suitable Auto-ID technologies have been selected. Subsequently, a concept for the automated data acquisition has been derived. Finally, the GUI of the assembly recorder was designed. A first validation was carried out in the Aachen demonstration factory. As an overall result, the assembly recorder enables the automatic generation of the entire assembly planning.

6 ACKNOWLEDGMENT

The authors would like to thank the German Research Foundation DFG for its kind support within the Cluster of Excellence “Integrative Production Technology for High-Wage Countries”.

7 REFERENCES

[1] Sherman, L.: If You´re in a Dogfight, Become a Cat! Strategies for Long-Term Growth. New York: Columbia University Press, 2017, p. 157

[2] Festge, R.; Malorny, C.: The future of German mechanical engineering. Operating successfully in a dynamic environment, 2014, p. 61

[3] van Iwaarden, J.; van der Wiele, T.: The effects of increasing product variety and shortening product life cycles on the use of quality management systems. In: International Journal of Quality & Reliability Management, Vol. 29, Issue: 5, 2012, pp. 470-500

[4] Diels, F.; Rudolf, S.; Schuh, G.: Highly iterative product development process for engineering projects. In: Applied Mechanics and Materials, Vol. 794, 2015, pp. 532-539

[5] Schuh, G.; Riesener, M.; Diels, F.: Structuring Highly Iterative Product Development Projects by Using HIP-Indicators. In: IEEM Industrial Engineering and Engineering Management, 2016, pp. 1171-1175

[6] Diels, F.; Riesener, M.; Schuh, G.: Methodology for the Suitability Validation of a Highly Iterative Product Development Approach for Individual Segments of an Overall Development Task. In: Advanced Material Research, Vol. 1140, 2016, pp. 513-520

[7] Schuh, G.; Prote, J.-P.; Dany, S.; Molitor, M.; Giner, B.: Process model for generative assembly planning in the highly iterative product development. In: Procedia CIRP, Vol. 72, 2018, pp. 363-368

[8] Schuh, G.; Blum, M.: Design of a Data Structure for the Order Processing as a Basis for Data Analytics Methods. In: Portland International Conference on Management of Engineering and Technology (PICMET), 2016, pp. 2164-2169

[9] Schuh, G.; Stich, V.; Basse, F.; Franzkoch, B.; Harzenetter, F.; Luckert, M.; Prote, J.-P.; Reschke, J.; Schmitz, S.; Tücks, G.; Weißkopf, J.: Change Request im Produktionsbetrieb. In: Tagungsband zum AWK Aachener Werkzeugmaschinen Kolloquium. Aachen, 18.-19. Mai. Aachen: Apprimus Verlag, 2017, p. 109-131

[10] Müller, R.; Hörauf, L.; Vette, M.; Speicher, C.: Planning and developing cyber-physical assembly systems by connecting virtual and real worlds. In: Procedia CRIP, Vol. 52, 2016, 35-40

[11] Feldmann, K.; Schmuck, T.; Brossog, M.; Dreyer, J.: Beschreibungsmodell zur Planung von Produktionssystemen. In: wt Werkstatttechnik online, Vol. 98, 2008, No. 3, p. 156-162

[12] Kärcher, S.; Cuk, E.; Denner, T.; Görzig, D.; Günther, L.; Hansemersmann, A.; Riesinger, G.; Bauernhansl, T.: Sensor-driven Analysis of Manual Assembly Systems. In: Procedia CIRP, Vol. 72, 2018, p. 1142-1147

[13] Manns, M.; Wallis, R.; Deuse, J.: Automatic proposal of assembly work plans with a controlled natural language. In: Procedia CIRP, Vol. 33, 2015, p. 345-350

[14] Li, Z.; Hong, J.; Xue, F.; Shen, G.: SWOT analysis and Internet of Things-enabled platform for prefabrication housing production in Hong Kong. In: Habitat International Vol. 57, 2016, p. 74-87

[15] Wang, Z.; Hu, H.; Zhou, W.: RFID Enabled Knowledge-Based Precast Construction Supply Chain. In: Computer-Aided Civil and Infrastructure Engineering, Vol. 32, Issue 6, 2017, p. 499-514

[16] Schuh, G.; Zeller, V.; Prote, J.-P.; Molitor, M.; Wenger, L.: Generative Stücklistenerstellung in der manuellen Montage. In. Zeitschrift für wirtschaftlichen Fabrikbetrieb, Vol. 112, 2017, Issue: 6, p. 392-395

[17] Schuh, G.; Prote, J.-P.; Gerschner, K.; Molitor, M.; Walendzik, P.: Technologiebasierte Externalisierung von Wissen. In: wt Werkstatttechnik online, Vol. 107, 2017, No. 9, p. 578-581

[18] Winkes, P. A.; Aurich, J. C.: Method for an enhanced assembly planning process with systematic Virtual Reality inclusion. In: Procedia CIRP, Vol. 37, 2015, pp. 152-157

[19] Finkenzeller, K.: RFID handbook. 3rd ed., reprinted. Chichester, West Sussex: Willey, 2006, p. 7 and p. 108-110 and p. 199-200

[20] Bartneck, N.; Klaas, V.; Schönherr, H.: Optimizing Processes with RFID and Auto-ID. Erlangen: Publicis Publishing, 2009, p. 5

[21] DIN Deutsches Institut für Normung e.V.: DIN SPEC 16599. Informationstechnik. Berlin: Beuth Verlag, 2017, p. 15

[22] Helmus, M.; Meins-Becker, A.; Laußat, L.; Kelm, A.: RFID in der Baulogistik. 1st ed. Wiesbaden: Vieweg, 2009, p. 199

[23] Kern, C.: Anwendung von RFID-Systemen. 2nd ed. Berlin: Springer, 2007, p. 15

[24] Eversheim, W.: Organisation in der Produktionstechnik 3. Arbeitsvorbereitung. 4th, edited and corrected ed. Berlin, Heidelberg: Springer-Verlag, 2002, p. 18

[25] Hahn, M.: Durchführung einer Machbarkeitsstudie zur Implementierung von Auto-ID Technologien in der Intralogistik eines Sondermaschinenbauunternehmens. Master thesis. Technical University of Dortmund. Dortmund, 2016, p. 50-59

8 BIOGRAPHY

Since 2002 Prof. Günther Schuh holds the chair of Production Engineering and at the same time he became a member of the Managing Board of the Laboratory of Machine Tools and Production Engineering WZL at RWTH Aachen University and the Fraunhofer IPT in Aachen. Since October 2004 he is also director of the Institute for Industrial Management FIR at the RWTH Aachen University and since 2009 Managing Director of Aachen Campus GmbH. From 2008 to 2012 Prof. Günther Schuh was Vice Chancellor of the RWTH Aachen University.

Since 2014 he is a member of the board of the National Academy of Science and Engineering acatech.

He is a member of several supervisory-, administrative- and advisory boards and is Chairman of the Management Board of e.GO Mobile AG since January 2015.

Dr.-Ing. Jan-Philipp Prote, studied industrial engineering with a specialization in mechanical engineering at RWTH Aachen University and industrial engineering at Tsinghua University in Beijing. He is chief engineer at the Laboratory of Machine Tools and Production Engineering WZL at RWTH Aachen University and since 2016 has been the head of the production management department.

Marco Molitor, M. Sc., studied mechanical engineering in at the University of Applied Sciences Koblenz and industrial engineering at the University of Applied Sciences Mannheim. He has been a research assistant at the WZL machine tool laboratory at the RWTH Aachen University and has been group leader of the Process Management research group in the Production Management department since the beginning of 2018.

Nils Lehde genannt Kettler, B. Sc., studied mechanical engineering at RWTH Aachen University. Since 2016 he is a project manager at the Demonstration Factory Aachen DFA and specializes in manual assembly systems.

Lean Management Approach to Optimize Engine Overhauling Processes

Jörg Niemann[1], Timm Seeger[1], Claudia Fussenecker[1], Martin Schlösser[1]
[1]Faculty of Mechanical and Process Engineering, Düsseldorf University of Applied Sciences, Düsseldorf, Germany

Abstract
In today's world of manufacturing, it gets increasingly important to supply a product with the right quality as fast as possible for a low price. The Lean Philosophy and its methods is one way to achieve this goal, especially in a world where customer requests get more specific. To satisfy the customer's demand, short lead times and lean processes are necessary. This paper is about Value Stream Mapping which is a useful method to reduce waste and optimize your processes, which will result in lower lead times and costs. The first part is about the background of Value Stream Mapping and gives a short introduction. In the second part, a practical example of an early stage engine overhauling process by a top in class supplier in South Africa is given, including the optimization of identified potentials by eliminating waste and applying the 5S.

Keywords
Value Stream Mapping, waste, Lean Manufacturing, 5S, optimization

1 INTRODUCTION

1.1 Lean Philosophy

The phrase, Lean Manufacturing, was created by a research group led by J. Womack at the MIT in the late 1980s and early 1990s and has been popular since they published the book "The machine that changed the world" in 1990. They compared Japanese to American companies and found out that the most efficient car manufacturer was the Toyota Motor Company with the Toyota Production System. [1] In a second book, published in 1996: "Lean Thinking: Banish Waste and Create Wealth in Your Corporation", they exposed the essential principles of the Lean Philosophy. [2]

According to Womack and Jones, there are five essential principles in every lean system:

1. Specify value from the customer's point of view
2. Identify the value streams and remove waste
3. Create continuous flow in the remaining value streams
4. Allow Customers to pull the products from the value streams
5. Keep improving the process until it is perfect (delivery of pure value and zero waste) [3]

A powerful method to realize step 2: "Identify the value streams and remove waste", is Value Stream Mapping. A value stream is defined by Rother and Shook in their book "Learning to See - value stream mapping and eliminate Muda" as "all the actions (both value added, and non-value added) currently required to bring a product […] from raw materials into the arms of the customer." [4]

1.2 Value Stream Mapping

A Value Stream Map visualizes a whole process a product takes: for example, from door to door of a facility or beyond. It is used to identify waste. Everything which doesn't add any value for the customer to the product can be seen as waste and have to be minimized. Additionally, to the process flow, a value stream shows the information and material flow throughout the process, by using fixed symbols and a fixed structure. [4]

The sequence of the Value Stream Map method can be portrayed in five steps:"

1. Value stream scope: Determine the value stream to be improved
2. Current state map: Understanding how things currently operate. This is the foundation for the future state
3. Future state map: Redesigning the process according to lean principles
4. Implementation plan: Developing a plan of implementation to realize improvements (what, who, when)
5. Implementation of the improvement plan: The goal of mapping" [5]

1.3 The different types of waste

The seven types of waste or as it is called from its japan origin Muda, which are supposed to be identified and eliminated in the Lean Philosophy with the help of Value Stream Mapping, are:

- Transport – unnecessary transport of material or people between processes

- Inventory – unneeded raw materials, semi-processed goods or finished goods which are not having value added to it
- Motion - unnecessary transport of material or people within processes
- Waiting – People or parts waiting for the next process steps
- Over-processing – processing over the need of customers' requirements
- Overproduction – production which produces quicker, faster or more than the customers requires
- Defects – quality defects of the products [6]

According to other sources there are various additional types of waste. Womack says there is another type waste, which is the manufacturing of products, which does not meet the customer's demand, since value is defined by the customer. [3] Other sources name another type of waste as "Skills", "Talent" or "Ideas". It describes the inappropriate use of the employer's skills, talents and ideas. Examples are, the use of an overqualified worker on an easy task or ignoring the ideas of employers for improvement. [7]

1.4 The 5S Method

The 5S method originally comes from Japan and is used to create a clean, efficient and safe workplace. The method contains five steps which are described by these terms:

- "Seiri" – Sort
- "Seiton" – Set in order
- "Seiso" – Shine
- "Seiketsu" – Standardize
- "Shitsuke" – Sustain [8]

A workplace in which 5S is applied and taken seriously will have positive effects on many levels of the company. For example, the lead time will be reduced, because of shorter transport times and less time for searching materials or tools. Another benefit could be more, space which increases safety and provides more comfort.

But, as with every Lean Method, it will only be improving the company and its processes in a long-term point of view when taken seriously and if it´s implemented in the company´s philosophy and worker's mind-sets.

2 VALUE STREAM MAPPING: A PRACTICAL EXAMPLE

As a practical example of Value Stream Mapping and the elimination of waste, a process in one of the leading manufacturing and overhauling companies for diesel engines in South Africa was observed. Their main business is doing full overhauls of big diesel engines for ships, trains, mining vehicles and engines for power generation. The observed process is an early stage process, in which a long block is used for the overhauling process. A long block is an almost finished engine excluding various parts which are, for example, electrical components, exhaust and intake components, fuel pumps and further components. It contains the full assembled block, usually including cylinder heads, the camshaft, the crankshaft and a lot of other components. By using this already overhauled long block, the lead time of an overhauling process can be shortened, since only the missing parts have to be removed from the customer's engine. The older long block is overhauled later in a different process. The process can start once the customer's engine arrives. The engine gets stripped down to long block level and the re-usable parts are cleaned and overhauled, while others are replaced. After the assembly is done to the customer's specifications, the engine gets painted and dyno tested.

The difficulties in this process is the wide variety of the different specifications: even if all the engines use the same long block, it is necessary to deliver the right parts for fitting the specification specific components at the right time within the process.

The value for the customer in this process is mainly focused on short lead times and a good quality since an engine, which is not on the road because of overhauling or repair, costs them money every hour.

2.1 Current state map

In this stage of the observed process, parts of the planned process were not done, because the customer's engine was not re-used. With Value Stream Mapping, five optimization potentials have been identified (see Figure 1):

1. Waiting time between the unpacking and preservation
2. Waiting time between the preservation and the part picking
3. Time between start of part picking and the start of the assembly
4. The assembly processes
5. Waiting time between assembly and test bench

In the next step, there is a closer look on the reason for a long waiting time or a delay to work out a reasonable way to optimize the process. Parts of the process, where waste or an optimization potential was observed, are marked with a "kaizen-burst" and numbered.

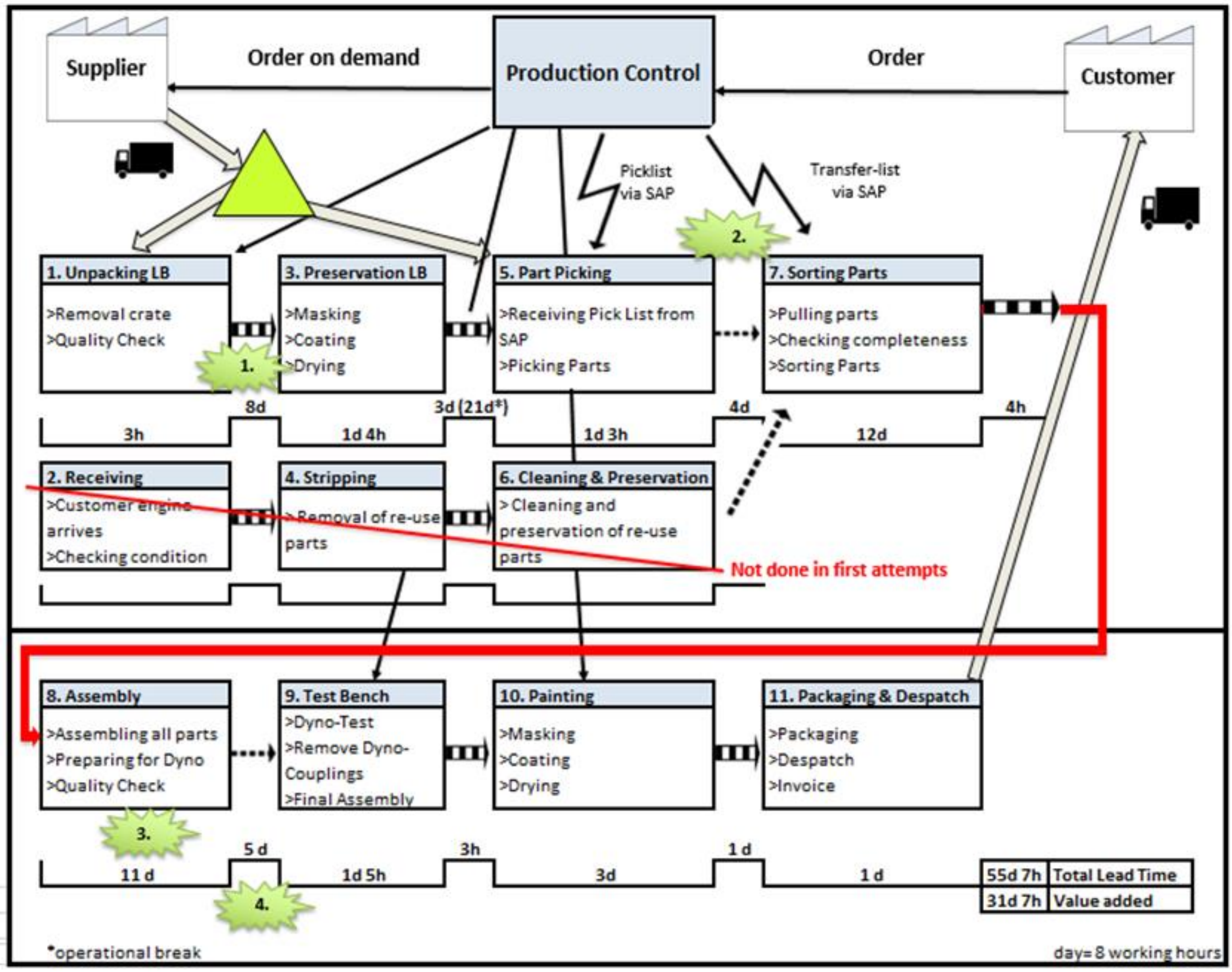


Figure 1 - Current value stream map

2.2 Elimination of waste and optimization

2.2.1 First Kaizen-Burst – Waiting time between unpacking and preservation:

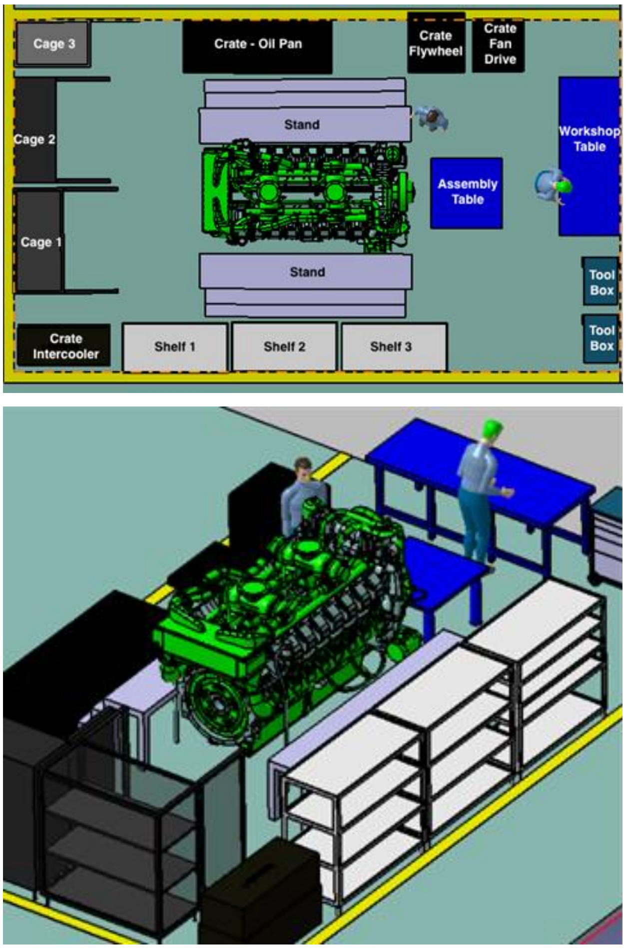
Problem: The Value Stream analysis showed that there was a long waiting time between the unpacking and the preservation of the Long Block which will be used for the overhauling process. The main reason for that was a misunderstanding of the scope of supply of the Long Block. The Production Management thought that the Long Block would be already preserved with a coating to prevent corrosion, unfortunately this was not the case. It took some time to get the right coating and to ensure the painting quality.

Solution: This problem can be solved by a safety stock of the right coating, since the long block will be delivered without coating in the future as well.

2.2.2 Second Kaizen-Burst – Long part sorting time

Problem: One main problem of the analysed process was a very long sorting time. In the usual overhauling processes, the workshop gets supplied with all the parts from the logistics office all at once. The parts get picked and are moved with a trolley into the workshop without any order. After that, the mechanics double check that all the parts from the picklist are supplied with the help of a transfer list. Afterwards, the mechanics sort all their parts by the different assembly groups in the shelfs. We found this process to be very inefficient, as the same parts are moved around at least three times before they reach their destination in the shelves and it's very exhausting to check, by hand, where one of the over 1000 different parts will be used on the engine, especially for small parts, such as screws and washers. Additionally, there were parts missing, which caused more delay without adding any value to the product.

Solution: To eliminate this waste of transport and to reorganize the process in a lean way, it makes sense to connect the picking process part in logistics with the part that sorts them into the shelves. Like that, the logistics worker puts the parts directly into the right spot in the shelves instead of a normal trolley to avoid this transport step. In the end, the shelves can be moved into the assembly bay. To shorten the process of ordering, additional information on the picklist is needed. This additional information is the assembly group which tells the logistics worker where to put the part in the shelves. This information can be easily obtained from the order and, like that, the picking order can be used for sorting instead of just ignoring



it.

Figure 2 - Current value stream map

2.2.3 Third Kaizen-Burst – The assembly

While measuring the assembly time and overserving the process, several problems have been identified:

- Too little space on the sides of the engine, if there are stands on both sides (necessary)
- Unnecessary shelf space available
- Problems to maintain the order of sorted parts in the shelves
- Crates and engines are hard to remove from bay
- Parts from crates had to be moved far around the bay with the overhead crane (wrong position of crates)
- Turning the engine was necessary
- Unsafe, due to too little space to handle big and heavy parts

To be able to find a better solution to this problem, a 3D CAD model of the current assembly bay was created (see Figure 2):

To optimize and reorganize the assembly bay, the 5S method was applied on the old layout (see Figure 3):

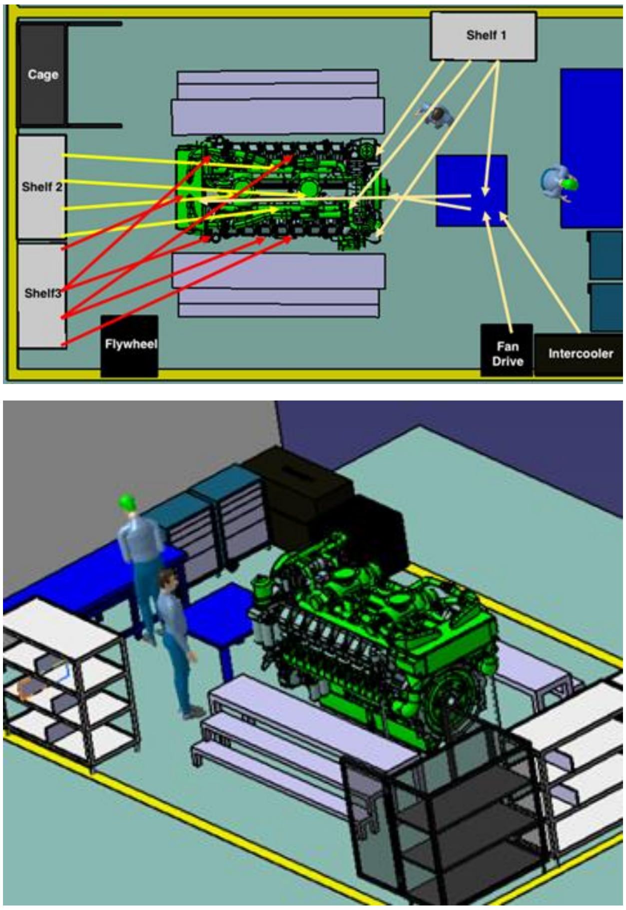


Figure 3 - Optimized assembly bay layout, created with CATIA V5

1. Sort: Every unnecessary item was removed from the bay, due to space optimization in the shelves and too much shelf space in general; as a result, “Cage 2” and “Cage 3” could be removed from the bay.
2. Set in Order: In this step, the necessary components were rearranged in a rational and safe way. To ensure easy accessibility to all the components, which must be removed from or moved into the bay, they are now positioned close to the shop floor side. Like this, all the crates can easily be moved into the bay with a forklift and the engine can be moved in and out without moving any shelves. To ensure workers’ safety, there are no components placed in front of the stands. The shelves are rearranged close to the engine and they will carry the parts which are used on this side of the engine. They will also be modified with dividers to ensure that every part has a place and that spares cannot get mixed up during the assembly process, to prevent long searching times.

3. Shine: During the observation of the assembly process, the working place was usually clean, and this point wasn't a big problem. However, the removal of packaging and crates should be done immediately, and this was not always the case.
4. Standardize: To ensure consistency of the first 3S's, all the components in the bay will be labelled as part of this bay and with its designation. Pictures of the correct bay layout will be suspended to the walls of the bay.
5. Sustain: Audits for safety, which also contain 5S aspects, are done on a regular basis.

The result is a much cleaner and safer assembly bay which is also optimized for an easy material flow.

2.2.4 Fourth Kaizen-Burst – Waiting time between assembly and the test bench

Problem: A long waiting time between the assembly and the testing of the engine on the dyno was caused by issues with the re-programming of the Engine Control Unit (ECU). Since it is not possible to run the engine without ECU, it was necessary to delay the testing until the problem was solved.

Solution: To lower the likelihood that such an event occurs again, the assembly order of the parts could be changed. It makes sense to assemble parts which are needed for the dyno testing first to ensure that all parts are working and available. The ECU should be programmed already when the customer engines arrive, and it must be done before the assembly starts. Another problem at this point of the process could be the availability of the test bench. Testing an engine takes at least one day and there are several engines tested every month. Daily communication between the mechanics, who does the assembly and the production, controls who manages the time scheduling for the test bench and is necessary to ensure that the test bench is available once the assembly is finished.

2.3 Future state map:

By comparing the Current State Map and the Future State Map, a possible reduction of the total lead time of 55 days and 7 hours to 30 days can be observed. The value-added time improves from 25 days and 5 hours to 21 days and 5 hours. A long-term goal should be an increase of the process flow by reducing the waiting times between the different steps. [9]

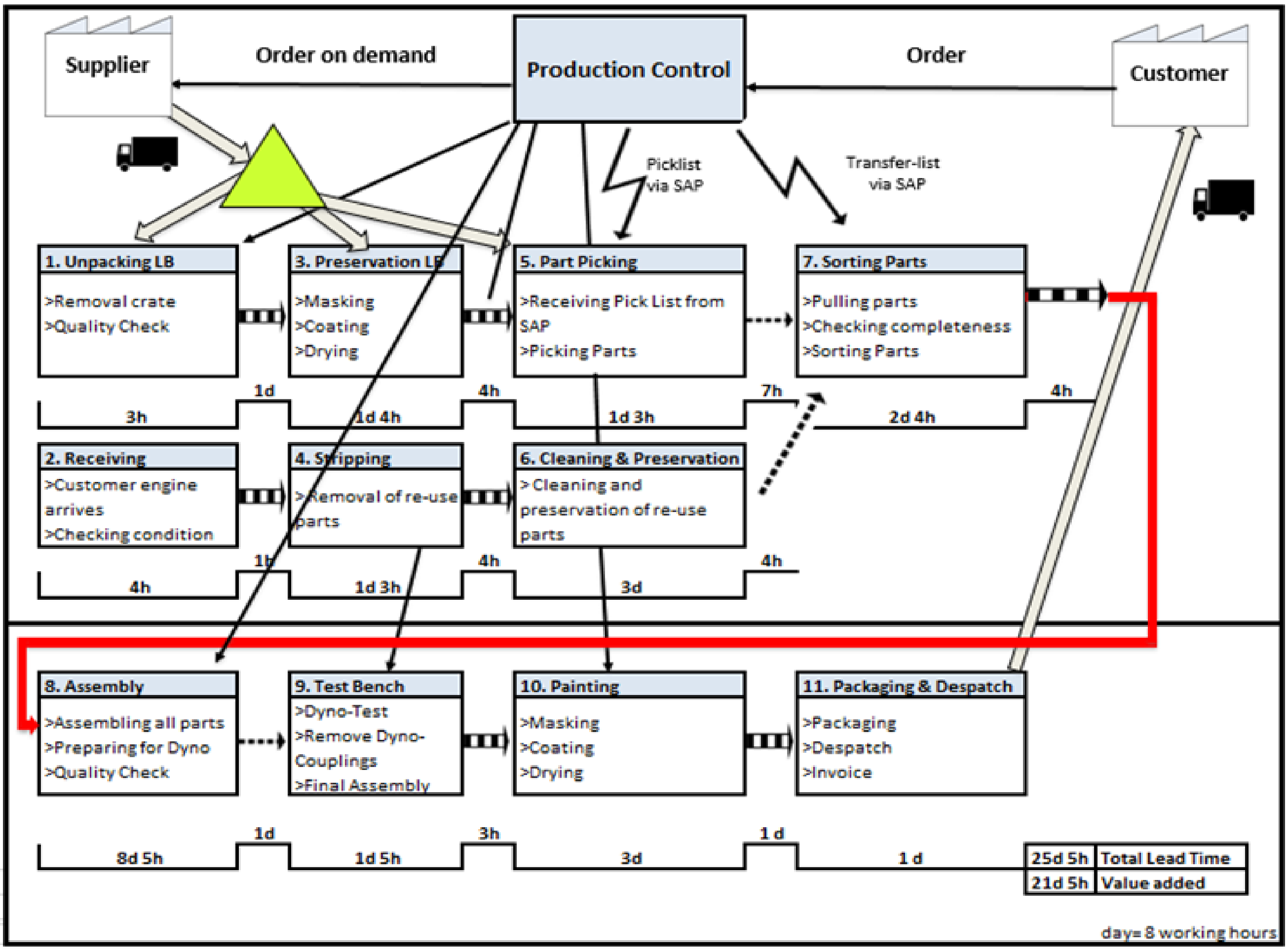


Figure 4 - Future value stream map

3 CONCLUSION

In general, Value Stream Mapping is a very powerful method to identify optimization potentials and to optimize. Its main benefit is the view on the whole process, which will let you understand where the critical steps are and will give you an idea how to improve these. [10,11]

Another advantage is the customer orientated definition of value used with Value Stream Mapping. With this perspective in mind, the improvements will not only save costs, but they will also bring you closer to fulfil what the customer wants. Due to the standardized symbols and easy layout of the map, is it possible to work on it in a team and to easily exchange results.

Problems with the method that were observed in this process, were with displaying parallel processes properly in the map. It is also very time consuming to gather the data for the current state map, but by considering the optimization potentials which can be achieved, it is worth it. Though applying value stream mapping itself will not guarantee successful improvements, it is necessary to combine it with other lean tools and methods and to integrate lean thinking into the whole company.

4 REFERENCES

[1] Womack, J.P., Jones, D.T., Roos, D., 1990, The Machine that changed the world.

[2] Dekier, L., 2012, The Origins and Evolution of Lean Management Systems,

[3] Womack, J.P., Jones, D.T., 1996, Lean Thinking: Banish Waste and Create Wealth in your Corporation

[4] Rother, M., Shook, J.; 1999, Learning to See – value stream mapping to add value and eliminate muda.

[5] Langstrand, J., 2016, An introduction to value stream mapping and analysis, Linköping University

[6] Conrad, R.W.,2016, 7 Arten der Verschwendung (7V), Springer

[7] Lean Consulting, in: https://leanconsulting.com/lean-resources/lean-templates/8-wastes-of-lean/, last acces 01.05.2018

[8] Neuhaus, R., Conrad, R.W., 2016, 5S als Basis des kontinuierlichen Verbesserungsprozesses ifaa

[9] Behle, R.; Niemann, J.: How meaningful is automation of assembly line supply. In: 2018 International Conference on Production Research – Africa, Europe, Middle East 5th International Conference on Quality and Innovation in Engineering and Management, July 25-26, 2018, Cluj-Napoca, Romania

[10] Niemann, Jörg: Die Services-Manufaktur, Industrielle Services planen –entwickeln – einführen. Ein Praxishandbuch Schritt für Schritt mit Übungen und Lösungen. Aachen, Shaker Verlag, 2016

[11] Niemann, Jörg; Tichkiewitch, Serge; Westkämper Engelbert: Design of Sustainable Product Life Cycles, Springer Verlag, Heidelberg Berlin, 2009

5 BIOGRAPHY

Jörg Niemann obtained his PhD degree in Manufacturing Engineering from the University of Stuttgart. In 2012 he was appointed Professor in Business Management and Mechanical Engineering at the Düsseldorf University of Applied Sciences, Germany.

Timm Seeger is a B.Eng. student at the University of Applied Sciences Duesseldorf. He is majoring in production optimisation and an expert in the field of production flow simulation.

Claudia Fussenecker holds a M.A. in European Management from the FH Bund in Brühl and a Diploma in Industrial Management from EUFH in Brühl. She is currently a scientific assistant in the Department of Mechanical and Process Engineering, focusing on Marketing and Services and Life Cycle Management.

Martin Schloesser holds a B.Eng. with specialization in Product Development and Production from the University of Applied Sciences Duesseldorf. He is currently a scientific assistant in the Department of Mechanical and Process Engineering, focusing on production simulation and Life Cycle Management.

Planning and Configuration of Agent Systems for Autonomous Production Control – Partially Automated Code Generation in an Early Planning Phase

L. Kiefer[1], D. Königl[1], C. Richter[1], G. Reinhart[1]
[1]Fraunhofer IGCV, Provinostraße 52, Augsburg 86153, Germany.

Abstract
The rising demand for customer innovated products causes an increasing complexity within production systems. Autonomous control systems (ACS) are recommended to deal with this complexity by distributed problem solving. The use of ACS is subject to fulfilling three fundamental requirements (identification, localisation, communication). This approach focuses the requirement communication and agents as an enabler within an alternative, non-ID-based ACS. To develop a suiting agent system, various methods have been developed. Each approach focuses on different conditions and later on tasks of the system. By defining requirements and conditions for attribute-based autonomous control systems, existing planning methods are analysed for their suitability. Based on this analysis, a method is selected and adapted to the specific development of ACS. The analysis and the resulting adapted method are presented within this paper.

Keywords
Autonomous Control, Agent Systems, Agent Planning, Configuration Agent Systems, Identification

1 INTRODUCTION

With increasing product diversity and market dynamics, companies are facing new market challenges. Shorter innovation cycles, customised products and highly fluctuating demands rise the complexity, which is no longer manageable with hierarchical architectures in production [1] [2]. One approach to master the complexity are autonomous manufacturing systems. Autonomous control requires the identification of all system elements [3, 4]. The object identification is typically performed indirect, which is based on additional data carriers and causes additional costs (e.g. RFID-tags). To avoid these carrier objects can be identified directly based on their attributes. Thereby, agent-based approaches are recommended to enable communication within this heterachic modelled system. The agents are used as a virtual representation of the production components. Agent systems, as well as the agents themselves, require a detailed planning to ensure their functionality [5]. Different methods for planning and developing agent systems have been developed, each suiting for specific tasks and environments [6]. However, the alternative autonomous control based on an direct object identification has not been focused yet. The definition of required additional functions and capabilities of agents in this scenario has to be performed. First, an introduction of autonomous control and agent planning methods is given.

2 AUTONOMOUS CONTROL SYSTEMS

In autonomous control systems, objects and machines are enabled to communicate and take decisions on their own [7]. For an autonomous control, the following criteria need to be fulfilled: identification, localisation and communication.

2.1 Identification

“Identifying means recognising, labelling or addressing an object within its scope using the required features” [8]. There exist two different types of identification procedures: indirect and direct identification [9]. Within indirect identification, objects are labelled with an ID to recognise them and read their data. Examples of indirect techniques are barcode and RFID-systems. Direct identification is based on an object’s attributes. Therefore, no additional data tag is required. Exemplary attributes are colour, shape, size and weight of the object [9]. These can be recognised by sensors of the production system and compared with attributes in a database for unique identification.

2.2 Localisation

For an autonomous control, the location of the object is one of the most important parameters for further processing [3]. One approach is the use of indoor-GPS combined with Auto-ID-techniques, in which a parallel localisation is possible [10]. Another opportunity is a constant saving of an object’s location and an update after every movement. Within the attribute based autonomous control approach, handling devices update the position and ensure a constant tracking of each object.

2.3 Communication

Communication serves to exchange information and represents a central requirement in autonomous production systems [4]. An approach to enable the communication of objects and machines in a production

system is to bind them to a virtual component that handles the interactions. For this purpose, several experts recommend the use of agents as a virtual twin [11]. Procedural communication, black-board-systems and message-based communication are several techniques for agents to communicate [12].
Within the *procedural communication*, components interact by calling another's procedure. A procedure is a particular program and has input parameters to detail a request and return parameter to respond [13].
The blackboard is a data store in *blackboard-systems*. Components access, read and modify the data store to share information, data and knowledge [13]. For agent systems, an agent writes information into the blackboard which can be read by another agent [14]. Optional coordinating components can verify the accesses to the blackboard and filter relevant information.
The concept of *message-based communication* takes place via the direct exchange of messages between two virtual components, also called sender and receiver. This type of communication is not fixed; objects can send messages to several recipients. Multicast messages have a limited number of known receivers while in broadcast messages the receivers and their number are unknown [15].

3 AGENT PLANNING METHODS

In agent-based ACSs agents represent objects and resources. The resulting multi-agent systems are an association of agents that interact and pursue specific goals. The capabilities of an agent system are beyond the individual capabilities or knowledge of the individual agents because subtasks can be performed by multiple agents at the same time [16]. The overall functionalities of the system and its performance is mainly depending on its structure, which has to be specifically planned.
The purpose of a development method is to assist the developer in his planning and development tasks throughout the software development process [16]. The basic structure of a development method consists of the analysis phase, the rough and detailed design as well as the implementation and test phase [16]. There are more than 80 different methods available for agent system planning grouped in knowledge-based and agent-oriented approaches [6]. Object-oriented as an extension of agent-oriented approaches use three data structures, referring to agents BDI-architecture [16].The current state of the agent stored in agent's beliefs (B). The agent's core desire (D) defining its behaviour and goals is stored in the targets. Intentions (I) and plans are hierarchical procedures implemented to achieve the goals of the agent. The consideration of theses structures by software developers enables a standardization, which refers to the design rules of the Foundation for Intelligent Physical Agents (FIPA).
Exemplary development methods are PASSI [17], Aspects [18], Mobmas [19] and Zeus [20]. These methods are suited for a general development of agent systems. Regarding the development of an agent system for an attribute based ACS, specific criteria have to be fulfilled. Besides the integration of specific agent tasks for the autonomous control the application of the method has to take place in an early planning stage of the production itself [21]. Therefore, the user of the method is probably not a software engineer but a factory planner. For this reason, the planning steps have to be simplified by integrating expert knowledge in different development tools. Furthermore, the development should be partially automated to reduce the complexity for the methods user. Existing development methods require a software developer and thus specific knowledge of the user.

4 PLANNING AND CONFIGURATION OF AGENT SYSTEMS

In the following sections the analysis of the existing methods and the adapted method are presented.

4.1 Development of the method

The method for planning and configuration of agent systems is systematically developed by using the Munich method model (MMM). The model provides an approach for developing a methodology for a specific use case and contains four phases [22]: Clarifying method's use, selecting suiting method, adapting method and applying method.
To clarify the *method's use* the first step is to define conditions of the method and to clarify the desired output of the procedure to be developed. In this case, the design of an agent-oriented autonomous control system for a production system based on a direct identification is the desired final output. Specific descriptions of the system are needed to derive the required system requirements. Furthermore, the product corridor and the type of products that are manufactured have to be known [21]. This knowledge includes production processes and product-specific quantities. This information is required for the definition of roles and agents.
Usually, agent systems are created by software developers [5]. In order to take advantage of an early planning phase, the method is more likely to be used by factory planners, at least initially. For this reason, a great deal of expert knowledge has to be integrated into the method.
To select a suiting method regarding the described use-case, evaluation criteria first have to be defined. Categorised evaluation criteria enable a structured evaluation. Exemplary rating categories are general, domain-related, design-related, system-related, flexibility-related and process-related criteria [16, 23]. In addition, the analysis of supporting tools is highly relevant for the selection.

From the perspective of autonomous control, not all evaluation criteria are relevant. In the following, a distinction is made and evaluated between general and autonomous control specific criteria. In agreement with experts and using a pairwise comparison, the criteria were weighted. An extract of the most important criteria out of the two categories is listed and described below. General criteria are:
Communication is an important part of agent systems. It examines the modelling of communication and how communication protocols and languages are taken into account.
Implementation evaluates the effort required to convert the models and diagrams created in the method into program code.
Reusability examines whether models, diagrams or elements can be stored so that they can be reused in subsequent projects, making them easier to implement.
Criteria regarding autonomous control:
Flexibility refers to the variety of agent's behaviours. In attribute-based autonomous control, agents possess the communication initiative. Hence, proactive agents are needed. To realise a suited behaviour in different production situations further characteristics of agents are required (e.g., selfish, cooperative).
Roles describe representatives of various production elements that enable production control through their negotiations. This criterion examines how the development method supports the definition, presentation and description of agent roles.
Abstraction level: The distinction of objects is based on object attributes. Resulting of the different attributes, agents have different configurations. It examines how the method supports agent modelling at agent-level (single view) and system-level (overall view). Due to the very large number of existing development methods, the evaluation focus on methods being used in the field of production control.

Method	Communication	Flexibility	Roles	...	Compatibility	Detail level	Specific knowledge	Implementation	Reusability	Tool support	Scalability	Value of benefit
PASSI	2	2	2		1	0	0	2	2	2	2	56
Aspects	2	0	2		0	0	0	2	2	2	2	50
Mobmas	2	2	2		0	0	0	1	2	1	0	48
Zeus	2	1	2		0	0	0	2	0	2	1	47

Table 1 - Extract of method's evaluation

Table 1 contains an extract of the evaluation of the existing methods regarding the development of an ACS. The full evaluation contained twenty methods. The individual score of the criteria (good - 2, okay - 1, bad - 0) multiplied by their weight corresponds to the value of the benefit.

PASSI is the most suitable method, followed by Aspects [18], Mobmas [19] and Zeus [20]. PASSI is very well documented in the literature and thus has a high level of comprehensibility. Furthermore, it is tested in a variety of applications and based on the UML notations and standards, which leads to a good acceptance of the method within an industrial environment [24]. PASSI supports the common features of an agent system (e.g., modelling communication, ontology and roles) and contains agent's design on different levels (agent- and system-level). The multilevel design is especially important due to the variety of system elements and their high connectivity in production systems. For this reason, the method tailored to the application is based on the process steps of PASSI but does not completely adopt them.
The results of the step *adapting method* are presented in chapter 4.2, while the *method's application* is beyond the scope of this paper.

4.2 Structure of the method

The procedure of the developed method, which is shown in Figure 1, splits in the three main phases *Analysis, Design and Implementation*. These phases correspond to the planning and configuration phases of PASSI. The individual phases enable a purposeful approach. However, as in PASSI it is possible to proceed iteratively. Therefore, especially the definition of clear in- and outputs of all sub phases has been focused.

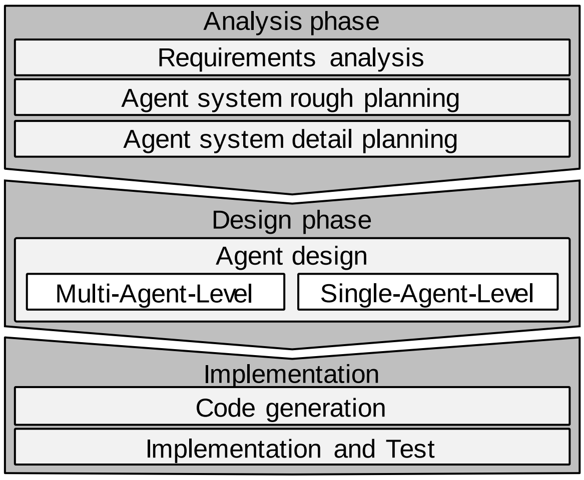


Figure 1 - Procedure of the method

The analysis phase starts with a first requirements analysis. On this basis the system is first roughly planned and after that in detail. This procedure corresponds to the factory planning process [25] and thus simplifies the utilization of the information generated there. The subsequent design phase is based on the agents in the system. In this case, as in PASSI, a multi-agent view and a single agent view is distinguished. The implementation phase finally includes the implementation and testing of the system. In the following, the phases are explained in detail.

4.2.1 Analysis phase

Input variables of the *requirement analysis* are production tasks (pre-planning) and the derived target concept (target planning) of factory planning. Also, various checklists are available to the developer for the requirements definition.
The analysis starts with a description of the goals of the production system. These are derived from the goals defined by management in factory planning. Based on the goal descriptions of the previous step, the technical requirements of the system are defined. Here, a distinction is made between functional requirements (what shall the system do) and non-functional requirements (how shall the system act). To facilitate this step, the developer can select different requirements from an already existing requirement pool. These requirements partially result of the alternatives regarding the production layout. In case of several parallel connected systems it is recommended to implement egoistic agents to increase to workload. The consequent requirement is the ability of an agent framework to involve egoistic agent behaviours. In the final step of the phase, functional and non-functional requirements are gathered in a list, which is finally checked by the involved developers.
The overall outputs of this phase are target descriptions and a list summarising all functional and non-functional requirements.
The *agent system rough planning* defines the syntax and the structure of the agent system related to the production. The function determination, the manufacturing process and the flowchart of the rough planning of factory planning form the input. Furthermore, the developer has various checklists available for the definition of system functions and the selection of a framework.
First of all the system functions are derived from the list of requirements created in the previous phase. A checklist simplifies this process step and ensures complete recording of the data. In the next step, interfaces and additional functions are defined (e.g., user interface, interfaces to handling devices). The developer is also supported by a list of predefined interfaces and functions. Subsequently, the functions are compared with the data of factory planning to check them for completeness. In the final step of the phase, a suitable framework is chosen to support the development of the system. The selection of the framework is depending on the system functions and is supported through a predefined but extendable list of frameworks.
Results of the agent system rough planning are descriptions of the identified system functions as well as descriptions of interfaces and additional functions. In addition, a suitable framework for the development will be selected during this phase.
The *agent system detail planning* is used to define the system in detail. Roles, agents and tasks are defined. The required information from previous phases and the parallel factory are the resource list incl. requirements, the interaction processes, the determined functions and the products incl. properties.
First, the elements of the production system are identified and categorised. Subsequently agent roles are defined related to the identified groups. Each role has to fulfil specific functions [5]. Predefined roles for production elements in an ACS are integrated, which can be adapted and extended. For further simplification, a tool is available to create roles and automatically convert them into program code without specific knowledge about agent systems and programming skills (see Figure 2).
To create the code the user has to determine the agents type. He can choose between physical and virtual agents. In case of a physical agent a dropdown list appears to choose between products and various resources within the production system. Depending on the resource several main and additional preconfigured tasks appear.

Objectagent				
Agent type	Physical agent		general desciption	
Physical component	Workpiece			
Tasks	Main tasks	registration …	Framework	JADE
	additional tasks	send identification attributes	File path	Link
Behaviour	Rule-based	FIFO LIFO ..	Interactions	Transportagent Machineagent ..
	situational	egoistic …	Parameter	size weight

Figure 2 - Code Generator

An analysis of manufacturing processes, following in italics, is the basis for the preconfigured role models. Depending on a resource's manufacturing process, its virtual component has corresponding tasks.
Primary forming implies the creation of a new object [26]. Therefore, a new agent has to be initialised. Primary forming agents have the task to create agents incl. their parameters and register them in the agent platform. *Reshaping* changes the shape of the body [26]. Corresponding agents need the ability to update other agents. In case an object is *separated*, one agent needs an update and another one has to be initialised or reactivated. The opposite process is *Joining* [26]. In this process, objects are put together. Hence, the agents have to be combined. Either by de-registering both agents and initialising a new agent, or by integrating one agent into the other, which leads to an update.
After the configuration is complete the user has to select a framework. Once the role is finished, it can be transformed into program code by a button. Roles are displayed in a UML-based role diagram with their dependencies. Depending on the produc-

tion program and the previously defined roles, the required agents are identified. Subsequently, tasks of the individual agents are specified and described using an activity diagram (UML). At the conclusion of this phase, the identified agents are assigned to the previously defined agent roles. For this, an agent pool is available, which contains all roles with the associated agents and eases the development of the agent classes by using predefined program templates (e.g., parameter, methods) which can be adopted for the agent classes and further specified. The templates focus on the use of Jade as an agent framework. Jade fulfils all FIPA rules and has a good performance [24, 27]. Therefore, it is very well suited for the use in an ACS. In case another framework is predefined or chosen due to specific properties of the production, it is necessary to adapt the syntax.

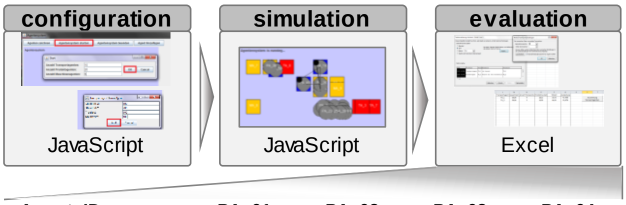
Descriptions of the system components, roles, agents and their tasks are the generated output of this phase. Further results are the role diagram and the generated agent pool.

4.2.2 *Design phase*

The *design phase* starts with the agent design at *multi-agent level*. The goal of this phase is to describe the process of the system and to work out required interactions between agents.
By clarifying relations within the later on production system using structure diagrams (UML) the required interactions are determined. Based on the identified interactions, the next step is to develop communication scenarios for each interaction and define the content language. It determines the minimum number of messages needed for interaction. This number depends on the chosen communication technique (see chapter 2.3). Afterwards, the developer has to detail every message regarding potential receivers, content and the ontology, which defines the vocabulary of the agents and describe the interaction protocol using sequence diagrams (UML).
Main output of this phase is a fully and standardised documented system incl. structure, communication and interaction protocols.
Within *single-agent-design*, agents are considered individually. Agent's tasks descriptions and the agent pool are the fundamental input.
To define the inner structure using the agent tasks descriptions and the agent pool is the first step of this phase. It describes which parameters, methods and internal classes an agent needs to fulfil his tasks. Subsequently, the inner processes of an agent are modelled using an activity diagramm (UML). The previous defined communication protocols are implemented as function blocks.
The output is the amount of specified agents whose structure and procedure are documented in UML diagrams.

4.2.3 *Implementation phase*

The final implementation phase of the method involves a complete implementation of the production system as well as a testing phase. As input, the generated UML diagrams, system descriptions and the prepared code are used.
The preconfigured program code is completed by inserting specific parameters. Placeholders and leading comments are integrated for this purpose of simplifying the code completion. After finishing the system and agent implementation, the developer programs different configurations to test the system. To evaluate the test runs another tool is available. It enables to determine the following values for each agent role (see Figure 3). The *total time* represents the time an agent has been registered in the system. Other relevant timespans are the *transport, processing and waiting time*, which divide the total time in value- and non-value-adding sections. Using these sections enables to calculate the *workload* of each component.



Agent_ID	PA_01	PA_02	PA_03	PA_04
total time	12,84	8,87	15,79	13,29
transport time	0,94	1,08	0,86	0,86
processing time	0,83	0,96	0,76	0,96
productive time	1,77	2,04	1,62	1,82
waiting time	11,07	6,83	14,17	11,47
workload	13,79%	22,99%	10,26%	13,70%

Figure 3 - Evaluation tool

The output of this final phase are the program code and its documentation as well as the evaluation of the test runs.

5 SUMMARY AND OUTLOOK

The presented procedure supports the development of an agent system as a major component of an attribute-based autonomous control. The use of the MMM lead to PASSI as basic method. The created checklists and especially the semi-automated role and code generation enable non-experts to create a testable agent system. The consideration of different agent levels allows both the individuality of production elements and their complex relationships to be considered during the development.

Next step is the extension of the various tools and resources regarding other Frameworks. In addition, an extension of the code generator by integrating the various UML diagrams even more into the process of programming is conceivable.

6 REFERENCES

[1] H. Wildemann, *Variantenmanagement. Leitfaden zur Komplexitätsreduzierung, -beherrschung und -vermeidung in Produkt und Prozess.* München. TCW-Verlag, 2016.

[2] Scholz-Reiter et al., *Einfluss der strukturellen Komplexität auf den Einsatz von selbststeuernden logistischen Prozessen,* 2006.

[3] A. Schuldt, "Multiagent Coordination Enabling Autonomous Logistics," *Künstl Intell*, vol. 26, pp. 91–94, 2012.

[4] K. Windt and M. Hülsmann, Eds., *Understanding autonomous cooperation and control in logistics. The impact of autonomy on management, information, communication and material flow.* Berlin, New York. Springer, 2007.

[5] G. Weiß and R. Jakob, *Agentenorientierte Softwareentwicklung. Methoden und Tools.* Berlin. Springer, 2005.

[6] Z. Akbari, "A survey of agent-oriented software engineering paradimg: Towards its industrial acceptance,", 2010.

[7] M. Hülsmann and K. Windt, *Understanding Autonomous Cooperation and Control in Logistics. The Impact of Autonomy on Management, Information, Communication and Material Flow,* 2007.

[8] DIN 6763, *Nummerung Grundbegriffe,* 1985.

[9] K. Krämer, *Automatisierung in Materialfluss und Logistik. Ebenen, Informationslogistik, Identifkationssysteme, intelligente Geräte.* Deutscher Universitäts-Verlag, 2002.

[10] R. F. Brena, J. P. García-Vázquez, C. E. Galván-Tejada, D. Muñoz-Rodriguez, C. Vargas-Rosales, and J. Fangmeyer, "Evolution of Indoor Positioning Technologies. A Survey," *Journal of Sensors*, vol. 2017, p. 21, 2017.

[11] B. Vogel-Heuser, J. Lee, and P. Leitao, "Agents enabling cyber-physical production systems," *at-Automatisierungstechnik*, vol. 2015, pp. 777–789, 2015.

[12] W. Brenner, R. Zarnekow, and H. Wittig, *Intelligente Softwareagenten. Grundlagen und Anwendungen.* Berlin u.a. Springer, 1998.

[13] T. Fischer, *Multi-Agenten-Systeme im Fahrzeugumschlag. Agentenbasierte Planungsunterstützung für Seehafen-automobilterminals*.

[14] W. Brenner, R. Zarnekow, and H. Wittig, *Intelligente Softwareagenten. Grundlagen und Anwendungen.* Springer, 1998.

[15] G. Görz, J. Schneeberger, and U. Schmid, *Handbuch der künstlichen Intelligenz,* 2012.

[16] VDI, *Agentensysteme in der Automatisierungstechnik - Entwicklung,* 2012.

[17] M. Cossentino and V. Seidita, "PASSI: Process for Agent Societies Specification and Implementation," in *Handbook on Agent-Oriented Design Processes.* M. Cossentino, V. Hilaire, A. Molesini, V. Seidita, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 287–329.

[18] M. Cossentino, N. Gaud, V. Hilaire, S. Galland, and A. Koukam, "ASPECS. An agent-oriented software process for engineering complex systems," *Autonomous Agents and Multi-Agent Systems*, vol. 20, pp. 260–304, 2010.

[19] Q.-N. N. Tran and G. Low, "MOBMAS. A methodology for ontology-based multi-agent systems development," *Information and Software Technology*, vol. 50, pp. 697–722, 2008.

[20] G. Weiß and R. Jakob, *Agentenorientierte Softwareentwicklung.* Berlin, Heidelberg. Springer Berlin Heidelberg, 2004.

[21] L. Kiefer, C. Richter, and G. Reinhart, "Attribute-based identification processes for autonomous manufacturing systems - An approach for the integration in factory planning mathods," in *12th CIRP Conference on Intelligent Computation in Manufacturing Engineering,* 2018.

[22] J. Ponn and U. Lindemann, *Konzeptentwicklung und Gestaltung technischer Produkte - Systematisch von Anforderungen zu Konzepten und Gestaltungslösungen.* Berlin, Heidelberg. Springer Berlin Heidelberg, 2011.

[23] Q.-N. N. Tran, G. Low, and M.-A. Williams, "A Feature Analysis Framework for Evaluating Multi-agent System Development Methodologies," in *Foundations of Intelligent Systems*, vol. 2871. G. Goos, J. Hartmanis, J. van Leeuwen, N. Zhong, Z. W. Raś, S. Tsumoto, E. Suzuki, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 613–617.

[24] K. Kravari and N. Bassiliades, "A Survey of Agent Platforms," *JASSS*, vol. 18, 2015.

[25] C.-G. Grundig, *Fabrikplanung. Planungssystematik - Methoden - Anwendungen.* München. Hanser, 2015.

[26] DIN 8593, *Fertigungsverfahren Fügen.* Berlin, 2003.

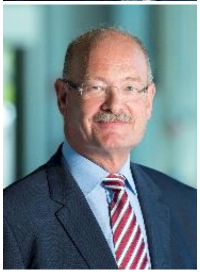
[27] J. W. Tweedale, "A Review of Cognitive Decision-making within Future Mission Systems," *Procedia Computer Science*, vol. 35, pp. 1043–1052, 2014.

7 BIOGRAPHY

Lucas Kiefer, M. Sc., graduated in Mechanical Engineering at the Technical University of Munich. Since 2015 he performs as a research associate and is group leader in the field of flexible systems engineering at the Fraunhofer IGCV.

Daniel Königl, B. Sc., graduated in computer science at the University of Augsburg. Since 2016 he works as a research assistant at the Fraunhofer IGCV.

Dr.-Ing. Christoph Richter graduated in Mechanical Engineering at the Technical University of Munich. Since 2016 he is head of the department for manufacturing systems control at the Fraunhofer IGCV in Augsburg.



Prof. Dr.-Ing. Gunther Reinhart holds the chair for Industrial Management and Assembly Technologies at the *iwb* of the Technical University of Munich. Furthermore, he is Executive Director of the Fraunhofer IGCV in Augsburg. The *iwb* and the Fraunhofer IGCV conduct research into assembly technology, automation, robotics as wells as production management and logistics.

Explorative Investigation of Application Scenarios for Smart Bin Systems

R. Vogt[1], V. Hummel[1], J. Schuhmacher[1], K.H. von Leipzig[2], L. Louw[2]
[1]ESB Business School, Reutlingen University, Germany
[2]Department of Industrial Engineering, University of Stellenbosch, South Africa

Abstract
Increasing flexibility, greater transparency and faster adaptability play a key role in the development of future intralogistics. Ever-changing environmental conditions require easy extensibility and modifiability of existing bin systems. This research project explores approaches to transfer the Internet of Things (IoT) paradigm to intralogistics. This allows a synchronization of the material and information flow. The bin is enabled by the implementation of adequate hardware and software components to capture, store, process and forward data to selected system subscribers. Monitoring the processes in the intralogistics by means of the smart bin system ensures the implementation of appropriate actions in case of defined deviations. By using explorative expert interviews with representatives from the automotive and pharmaceutical industries, seven practical application scenarios were defined. On this basis, the requirements of smart bin systems were examined. For each individual case of application, a system model was created in order to obtain an overview of the system components and thus reveal similarities and differences. Based on the similarities of the system models, a general requirement profile was derived. After the hardware components of the bin system had been determined, a utility analysis was carried out to find the adequate IoT software. The utility analysis was conducted with a focus on data acquisition and data transfer, data storage, data analysis, data presentation as well as authorization management and data security. The results show that there is great interest in easily expandable and modifiable bin systems, as in all cases, the necessary information flow in the existing bin system has to be improved by means of new IoT hardware and software components.

Keywords
Intralogistics, Smart bin system, IoT Software, Modifiability of bin systems

1 INTRODUCTION

Many companies intend to optimize certain parts of their intralogistics network, and thus their service processes face a variety of technical and business challenges [1, 2]. Of particular interest is the management of bins, including the planning, controlling and monitoring of bins and inventories. The non-transparent flow of material, bins and information, the lack of consistent object acquisition, the unreliable acquisition of bins, missing real-time related data and the shortage of visual support for work processes pose challenges in bin management [3]. According to a study by the Institute of Computer Science and Society at the University of Freiburg regarding the introduction of smart bins, the following obstacles are considered major challenges [4, 5].

According to the survey, 68% of the companies identified the complex integration into cross-company business processes as a central challenge. In addition, the integration into the existing IT infrastructure is perceived by 55% and into the internal business processes by 48% of the participants as a major hurdle. Furthermore, half of the companies do not see the possibility of measuring the benefits of innovative bin systems (50 %). [4, 5]

Integrated bin systems, which generate data at a field level by means of appropriate sensors and actuators and forward the information to the superordinate planning and control systems, hardly exist in intralogistics. A more detailed classification of the role of data in the provision of operational services can be found in a white paper published by Otto et al. [6].

According to Strassner and Fleisch [7] automation must be enhanced in order to reduce manual workload in intralogistics. However, it must be noted that potential savings through the implementation of innovative technologies decrease as the level of automation increases. Figure 2 shows the correlation between costs incurred depending on the degree of automation. The costs are divided into IT costs, workload and potential costs resulting from errors.

A potential benefit of smart bin systems is the increase in process quality, since fewer errors can occur compared to manual processes and permanently automated processes are possible [8]. In the case of less automated processes, e.g. where the use of other Auto ID technologies does not work, savings in labour can exceed the costs of a smart bin system. In this case, the optimum level of automation, without taking into account the potential costs resulting from errors (marked in Figure 1 with 1), is where the sum of the workload and IT costs is minimal. In addition, high potential costs resulting from errors can justify a higher degree of automation. The optimum degree of automation (marked in Figure 1 with 2) is where the sum of all displayed costs is minimal.

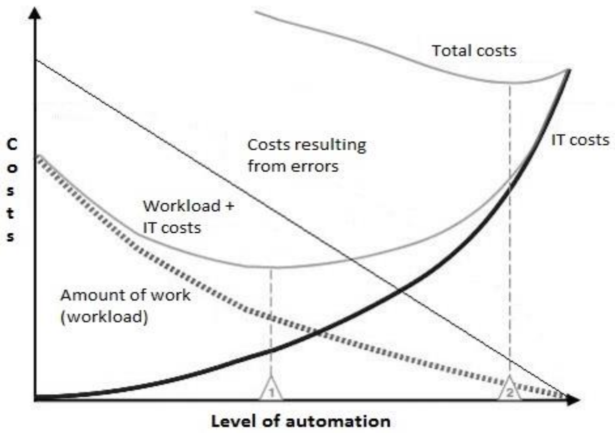


Figure 1 - Correlation between IT costs, workload and potential costs resulting from errors depending on the degree of automation [7]

2 LITERATURE REVIEW

The following sections provide an accurate classification of the terminology used in this paper. In addition, it shows which literature is used as basis.

2.1 Definition of smart bin and smart bin system

The term smart bin used in this research project includes the two definitions given by Bogatu [9]:

- Tagged bin: A reusable bin that can be uniquely identified by means of an AutoID technology and that can receive, store and communicate the relevant process data autonomously, but not entirely without external influence.
- Smart bin: This includes and extends a tagged bin. A smart bin can record process-related information as well as mechanical and hygienic bin conditions independently or through an external influence. At the same time, the bin is able to evaluate both the data received from outside and the data generated independently, to enrich it with further information and to transmit it to the outside via several communication channels.

According to Böhm et al. [10] and Ropohl [11] a smart bin system is a:

- business information system
- socio-technical (human-machine) system in which the system element "machine" consists of the smart bin and the associated supporting IT infrastructure, while the system element "human" includes the employees participating directly in the logistical processes as well as the management who are responsible for the implementation and improvement of logistics processes.

In addition, its primary purpose is to support the planning and operational execution as well as the management levels in the logistics division.

2.2 *Classification of smart bin systems*

The development towards smart products and objects raises the question of an objective system for the evaluation and classification of the level of intelligence. A system described by Tüllmann et al. to determine the level of intelligence of products and objects has been further developed by the authors into a morphological box as shown in Figure 2 [12]. By means of this modified system, it is possible to evaluate and compare the developed smart bin systems.

Morphorlogical Box	Classification of smart bin systems	Date	23.08.2018	ESB
	Application area: ASA and ASP	Editor	Raphael Vogt	INDUSTRIAL ENGINEERING

	Characteristics / Degree of characteristics	Alternatives 1	2	3	4	5
A	Identification and smart data processing	No use of sensors and actuators	Sensors/ actuators are integrated	Bin system continues to process data	Data processing and use of data for analysis purposes	Independent actions based on newly gained data
B	Memory	No data storage	Transponder with electronic data storage	Additional data memory integrated in the bin system	Cloud-based data storage	Cloud based data storage, filtering and sharing
C	Interaction / Communication	No interface integrated	Data input possible	Data input and output possible	Connection to network via Ethernet interface	Connection to net-work (eg WLAN or mobile radio)
D	Additional services	No additional services	Manually via online platform	Manually at the bin system	Automatic by bin system with manual control	Automatic by bin system in real time
E	Control system / monitoring	No monitoring of the operating state	Monitoring of the operating state	Monitoring of operat-ing state and report of breakdowns	Self-diagnosis if error message	Independent error management
F	Networking	No networking	Information exchange via e-mail or telephone	Uniform data formats, rules and interfaces for exchange	Uniform data formats and inter-divisional servers	Inter-divisional fully networked IT solution

Figure 2 - Morphological box for the classification of Smart Bin systems based on Tüllmann et al. [12]

3 CONDUCTING THE EXPLORATORY STUDY

By using explorative expert interviews with representatives from the automotive and pharmaceutical industries, seven practical application scenarios have been defined. On this basis, the requirements of smart bin systems are investigated. For each individual case of application, a system model is then created in order to obtain an overview of the system components and thus reveal the similarities and differences. Based on the similarities of the system models, a general requirement profile has been derived. After the hardware components of the bin system have been determined, a utility analysis is carried out to find the adequate IoT software.

3.1 Practical application scenarios based on expert interviews

Seven experts were interviewed in an approximately one-hour telephone interview. Four of the experts represent the automotive industry and three the pharmaceutical industry. In the selection process, it was ensured that the expert had a background in the strategic management of intralogistics. The main focus of the interview guideline was on improvement potential in bin management, strategic implementation and technological realisation. A particular focus of the present paper is on application scenarios of the automotive (ASA) and pharmaceutical (ASP) industries in Germany.

This is due to the fact that the German automotive industry is of great significance for the German as well as the European economy. The number of logistics processes in the automotive industry has been increasing for years. The main goal of automobile companies is to optimize the flow of information between manufacturers, suppliers and logistics service providers. Companies need transparency throughout the entire ordering and delivery process. [13]

The pharmaceutical industry is the third largest industry in Germany after the automotive and mechanical engineering industries (in terms of sales figures). However, the net value added per employee is the highest. The pharmaceutical industry is growing largely independently of economic development and has a high research intensity and innovation pressure because it is an industry of highest complexity. [14, 15]

3.1.1 ASA 1: Monitoring of C-parts

ASA1 describes an application scenario in which bins are stored in drawers of cabinet systems. The C-parts in the bins are monitored. Bins for which monitoring is required are equipped with camera and indoor localisation module, MCU, RFID tag and battery. If a minimum stock level is reached, an automatic message is sent to the warehouse keeper. Hence, it is a consumption- and order-dependent parts procurement.

3.1.2 ASA 2: Condition monitoring in after-sales

This application scenario focuses on the monitoring of high-quality spare parts in after-sales. The primary concern is shock and temperature monitoring as well as theft protection. In addition, the tracking of spare parts from the storage compartment to packaging and preparation for shipment is to be guaranteed. This ensures that on-time delivery can be met. In other words, it monitors whether the part is leaving the plant on time and undamaged after receiving the order.

3.1.3 ASA 3: Tracking of sequence containers

The bin system is used to track empties in the assembly hall to counteract the shortage of sequence containers in supply centers. The sequence containers are stocked in the supply centers with the parts depending on the sequence of the assembly line. The sequence containers are then transported to the corresponding assembly line by a tugger train. The cycle times at the assembly line can be controlled by the use of an indoor localization system. In addition, there is an increased flow of information for the tugger train drivers. The part picking is monitored to obtain additional information on how long the individual assembly step takes and when all parts are taken out to return the sequence container.

3.1.4 ASA 4: Tracking of assembly line stoppers

In ASA4, parts, which are essential for the further assembly of the vehicle, are monitored separately. The purpose is to be constantly informed about the condition and position of the highest priority parts. In addition, the number of stored parts (safety stock) shall be reduced due to the increased flow of information. This enables a position monitoring and an insight into whether the parts are at the correct assembly station.

3.1.5 ASP 5: Condition monitoring of drugs

The first application scenario in the pharmaceutical industry describes the monitoring of drugs, which react sensitively to temperature, pressure and humidity. In addition to room monitoring of the drug storage area, bin monitoring is to be introduced for selected bins. A selective monitoring and recording of environmental changes is possible. In addition, a seamless monitoring of drugs outside the drug storage, like in the dispatch department, is becoming feasible.

3.1.6 ASP 6: Process control & safety in service

The next application scenario of the pharmaceutical industry is the repair of instruments from the Healthcare and Pharma sector. When the instruments are delivered, it is uncertain whether sterilization has already been done. For this reason, the instruments are assigned to a bin with neutral color at goods receipt. The respective process step is indicated or changed optically via a display on the bin. The normal warehouse goods are stored in bins with a blue display, for hazardous goods yellow is used, high-risk goods are marked red and the instruments, which have already been evaluated and cleaned, are marked green. Additional indoor tracking ensures that high-risk goods are handled quickly and correctly. In addition, it guarantees that a complete proof of the process chain can be shown at Medical Devices article audits.

3.1.7 ASP 7: Storage of pharmaceuticals based on action and warning limits

Pharmaceuticals, which are particularly sensitive to temperature, are stored in different bins depending on action and warning limits. A pharmaceutical is a drug used as a medically active ingredient in the manufacturing process of a drug [16]. If a predefined limit is exceeded or not reached, the system displays the necessary action to the manager. If the action is not executed on time and the limit has been exceeded or undercut for too long, the system displays a warning and the pharmaceuticals must be discarded. Managers have the authorizations to make changes/entries in the system. They receive action and warning messages to take appropriate actions.

3.2 Derivation of industry-specific requirements

After evaluating the industry-specific requirements, it is striking that the focus in the automotive industry is on bin management and tracking. Filling level measurement, which means the monitoring of a certain minimum stock level, also represents an important factor. The function of indoor localization is mainly used to monitor process time and adherence to delivery dates.

In the pharmaceutical industry, however, the focus is on environmental monitoring. The continuous recording and storage of process information is particularly important for regularly scheduled audits. If certain environmental parameters are not met, the corresponding action messages must be sent to the responsible employees.

The evaluation of the expert interviews shows that a modular design of the bin system is required in both industries. This makes it possible to modify the smart bin system and adapt it to other circumstances. Furthermore, in the experts' opinion it is important that the module is not firmly integrated into the bin. This allows to equip only certain bins of a bin cycle with the module and to remove the module when the bin is leaving the company and a supply chain monitoring is not required. Another important component is the location-independent use. This mainly concerns the energy supply.

3.3 Hardware selection

The following sections provide a detailed technology selection and the development of a first prototype.

3.3.1 Technology selection

Before the experts were contacted, potential technologies were selected based on an internet research. The technologies highlighted in grey in Tables 1-5 are the selection of technologies that experts consider to be the most appropriate for the use in an operating environment.

Wi-Fi (IEEE 802.11 a/b/g/n)
Wi-Fi Direct à peer-to-peer specification
Mobile communications
ZigBee (IEEE 802.15.4)
Bluetooth (IEEE 802.15.1)
Bluetooth 4.0 LE Technology

Table 1 - Selection of communication technologies

RFID (low, high or ultra-high frequency)
NFC
Barcode, QR-Code

Table 2 - Selection of identification technologies

Laser Tracker & iGPS
Assisted global positioning system
Ultrasound systems (Telocate, Cricket or IMAPS)
Radio-network-based positioning (RADAR)

Table 3 - Selection localization technologies

Integrated camera
Weight sensors
Infrared light barrier

Table 4 - Selection of technologies for filling level measurement

Temperature monitoring
Pressure monitoring
Shock monitoring
Humidity monitoring

Table 5 -Selection of additional sensors

3.3.2 Outcome of the hardware selection

A first prototype, using the selected technologies, was realized with the Rasperry Pi Zero as shown in Figure 3 [17]. Further variants investigated in the course of this project were the Arduino uno, NodeMCU and ZF Tagfinder with deTAGtive TAGs. [18, 19, 20] However, the Raspberry pi zero proved to be the most suitable solution due to the BlueDot sensors available, the CSI camera connector on the board, the Wifi and Bluetooth LE module and the very low energy consumption. Furthermore, the Raspberry pi zero is compatible with the Losant IoT platform described later in this paper.

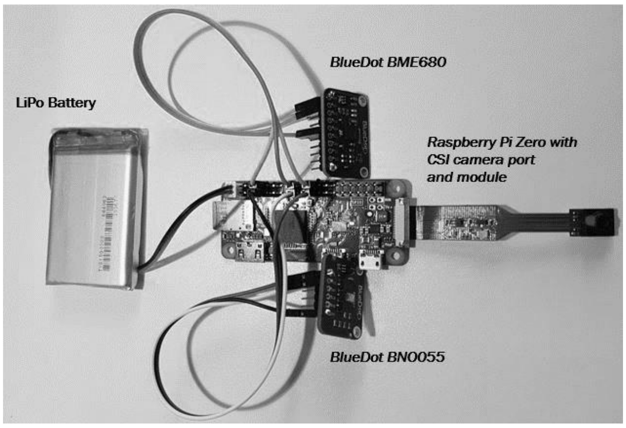


Figure 3 - Setup of the smart bin module

3.4 Utility analysis in order to identify suitable software

The utility analysis aims to identify potential software solutions for the industry-specific applications as described beforehand. By analyzing the platforms available on the IoT market, the work done can serve as a decision-making guide in the selection of a suitable IoT platform. Depending on the application and prioritisation, the benefit analysis can be adapted. However, not all possible platforms are examined. Due to the growing number of products on the IoT market, only the major IT companies are considered in the analysis. In addition, the IoT platform must be well established on the IoT market for a certain period of time.

3.4.1 Conducting the utility analysis

The following selection of IoT platforms is examined in more detail in the course of the paper: PTC ThingWorx, Bosch IoT Suite, IBM Bluemix IoT Zone, HPE Universal IoT Platform, Oracle IoT Cloud Service and Losant IoT Platform.

The necessary expertise for the analysis is accomplished by means of Internet research, telephone calls with qualified employees and, if possible, by means of testing using trial versions provided. The analysis of the IoT platforms is carried out on the basis of the following four categories shown in Table 6. The individual categories are divided into more detailed requirements. According to the categories and requirements in Table 6, the IoT platforms are analyzed for advantages and disadvantages, followed by a pairwise comparison and a utility value analysis.

Data acquisition and transfer	1. M2M device integration 2. Integration of environmental data (e.g. production plan) 3. Integration of IT systems 4. Transformation of data for filtering of certain information out of raw data (ETL process)
Data analysis and processing	1. High compatibility in handling data volumes and diversity 2. Finding trends and patterns 3. Monitoring and reporting 4. Alarm functions (e.g. SMS)
Data presentation via user interface (GUI)	1. User-friendly end-user GUI for presentation of information 2. User-friendly admin GUI (extending IoT platform) 3. Integration of PC or Tablet
Authorization management, data security and protection	1. Encrypted transmission technologies to prevent intruders 2. User and role concept 3. Adherence of data privacy 4. Structure of license model

Table 6 - Defined categories with corresponding requirements

3.4.2 Outcome of the utility analysis

If the requirements, elaborated application scenarios and hardware selection are taken into account, the Losant IoT platform performs best with 6.4 out of 10 points. This value includes the weighting of the individual requirements determined in a previous pair wise comparison. The final results of the utility analysis are illustrated in Figure 4.

		PTC ThingWorx		Bosch IoT Suite		IBM Bluemix		HPE IoT Pl.		Oracle IoT Cl.		Losant IoT Pl.	
	Weighting	Average of ratings	Value	Average of ratings	Value	Average of ratings	Value	Average of ratings	Value	Average of ratings	Value	Average of ratings	Value
M2M Device Integration	12.38%	7	0.87	8	0.99	6	0.74	5	0.62	6	0.74	9	1.11
Further environmental data	7.14%	6	0.43	5	0.36	7	0.50	5	0.36	4	0.29	5	0.36
Further IT systems	8.57%	7	0.60	6	0.51	6	0.51	5	0.43	8	0.69	3	0.26
Information filtering	3.33%	4	0.13	7	0.23	6	0.20	6	0.20	7	0.23	5	0.17
Working with Big Data	2.86%	8	0.23	5	0.14	4	0.11	6	0.17	5	0.14	5	0.14
Predictive analytics	2.38%	6	0.14	5	0.12	7	0.17	8	0.19	3	0.07	3	0.07
Monitoring and reporting	8.57%	5	0.43	7	0.60	5	0.43	6	0.51	4	0.34	6	0.51
Alarm function	11.90%	6	0.71	7	0.83	6	0.71	6	0.71	6	0.71	6	0.71
End-user GUI	10.00%	8	0.80	1	0.10	6	0.60	4	0.40	6	0.60	7	0.70
Admin-GUI	0.95%	7	0.07	1	0.01	4	0.04	6	0.06	5	0.05	6	0.06
Integration PC, Smartphone	7.62%	7	0.53	4	0.30	6	0.46	6	0.46	5	0.38	6	0.46
User and role concept	4.76%	7	0.33	7	0.33	6	0.29	6	0.29	7	0.33	6	0.29
Encrypted transfer technology	3.33%	6	0.20	6	0.20	7	0.23	6	0.20	6	0.20	7	0.23
Data privacy	3.33%	4	0.13	9	0.30	5	0.17	6	0.20	6	0.20	5	0.17
License cost model	12.86%	3	0.39	3	0.39	6	0.77	3	0.39	3	0.39	9	1.16
	Sum		6.00		5.42		5.93		5.18		5.37		6.40

Figure 4 - Result of the utility analysis

4 CONCLUSIONS AND OUTLOOK

This paper provides an overview regarding the introduction of smart bin systems for companies in the automotive and pharmaceutical industries. A possible solution is approached from an operative point of view. Based on expert interviews, possible application scenarios and the corresponding requirement profiles were derived. This enabled a targeted technology and subsequent hardware selection. Afterwards, a suitable software was selected by means of an adaptable utility analysis. The methodology presented in this paper can be adapted to specific application scenarios and requirement profiles. The results of the internet research and expert interviews indicated great interest in expandable and modifiable bin systems.

In the next steps of the research project, the prototype will be introduced to the Logistics Learning Factory (LLF) of Reutlingen University for validation. In the LLF environment, the defined requirement profiles will be examined and evaluated in more detail. In this way, the specified requirements can be assessed in terms of their relevance and implementation potential. Furthermore, previously unrecognized requirements for the functionality of the smart bin system are to be reassessed. At the same time, the prototype should point out the limits of the smart bin system and form the basis for further research projects.

5 REFERENCES

[1] Prives, A., 2016, Systemkonzept zur Steigerung logistischer Effizienz im Lebensmitteleinzelhandel durch Einsatz intelligenter Behälter, dissertation, Universität München, pp. 70-78.

[2] Emmerich, J., Roidl, M., Bich, T. and ten Hompel, M., 2012, Development of Self-sustaining Intelligent Load Carriers Using the Example of the inBin, pp. 1-3.

[3] Sydow, A., 2017, Dynamische Behälterlogistik: Konzeption eines IT-gestützten Management- und Entscheidungsunterstützungssystems für automobile Mehrwegbehälterkreisläufe, dissertation, Universität Clausthal, p. 6.

[4] Graef, H. and Wesemannruzicka, J., 2010, Kanban mit RFID: Verbindung der Materialflussebene mit der Geschäftsprozessebene, in HMD Praxis der Wirtschaftsinformatik 47(2), pp. 36-48.

[5] Strüker, J., Gille D. and Faupel T., 2008, RFID-Report 2008: Optimierung von Geschäftsprozessen in Deutschland, in Sonderpublikation der VDI Nachrichten, Inst. für Informatik und Ges., Abt. Telematik, Freiburg.

[6] Otto, B., Jürjens, J., Schon, J., Auer, S., Menz, N., Wenzel, S. and Cirullies, J., 2016, Industrial Data Space: Digitale Souveränität über Daten, Fraunhofer-Gesellschaft, Munich.

[7] Strassner, M. and Fleisch, E., 2005, Innovationspotenzial von RFID für das Supply-Chain-Management: WI-Schwerpunktsaufsatz, in Wirtschaftsinformatik 47(1), pp. 45-54.

[8] Verband der Automobilindustrie, 2016, RFID im Behältermanagement der SC, version 2.2, p. 8.

[9] Bogatu, C., 2008, Smartcontainer als Antwort auf logistische und sicherheitsrelevante Herausforderungen in der Lieferkette - Auswirkungen und Handlungsempfehlungen für die Wertschöpfungskette der Logistik, dissertation, Technische Universität Berlin.

[10] Böhm, R.; Fuchs, E. and Fischer, M., 2002, System-Entwicklung, in Wirtschaftsinformatik, 5. edition, vdf Hochsch.-Verlag, ETH, Zürich.

[11] Ropohl, G., 2009, Allgemeine Technologie - Eine Systemtheorie der Technik, 3. edition, Univ.-Verlag Karlsruhe.

[12] Tüllmann, C. Prasse, C., Sagner, D. and Piastowski, H., 2016, Future Challenges in Logistics and Supply Chain Management: Prozesse durch Digitalisierung nachhaltig optimieren, Fraunhofer-Institut für Materialfluss und Logistik IML, Dortmund, pp. 9-12.

[13] Verband der Automobilindustrie e.V, 2016, Wissmann: Automobillogistik wird immer globaler und komplexer: Forum Automobillogistik und BVL, Frankfurt am Main. [online] Available: https://www.vda.de/de/presse/Pressemeldungen/20160203-automobillogistik-wird-immer-globaler-und-komplexer.html [Accessed 17 Aug. 2018].

[14] Philipp, P. M. and Hettinger, F., 2017, Die Zukunft der Pharmaindustrie liegt in der Vereinfachung, [online] Available: http://www.chemanager-online.com [Accessed 17 Aug. 2018].

[15] Statista, 2017, Statista-Dossier zum Thema Pharmaindustrie in Deutschland, Article number: did-14282-1.

[16] Bayer AG, 2011, Galenik: Die Chemie der Medikamente, Forschungsmagazin „research“, 23 edition, p. 1a.

[17] Raspberry Pi Foundation, 2018, Raspberry Pi Zero W, [online] Available: https://www.raspberrypi.org/products/raspberry-pi-zero-w/ [Accessed 20 Aug. 2018].

[18] Arduino, 2018, Arduino, [online] Available: https://www.arduino.cc [Accessed 9 Aug. 2018].

[19] NodeMcu Team, 2014, NodeMcu, [online] Available: http://nodemcu.com/index_en.html [Accessed 20 Aug. 2018].

[20] Korzer, P.; Schlesinger, B.; Boehm, F.; Becker, M. and Bogoczek R., 2017, Condition Monitoring: Prozessleitfaden und Klickanleitung.

6 BIOGRAPHY

Raphael Vogt holds a B.Eng degree in Mechanical Engineering from the University of Applied Science in Konstanz. He is currently doing his MSc degree in Digital Industrial Management and Engineering at the ESB Business School in Reutlingen. His major field of research are smart bin systems in intralogistics.

Vera Hummel, Prof. Dr.-Ing., Dipl.-Ing., has been a professor at the ESB Business School, Reutlingen University since 2010. She is a founding member of the “Initiative of European Learning Factories” and the initiator of the "ESB Logistics Learning Factory“ for education and industry training at the Reutlingen University.

Jan Schuhmacher (MSc.) is a research associate at ESB Business School and part of the research, training and consultant group of ESB Logistics Learning Factory. His major field of research is the design and optimization of changeable intralogistics systems.

Konrad von Leipzig obtained both his B.Eng and M.Eng degrees in Industrial Engineering. He later also completed a B.Comm degree. He has been a lecturer at the Department of Industrial Engineering at Stellenbosch University since 1987. He forms part of the team undertaking the realisation of the Learning Factory. His research areas include process, supply chain, logistics and financial management.

Dr. Louis Louw is a senior lecturer in Industrial Engineering at Stellenbosch University. He obtained the degrees B.Eng and Phd in Industrial Engineering. His research fields include Enterprise Engineering, Innovation and Operations & Supply Chain Management A key focus is on innovation, digitalisation and automation in the supply chain. His industry experience spans over 12 years of business consulting working. He is also responsible for establishing a Learning Factory.

Approach and Tools for Business Model Development in Context of Industry 4.0

V. Hummel[1], M. Guldin[1]

[1]ESB Business School, Reutlingen University, Germany

Abstract

Due to Industry 4.0, the full value creation has the chance to undergo a fundamental technological transformation, the realisation of which, however, requires the commitment of every company for its own benefit. The new approaches of Industry 4.0 are often hardly evaluated, let alone proven, so that SMEs in particular often cannot properly estimate the potentials and risks, and often waiting too long with the migration towards Industry 4.0. In addition, they often do not pursue an integrated concept in order to identify possible potentials through changes in their business models. . As part of the research project "GEN-I 4.0 – Geschäftsmodell-Entwicklung für die Industrie 4.0" ", the ESB Business School at Reutlingen University of Applied Sciences and the Fraunhofer Institute for Industrial Engineering and Organization FHG IAO were engaged by the Baden-Württemberg Foundation from 2016 to 2018 to develop tools and an approach how the local economy can develop digital business models for itself in a methodical, beneficial and targeted manner. Through international analyses and interviews GEN-I 4.0 gained and concretized the knowledge required for the evaluation and selection of solutions and approaches for the transfer to develop digital business models. Together with the know-how of the project partners on Industry 4.0 and business model development, the findings were incorporated into the development of two software tools with which SMEs are shown the potentials of Industry 4.0 for their individual business model, online and in self-assessment, and given a comprehensive structured, concrete approach to development, as well as their individual risk. Users of the tools are supported by the selected platform for the networking of different players to implement innovative business models accompanied by coaching concepts for the companies in the follow-up and implementation of the assessment results.

Keywords

Industry 4.0, Business Model Development, Self-Assessment, Comprehensive Approach

1 INTRODUCTION

The manufacturing industry and business-related services, which make a significant contribution to the economic success of the state of Baden-Württemberg [1], were missing at the time of application valid methods for individual potential assessment, transfer formats that facilitate access to research facilities and results for companies and support them in evaluation and implementation of digital business models [2, 3, 4]. On the research side, only the BMBF project GEMINI was initialized in the addressed field of research at the beginning of 2016, which tries to develop a complex set of instruments for the development of business models in the context of Industry 4.0 for completely new business ideas [5]. The migration of existing business models, as is the case of GEN-I 4.0, was not in the focus. Methods, which support companies in the evolution of their processes and business models towards Industry 4.0, had to be able to include the broad spectrum of emerging technologies and to correlate them with the already existing individual value creation drivers and the existing process and technology maturity levels and to additionally formulate a target state for the future business model and a realistic and step-by-step path leading there.

A clear description of the elements and the development of business models for Industry 4.0 was completely missing at the time of application. While the GEN-I4.0 project was able to address numerous mentioned deficits from the point of view of the state of the art in science and technology as part of its work program, other research institutes and consulting firms have also significantly intensified their activities with regard to digital business models. In the meantime, the RWTH Aachen University, the University of Potsdam, the University of Bayreuth and the Quadriga Hochschule Berlin are offering courses, seminars and, in some cases, entire study courses on the subject of business models and digital transformation [6]. The Industry 4.0 platform operated by the VDI set up a working group for digital business models in 2018 to "analyse the use cases available on the market and develop [...] building blocks, mechanisms and typologies" [7] - a task that the GEN-I4.0 project has already completed.

At the same time, the level of industrial implementation in relation to digital business models has not followed the same trend during the project period. Thus Riemensperger et.al [8] reports under the title "Digital business models without business" that industrial companies, similar to the findings of the

GEN-I4.0 project team, on average still do not generate any significant revenues from digitisation activities and digital business models. To this extent, Kreimeier [9] also reports, likewise in line with the findings of this project, that the majority of manufacturing companies have now appointed and allocated digitisation managers and budgets, "but only a few companies see the possibility of penetrating new business fields. [...] "New business models would [...] have no priority" [6] names too intensive involvement of management in day-to-day business as an obstacle to dealing with new business models - but at least reports on the need for further training programs on the subject of business model modelling.

2 PROJECT AND RESEARCH GOALS

The creation of value in Germany has the chance of a fundamental technological change through Industry 4.0, the realization of which, however, requires the commitment of every company to develop its individual potential. The new approaches of Industry 4.0 are often hardly evaluated, not to mention proven, so that especially SMEs lack not only an overview but also knowledge regarding their interaction, their potentials and risks. They therefore run the risk of waiting too long with the migration towards Industry 4.0, of not pursuing a holistic concept or of not dealing with possible changes in the business model in a structured way.

2.1 Project description

For exactly this purpose, GEN-I 4.0 gained and specified the knowledge required for the evaluation and selection of solutions and approaches for the transfer into digital business models through national and international analyses and interviews. It is integrated together with the know-how of the project partners on Industry 4.0 and on business model development in two software tools, with which SMEs are shown the potential of Industry 4.0 for their individual business model online and in self-assessment and can so determine a structured, concrete way to develop it and determine their individual risk. The users of the tools are informed and networked through a community platform and continuous transfer events. Ultimately, they also benefit from a coaching program that supports them in implementing the results from the application of the developed tools.

2.2 Research Goals

The individual elements to be developed are components of a holistic approach for the transformation or further development of existing business models on the basis of so-called enabling technologies in classical industries of significant relevance to the economic performance of Baden-Württemberg. This approach is flanked by networking events and workshops for the evaluation and dissemination of the developed tools and methods, for the networking of know-how providers in the individual enabling technologies as well as for the competence development in further methods generally relevant for the development of business models. The result is a validated, evaluated and disseminated repertoire of tools to support a methodical and structured approach to the digitisation of existing business models. According to the project proposal, the overriding objectives can be summarized as follows:

- Creating understanding of what characterizes successful business models. Not only the respective business model itself is relevant, but also its specific ecosystem (regional aspects, partners from research, investors, politics, associations, etc.).
- Development and validation of an online self-assessment and risk management tool.
- Provision of a platform for the networking of different players for the realization of innovative business models and coaching of companies in the follow-up and implementation of assessment results.

This paper presents the approaches and research results regarding the cascade, self-assessment tool, the web-based information and networking platform as well as the holistic approach model.

3 RESEARCH APPROACH AND RESULTS

Founded by Baden-Württemberg Stiftung gGmbH from 2016 till 2018 the following research work with corresponding results were carried out.

3.1 Create understanding for digital business models

The first research work focused on the research and identification of typical starting points and approaches for the digitisation of a business model as well as the identification of technologies whose possibilities and potentials for the digitisation of the business model are based on. In a first step, 94 already available German and English studies, white papers, strategy papers, dossiers, articles, lecture papers and book chapters related to Industry 4.0 and digitisation were catalogued and analysed for the processing. By a qualitative literature analysis of the material with the help of the analysis software MaxQDA it was possible to identify (127) relevant technology terms in the context of Industry 4.0. In the second step, the general significance of these Industry 4.0 technologies was first examined using a software-supported quantitative frequency analysis.

Subsequently, a cascading technology model (see Figure 1) could be derived and developed, which consists of basic technologies and methods, tools/devices, concepts and approaches, use cases and business models. For this purpose, the identified technology terms were subjected to classification due to their fundamental reciprocal inconsistency and can now be described as follows: Numerous identified terms can be assigned to the group of basic technologies and methods that are not functional in themselves, but generically define a procedure or a standard. Tools and devices, on the other hand, integrate and partly use several such basic technologies into concrete functional units that can be used for the provision of services or added value and fulfil a specific purpose. This application pursues a specific concept or approach from the context Industry 4.0. The transfer of this concept including the required tools/devices and basic technologies into the concrete company-specific context generates a use case.

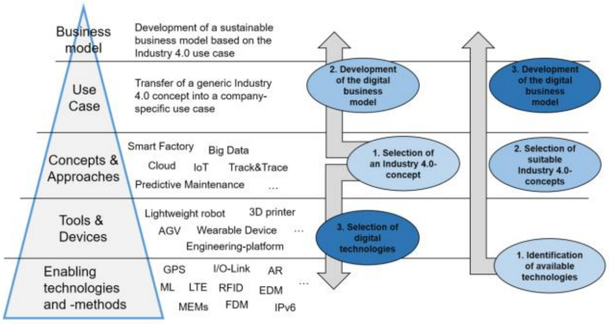


Figure 1 - Cascade of technology terms and possible approaches (2) for business model development

Two approaches for the development of a digital, technology-based business model could be identified (see Figure 1): On the one hand, the identification of a promising Industry 4.0 concept for a digital business model can result in its development. On the other hand, the identification of already existing technologies can be a starting point to develop a business model from such Industry 4.0 concepts, which build on those already existing technologies.

Within the scope of the literature analysis described, additional (39) application examples for digital/digitized business models could be found. Based on these examples, the correlation between the identified specific digital business models and the generic Industry 4.0 concepts and approaches could be analysed for the first time. Within the framework of this activity, (13) I4.0 concepts could then be identified, which have relevance and significance for the digitisation of existing business models and which were combined into (12) concepts on the basis of functional overlaps. All industry I4.0 concepts with business model relevance are explained in detail in the study (see study published https://www.digitale-geschaeftsmodelle-bw.de).

To validate the results to date, experts were identified in the next step for the planned baseline and depth surveys. Two comparison groups were defined for the survey. One group has not yet introduced a digital business model, while the other has done so. The approach with two comparison groups was chosen because it was established in the previous work that the personal affinity and attitude to the relatively abstract topic of digitisation could have a strong impact on the survey results.

Initially, this survey proved that the business models of the companies interviewed can all be assigned to the generic I4.0 concepts identified as relevant to the business model in the initial research findings. The complete results of the study are published at https://www.digitale-geschaeftsmodelle-bw.de.

3.2 "GEN-I Scheme" a penetration degree model

In the first step, existing maturity models for the description of Industry 4.0 or digitisation maturity were examined. Based on the analysis, it was possible to develop an adapted, six-level penettration degree model based on the Forrester IoT model and validate it on the basis of the (35) digital business models examined (see figure 2).

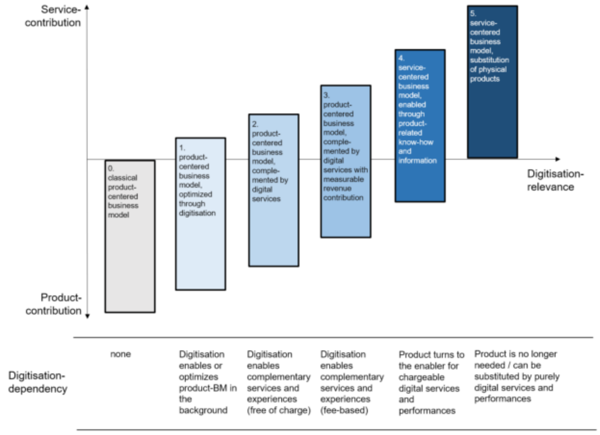


Figure 2 - GEN-I Scheme for categorizing digital business models by digitisation penetration

The model is based on the finding that as the degree of digitisation increases, the relative position of the original physical products in relation to the digitally provided service share of the total value added decreases - i.e. the product significance decreases and the relevance to digitisation increases. In extreme cases, the physical product would then be obsolete, and the original (product) service can be provided purely service-based and digitally. In mixed forms, product and service act as 'service bundles' - digital services can represent value-added services complementary to a physical product, for example, or the product becomes a pure platform for the provision of services, as in the case of a smartphone, but is increasingly relegated to the background. With the help of the penetration degree model, it is possible to classify companies and the associated business model according to the degree of digitisation or to gain orientation when developing a new business model. Experience has shown, however, that the complete digitisation of the business model is diametrically opposed to the continued interest of every producing company and should certainly not be pursued as a self-purpose for the sake of digitisation. In this respect, the 'ideal' degree of digitisation must be determined methodically. In addition to the (12) identified concepts for digital business models in the

Industry 4.0 context, repetitive archetypes could also be identified, such as those described by the Business Model Innovation Lab of the University of St. Gallen. For example, the St. Gallen archetype "Leverage customer data" is a regular core element of the Industry 4.0 concept "Big Data", the archetypes "Guaranteed availability" and "Pay-per-use" are an essential component of the "Everything-as-a-Service" concept. The archetypes "user designed" and "mass customization" can be found in the Industry 4.0 concept "individualization/ personalization". The business model archetypes "crowdsourcing" and "crowdfunding", on the other hand, are components of examined business models that follow the "Community & Open Innovation" concept. This suggests that digital business models do not necessarily differ conceptually from conventional service-oriented business models, but only that the performance is digital.

3.3 Self-Assessment Tool

In order to reduce obstacles for companies in the analysis of their own potentials with regard to the digitisation of their business model and to offer concrete support in the design of business model options, two concrete software tools (Self-Assessment Tool and Risk Management) were developed, validated and evaluated in this project step. This paper deals in detail with the Self-Assessment Tool.

As described in 3.1, a total of (12) relevant Industry 4.0 concepts relevant to the business model could be identified at the time of the study. The aim of the self-assessment tool is to determine the potential suitability of these concepts for the user (or his company) through a structured questioning process in order to develop and implement new digital business models based on them. The first of the two approaches for business model development outlined in the technology cascade (see Figure 1) was used. The aim here is not to configurate and deliver to the user a comprehensive proposal for the digitisation of the business model, but rather to give thematically oriented impulses, hints and food for thought for their own further intensification with the respective Industry 4.0 concept.

For this purpose, criteria have been developed for each Industry 4.0 concept, on the basis of which the suitability can be tested, as well as a question process, comprising a total of (27) questions, which determines whether these criteria apply to it. After its conception in the .NET framework, the tool was implemented by a software service provider (see figure 3).

The question logic covers the question categories: company, enabler technologies as well as customer relationship, utilization concept and service. Building on this, the correlation between answer characteristics and potential suitability of the individual Industry 4.0 concepts is then implemented, which ultimately indicates the concepts depending on the answers given by the user.

During the development and design of the Self-Assessment Tool, the main focus was on an independent, i.e. intuitive, usability and at the same time universal validity of the tool.

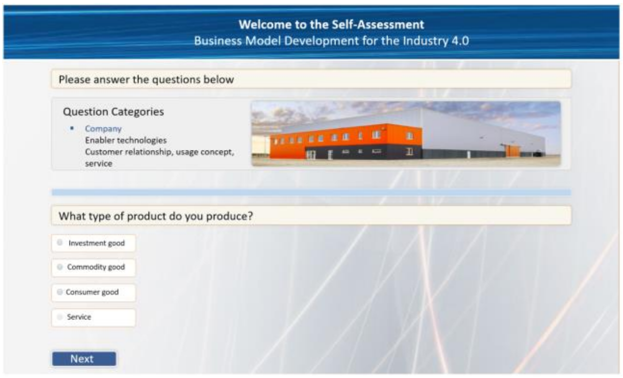


Figure 3 - Self-Assessment Tool

As a result, after completion of the self-assessment tool (which is available on the online platform), the user will also get optically highlighted those Industry 4.0 concepts, which are most interesting for him on the basis of the questions answered. These highlighted concepts of the Self-Assessment Tool are directly linked to the online platform (3.4) of the project (see figure 4). The online platform provides in-depth information on each of the Industry 4.0 concepts and suitable good practice examples from industry. The user thus has the opportunity to expand his knowledge with regard to the proposed Industry 4.0 concepts and receives further impulses for a possible digitisation of the business model.

The tool could be extensively tested during networking events as well as in the context of company coaching. On the basis of the test results and experiences, the self-assessment tool was optimized with regard to question logic and optics.

3.4 Online-Platform

The online platform is based on a comprehensive concept, which was developed to enable companies to collaborate situationally with all members or participants of their individual "ecosystem". So-called communities can be formed for this purpose, which can be staffed within or across companies. Individual communities can be merged and also separated again, for example to interact briefly for idea generation processes. The functionality of a community follows the idea of a smartphone, on which "apps" can be installed and later removed as required. These apps can be used for data exchange, discussion, brainstorming, appointment coordination or media visualization.

The basic technical framework and the provision are provided by INNO-FOCUS Business Consulting GmbH. With the help of the platform, innovation projects can be handled efficiently, the people involved can be networked and informed, and knowledge can be shared and multiplied.

The online platform (see Figure 4) thus also acts as a link between self-assessment and the risk management tool by making it possible to document initial ideas for new business models resulting from the use of the self-assessment tool and to further detail them, including the relevant stakeholders, until they can be modelled with their attributes in the risk management tool. Essential functions of the platform include forums, document management, survey tools, event planning, calendar, ideas.

Figure 4 - Platform (www.digitale-geschaeftsmodelle-bw.de)

Furthermore, good practices are available to the users of the platform (30), through which initial impulses for ideas for business model digitisation are given and are linked with the self-assessment tool. In addition, the self-assessment tool and the risk management tool were integrated into the start page after log-in and are therefore freely accessible to all users of the online platform. Furthermore, a separate area was created f*or each of the (13*) identified Industry 4.0 concepts, in which specific information and documents for knowledge transfer are stored.

The use of the platform has shown that it makes a functional contribution to the solution of subtasks along a business model development process, but cannot completely substitute the immersive and creative processes that take place in the context of idea-finding workshops, interdisciplinary discussions and haptic prototyping. In this respect, the platform represents a multifunctional tool for situational project support and documentation in addition to real development processes.

3.5 Holistic approach

Through the networking and testing events, it became clear that the individual elements developed, such as the cascade, the GENI scheme, the tools and the platform can only fully develop their potential if they are embedded in a holistic approach.

In the holistic approach, the developed elements were combined with the approach of Design Thinking and the Business Canvas model.

In the first step, users (companies) can intuitively assess the individual potential of a company using the self-assessment tool for their company. To support the further idea development process, the online platform with industry-specific success stories and concept-specific information is available to the users.

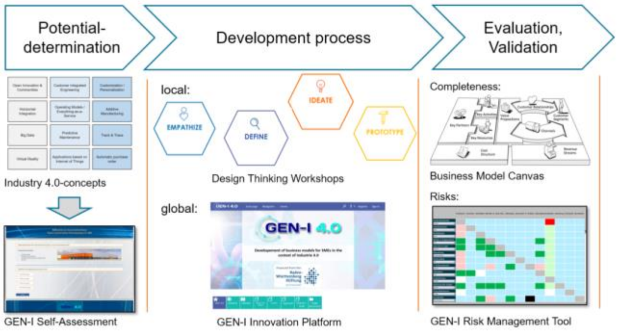


Figure 5 - Relationship between Partial Solutions and the GEN-I Approach

For the innovation project to digitize the business model, companies can enter into situational collaboration with all members or participants of their individual "ecosystem" by creating specific communities. For the creative, workshop-oriented development of the business model, however, the project partners are also available to the companies at the same time. For the evaluation and risk minimization of the designed business model the risk management tool is used and the completeness is checked workshop-oriented with the help of the Business Model Canvas.

In particular the testing companies were able to convince themselves of the conceptual consistency of the holistic approach model and were surprised by the degree of innovation and the speed of the ideas for sustainable business models developed along the development process.

4 CONCLUSIONS

The developed individual tools as well as the manifold findings and research results represent the components of a holistic and in this form a new approach for Business Model Development in Context of Industry 4.0. The results of the research could be made clear and experienced in different intensities for a multitude of companies and interested parties. In addition, it became apparent that the elements cannot be applied to any company, but that they provide suitable results, especially for companies with physical products or physical production - which, however, also corresponds to the target group of the research work in the GEN-I4.0 project.

The dynamics of the thematic in the field of research is also shown by the fact that additional new technologies, such as crypto currencies, machine learning and artificial intelligence, were at times also strongly discussed with regard to their impact on business models. This highlights the need for continuous adaptation of the results achieved.

5 OUTLOOK

As described in Chapter 3.5, the developed approach will be used in the further year 2018 as well as in 2019 in additional events within the transfer activities of the »Reutlinger Zentrum Industrie 4.0« under the motto "From the region for the future" for further validation and optimization of the research results. New technologies with an influence on digital business models will be analysed and integrated into the individual elements. In this way, the achievement of the overriding goals of the applied research project GEN-I4.0, i.e. to offer SMEs effective support in business model development and at the same time to create awareness of the importance of digital business models, can be sustained.

6 ACKNOWLEDGEMENTS

We acknowledge the cooperation of the Fraunhofer IAO, Stuttgart Germany and would like to express our gratitude for the support of the Baden-Württemberg Stiftung gGmbH in enabling this research work.

7 REFERENCES

[1] Lucke, D., Görzig, D., Kacir, M., Volkmann, J., Haist, C., Sachsenmaier, M., Rentschler, H., 2014, Strukturstudie “Industrie 4.0 für Baden-Württemberg” - Baden-Württemberg auf dem Weg zu Industrie 4.0, Ministerium für Finanzen und Wirtschaft Baden-Württemberg and Fraunhofer-Institut für Produktionstechnik und Automatisierung IPA (Eds.), [online] Available: www.mfw.baden-wuerttemberg.de [Accessed 12 Sep. 2018]

[2] Geissbauer, R., Schrauf, S., Koch, V., Kuge, S., 2014, Industrie 4.0 - Chancen und Herausforderungen der vierten industriellen Revolution, PwC and Strategy& (Eds.), [online] Available: https://www.strategyand.pwc.com/media/file/Industrie-4-0.pdf [Accessed 20 Aug. 2018]

[3] Zanker, C., Som, O., Buschak, D., 2014, Industrieller Mittelstand: Spitzenstellung in Gefahr? – Analyse zur Innovationsfähigkeit kleiner und mittlerer Unternehmen in der Metropolregion Stuttgart, Industrie- und Handelskammer Region Stuttgart (Ed.), [online] Available: https://www.stuttgart.ihk24.de/Fuer-Unternehmen/innovation/Aktuelles/Studie_Innovationsfaehigkeit/664670 [Accessed 22 Aug. 2018]

[4] Heger, R., Kelkar, O., Dao. D.-K., 2014, Studie Industrie 4.0 - Eine Standortbestimmung der Automobil- und Fertigungsindustrie, MHP - A Porsche Company (Ed.), [online] Available: http://www.mhp.com/de/studien/ [Accessed 20 Aug. 2018]

[5] Bundesministerium für Wirtschaft und Energie, AUTONOMIK für Industrie 4.0, 2016, [online]

[6] Heinrich, C., Gärtner, C., Lopper, E., Bolesta, W., 2018, Chief Digital Officer Studie 2018 - Wer steuert die Umsetzung der Digitalstrategie - und wie?, Quadriga Hochschule Berlin GmbH (Ed.), [online] Available: https://www.quadriga-hochschule.com/sites/default/files/cdo_2018_ansicht_12.pdf [Accessed 21 Aug. 2018]

[7] Plattform Industrie 4.0: Impulspapier „Innovationstreiber für digitale Geschäfts-modelle“, Juli 2018.

[8] Riemensperger, F., Pfannes, P., Bongardt, S., 2017, Deutschlands Top 500 - Digitale Geschäftsmodelle ohne Geschäft?, Accenture (Ed.), [online] Available: https://www.accenture.com/de-de/top500-2018 [Accessed 20 Aug. 2018] pp.6

[9] Kreimeier, N., 2018, Digitalchefs zeigen wenig Interesse an KI-Projekten, Capital (Ed.), [online] Available: https://www.capital.de/wirtschaft-politik/digitalchefs-zeigen-wenig-interesse-an-ki-projekten [Accessed 21 Aug 2018]

8 BIOGRAPHY

Vera Hummel, Prof. Dr.-Ing., Dipl.-Ing.(FH), has been a professor at the ESB Business School, Reutlingen University since 2010. She is a founding member of the “International Association on Learning Factories IALF” and the initiator and head of the "ESB Logistics Learning Factory“ on the campus of Reutlingen University. Head of Reutlinger Zentrum für Industrie 4.0 (RZI 4.0).

Marc Guldin (BSc.) is a research associate at ESB Business School and part of the research, training and consultant group Reutlinger Zentrum für Industrie 4.0 (RZI 4.0). His major field of research is the design and implementation of digital business models as well as the know-how transfer and development in companies for the implementation of Industry 4.0 solutions.

Internet Platform for the Implementation of Industry 4.0 in SMEs

C.Cevirgen[1], D.Herberger[1], S.Schenk[2], P.Nyhuis[1]
[1]Institute of Production Systems and Logistics, Leibniz University Hanover, Germany
[2]GREAN GmbH, Hanover, Germany

Abstract
Manufacturing enterprises are faced with major technological changes due to Industry 4.0, which influence the whole competition. New information and communication technologies link enterprises, people, machines and are creating new potential in the manufacturing sector. Large enterprises have the opportunity to entrust own divisions, who efficiently introduce new technologies into their production operations. However, small- and medium-sized enterprises (SMEs) are limited in their ability to introduce Industry 4.0 because of their financial and human capacities. As SMEs cover a significant part of the productive economy, it is imperative that they do not lose touch with these industrial changes despite their difficult starting position. Existing approaches to support the implementation of Industry 4.0 in SMEs rely on consulting services or software tools, which are often cost-intensive and generic in nature. These approaches are only partially applicable for SMEs, since external consultants and generic software approaches do not fully take the individual development status of SMEs into account. In order to meet current and future market requirements for the production and logistics processes of SMEs, new approaches must be developed that support each SME in the implementation of Industry 4.0 according to its individual possibilities and objectives and therefore close the gap in existing approaches. An internet-based platform, in which the user gets information and explanation of measures for a sustainable production, can support the implementation of Industry 4.0. Particularly decisive for the success of a platform is the didactic preparation of the platform content, so that no prior knowledge is required from the user and that the effort of using the platform kept to a minimum. This article gives a brief summary of the structured procedure for creating a project platform that closes the research gap. Based on the scientific approach, an internet-based demonstrator is going to be developed that will individually assist SMEs in the implementation of Industry 4.0 in the future.

Keywords
SMEs, Industry 4.0, internet platform, maturity level model, information and communication technologies

1 INTRODUCTION AND NEED FOR RESEARCH

The implementation of Industry 4.0 requires not only the development of new technologies, but also the efficient and targeted introduction into the enterprises. For example this includes new information and communication technologies (such as decentralised networked sensor technology), which provide information in real time and are seen as a basic prerequisite for Industry 4.0 [1]. The integration of such technologies enables the development of innovative business models as well as significant optimisation potentials in production and logistics [2, 3]. Due to the limited financial and personnel resources small and medium-sized enterprises (SMEs) face an immense challenge when implementing Industry 4.0. For this reason, the majority of SMEs are still approaching Industry 4.0 very cautiously [3–6]. So far, only about five percent of German SMEs are comprehensively networked and only one in three SMEs deals with Industry 4.0 [5]. It is obvious that the spread of Industry 4.0 depends on the size of the enterprise. In addition, other factors such as the sector or the enterprise strategy are also decisive, which is why all enterprises have different levels of maturity (development status) regarding to digitally supported production [1, 7, 8]. Therefore, each enterprise has its own framework conditions, which is to consider when successfully implementing Industry 4.0. As a result, it should be noted that the implementation of a high-end solution (e.g. fully networked, intelligent systems) is not meaningful or effective for all enterprises. In particular, SMEs can achieve a considerable increase in productivity with a small investment through the successively implementation of low-end solutions (cost-effective and easy-to-implement approaches).

As SMEs cover a significant proportion of the industrial sector [9], it is important to ensure in the future that they do not lose the connection to digitisation. Therefore, the Centre for European Economic Research recommends, that small and medium-sized enterprises need to be more aware of the possibilities offered by digitisation by using best practice examples [10]. Enterprises and especially SMEs need support that requires only minimal prior knowledge and little effort in use (low threshold) in order to advance the implementation of Industry 4.0. Table 1 gives an overview of some relevant Industry 4.0 implementation projects initiated by various programmes (as far as the results are known today). Twelve projects were examined with regard to their addressing of a cross-sector application, competence development, the conception and/or communication of introduction strategies, the provision of the results on a platform for self-application as well as the integration of enterprise-specific solution approaches.

This overview shows that no project exists yet which addresses all of the mentioned criteria. Furthermore, the existing projects are limited to the provision of information and support by external consultants, which are in conclusion cost-intensive or a too generic approach. This research gap shows that new approaches must be developed that support each SME in the implementation of Industry 4.0 according to its individual possibilities and objectives.

Subject / Project	Cross-sector application	Competence development	Introduction strategy	Platform for self-application	Enterprise-specific solutions
BMBF: IWEPRO	◑	○	◑	○	○
BMBF: BiMo, SE, TTvor	◑	◑	◑	○	○
BMBF: TA, **BMWi:** APPsist	○	◑	○	○	○
BMBF: SO, VorZug **BMWi:** GEMINI	◑	○	◑	○	○
BMWi: Plattform i40	◑	●	◑	◑	○

○ not addressed ◑ partly addressed ● core issue

Table 1 - Overview of relevant research projects [2, 11, 12]

2 IDEA TO CLOSE RESEARCH GAP

This article illustrates the development of a low-threshold, internet-based project platform to better promote the introduction and implementation of Industry 4.0 in SMEs. In this context, low threshold means that the users do not have to rely on prior knowledge, nor does it involve a great deal of effort to use the platform or to obtain relevant information. The basic preliminary considerations for a platform for the introduction and implementation of Industry 4.0 adapted to SMEs are illustrated in Figure 1.

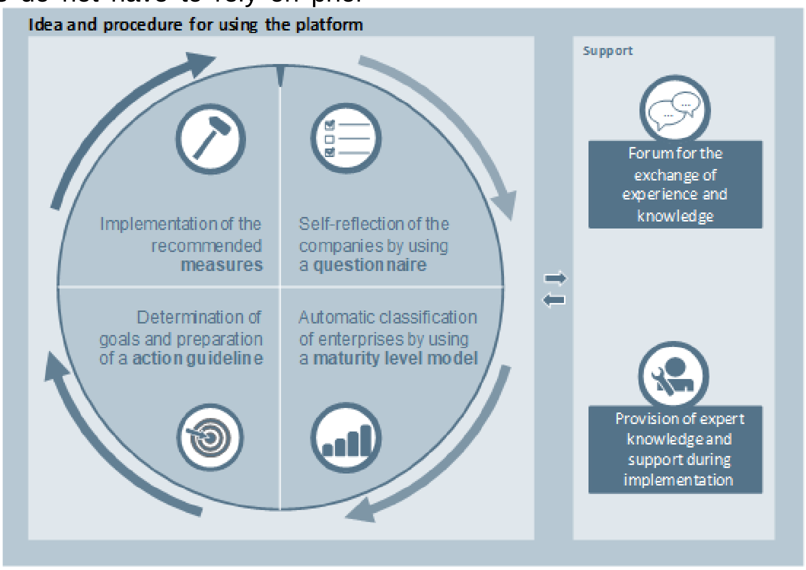


Figure 1 - Idea and procedure for using the plattform

In order to achieve the best possible adaptation to the needs of SMEs, the enterprises are guided first through a questionnaire to describe their general situation and to determine their degree of maturity on the platform (Figure 1). This includes, for example, the industry affiliation, the enterprise size, possible problems as well as the digitisation maturity in different areas of the enterprise. For the creation of the maturity level model, numerous preliminary works can be used [7, 8, 13], in order to develop a model adapted to the platform. The result of the maturity level determination, which is handed out automatically to the enterprises, represents a structured, comprehensible and descriptive overview of the actual situation. Based on this, SMEs are provided with an individual guideline for the targeted introduction and implementation of Industry 4.0 in their own enterprise. This contains promising measures for the improvement of the development conditions, for the enterprise relevant information about technologies, procedures and principles of Industry 4.0, selected methods and tools, various best practice examples, necessary milestones for the introduction, helpful video tutorials, linkage to suitable training courses and seminars as well as the collected experience knowledge of the platform users.

The last step of the process is the use of the knowledge to transfer the recommended measures into the enterprise. The implementation can be launched independently by the enterprise or by the use of a support area (see Figure 1). In this support area external consultants further increase the efficiency of the application through additional explanations and their experience. Following the implementation phase, a new actual state is reached, whereupon the iterative process can be run through again.

Generally understandable and comprehensible explanations of the questionnaire, the action guideline and the associated contents help to sensitise enterprises to the topic of Industry 4.0 and to point out relevant contexts (such as the necessity of lean production for the successful implementation of Industry 4.0) [14–16].

These explanations, supplemented with a clear and intuitive structure of the online presence, are essential for a low-threshold and simple use of the project platform by the employees of SMEs. Another benefit is the direct qualification of the employees, which plays a decisive role for the successful introduction and implementation of Industry 4.0 in the enterprise [17, 18]. Among other things, the difficulty lies in creating the necessary technological openness and allowing employees to participate in the introduction of approaches and solutions [19, 20]. The video tutorials offered on the platform support not only the detailed explanations of the methods but also the tools and contexts. A clear visualisation of the contents increases the acceptance of the employees and thus integrates them fully into the change process. A clear structure, a clear and appealing visualisation as well as clear and comprehensible explanations are necessary in order to fulfil the prerequisite of low threshold.

3 PROCEDURE MODEL OF THE INTERNET PLATFORM

In order to achieve the required results, six sub-steps are needed. Figure 2 visualises each step. In addition, the steps are described in detail below.

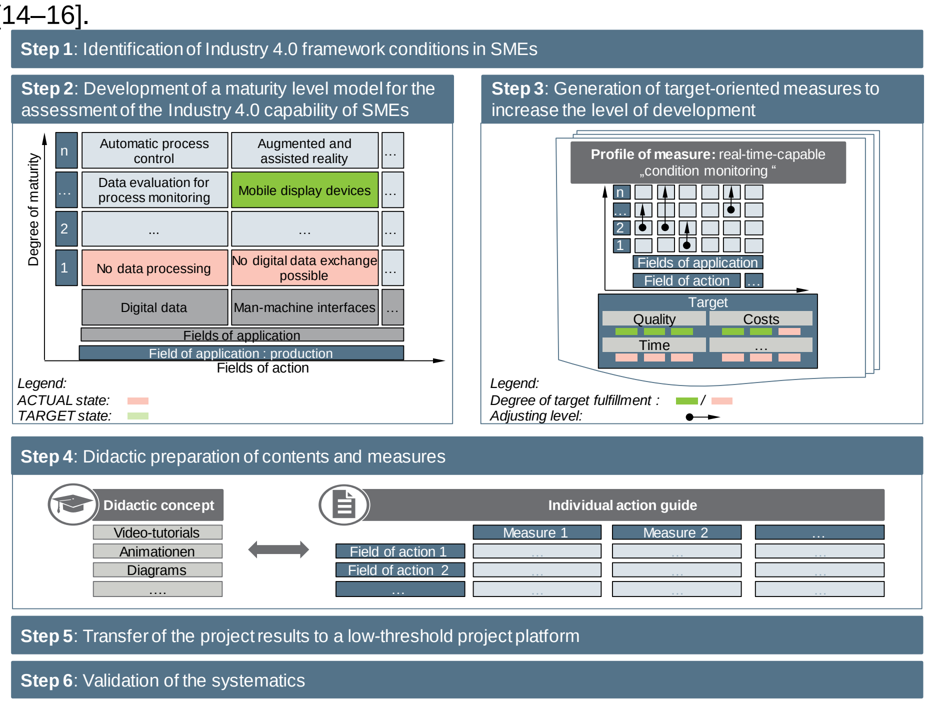


Figure 2 - Procedure model to reach the objective

In **step 1**, relevant requirements, applications, methods and tools of Industry 4.0 for SMEs are identified. In addition, the company areas influenced by Industry 4.0 as well as fields of action (e.g. production, logistics, business models) are detected. A status quo check is conducted on the current development status of Industry 4.0 in SMEs and on measures that are already implemented to increase the maturity level. The information required for this will be determined by expert interviews with actors from industry, by technology scouting and by literature research and prepared accordingly. The experience knowledge available from the project partners is used to supplement all of the information.

In **step 2**, the results from the first step are transferred into a maturity level model. This model classifies SMEs according to their stage of maturity (from "the basic prerequisite for the implementation of Industry 4.0 is not given in this area" to "the SME in this area is at the current state of research regarding Industry 4.0"). For this purpose, it is necessary to identify requirements for this maturity level model, to check the validity of existing maturity level models and to develop a maturity level model based on these findings and oriented towards the new platform. Furthermore, a questionnaire will be developed, which is capable of reflecting the actual situation of SMEs in the maturity model.

The measures identified in the first step, which have already been implemented to increase the level of development in SMEs, are used as a basis for achieving the objective in **step 3**. By critically questioning the procedures, the measures are specified and reflected on defined maturity levels. Missing measures are developed by the project partners, described comprehensibly and also assigned to the maturity levels. Subsequently, the measures as well as the related fields of action and maturity levels are analysed with regard to their interactions. The results are then described and clearly visualised (e.g. in cause-effect diagrams).

In **step 4**, the identified, relevant contents for the project platform (step 1), such as practical examples, methods, tools and principles of Industry 4.0, are explained and visualised. In video tutorials, animations and diagrams, the contents are made accessible to users in a simple, clear and comprehensible way. It is important, that neither prior knowledge nor a large amount of effort is necessary in order to make the collected knowledge accessible on the platform. The measures collected in step 3, their benefits and the steps necessary to implement them are prepared didactically in order to integrate them in a clear way into the platform. Subsequently, a framework for action guidelines will be developed, which, will be filled company-specifically and distributed to the platform users by evaluating the potentials in the maturity model (step 2) and the combination of associated measures (step 3).

In **step 5**, a content empty, structured and clear layout for the platform is first designed, which should enable intuitive use. This layout can then be filled with the generated contents of the first four steps. The necessary links between the questionnaire, the maturity model, the associated measures and the framework of the action guide are created so that the user is automatically provided with a company-specific action guide. The contents on the platform are supplemented by a support area for consultants from science and/or practice as well as various forums for the users. In order to ensure applicability and practicability, IT experts may be involved to support the project partners.

Finally, a general system test is carried out in **step 6**. This ensures that the defined service components in the platform run without errors and that they can be used smoothly by both users and system administrators. Subsequently, all contents and functions are checked for completeness and conformity. Internal case studies support the identification and elimination of errors. Finally, external case studies are conducted with practitioners to simulate as many different use cases as possible. Identified errors are corrected iteratively in all validation steps. At the end of the procedure, the user release takes place.

4 SUMMARY AND OUTLOOK

Due to personnel and financial constraints, SMEs face an immense challenge in taking their first steps towards Industry 4.0 on their own. In order to meet current and future market requirements, it is essential for SMEs to participate in the industrial change process. For this reason, the presented procedure fo an internet-based project platform has been developed to promote the introduction and implementation of Industry 4.0. Neither specific prior knowledge nor a great deal of effort is required to use the platform.

In principle, the elaborated internet-related project platform can represent a solution for a successful implementation of Industry 4.0 in SMEs. The entire procedure is being developed at the Institute of Production Systems and Logistics at Leibniz University Hanover. Currently, the project is in the early development phase. First results regarding the technology scouting and the maturity model are available and have to be coordinated with the appropriate improvement measures. First concepts for the didactic preparation have already been discussed. This preliminary work still has to be compiled into an overall concept. Finally, the results must be validated in close cooperation with industrial partners and converted into an internet-based demonstrator.

5 REFERENCES

[1] McKinsey Digital (Hrsg.): Industry 4.0 – How to navigate digitization of the manufacturing sector, 2015

[2] BUNDESMINISTERIUM FÜR BILDUNG UND FORSCHUNG: Industrie 4.0 – Innovationen für die Produktion von morgen. Bonn, 2015

[3] ROTH, A. (Hrsg.): Einführung und Umsetzung von Industrie 4.0: Grundlagen, Vorgehensmodell und Use Cases aus der Praxis. 1. Aufl. 2016. Berlin, Heidelberg: Springer Berlin Heidelberg, 2016

[4] BUNDESMINISTERIUM FÜR BILDUNG UND FORSCHUNG (BMBF) (Hrsg.): Die neue Hightech-Strategie: Innovationen für Deutschland. Stand: August 2014. Berlin, 2014 (Die neue Hightech-Strategie)

[5] SCHRÖDER, A.: The challenges of Industry 4.0 for small and medium-sized enterprises. Bonn, 2016

[6] SOMMER, L.: Industrial Revolution - Industry 4.0: Are German Manufacturing SMEs the First Victims of this Revolution, 2015

[7] LICHTBLAU K.; STICH V.; BERTENRATH R.; BLUM M.; BLEIDER M.; MILLACK A.; SCHMITT K.; SCHMITZ E.; SCHRÖTER M.: Industrie 4.0 - Readiness. Aachen, Köln, 2015

[8] VDMA, KIT KARLSRUHE (WBK), TU DARMSTADT: Leitfaden Industrie 4.0 - Orientierungshilfe zur Einführung in den Mittelstand. Frankfurt am Main: VDMA Verlag GmbH, 2015

[9] SMIT, J.; KREUTZER, S.; MÖLLER, C.; CARLBERG, M.: Industry 4.0. Brussels, 2016

[10] SAAM, M.; VIETE, S.; SCHIEL, S.: Digitalisierung im Mittelstand: Status Quo, aktuelle Entwicklungen und Herausforderungen: Forschungsprojekt im Auftrag der KfW Bankengruppe. Mannheim, 2016 (acatech STUDIE)

[11] BUNDESMINISTERIUM FÜR WIRTSCHAFT UND ENERGIE: Autonomik für Industrie 4.0. Berlin, 2015

[12] BUNDESMINISTERIUM FÜR WIRTSCHAFT UND ENERGIE: Internetpräsenz der Plattform Industrie 4.0. URL https://www.plattform-i40.de

[13] SCHUH, G.; ANDERL, R.; GAUSEMEIER, J.; TEN HOMPEL, M.; WAHLSTER, W.(Hrsg.): Industrie 4.0 Maturity Index: Die digitale Transformation von Unternehmen gestalten. München: Herbert Utz Verlag, 2017 (acatech STUDIE)

[14] BICK, W.: Warum Industrie 4.0 und Lean zwingend zusammengehören. In: VDI-Z 156 (2014), Nr. 11, S. 46–47

[15] KASPAR, S.; SCHNEIDER, M.: Lean und Industrie 4.0 in der Intralogistik. In: Productivity 20 (2015), S. 17–20

[16] SANDERS, A.; ELANGESWARAN, C; WULFSBERG, J.: Industry 4.0 implies lean manufacturing: Research activities in industry 4.0 function as enablers for lean manufacturing. In: Journal of Industrial Engineering and Management 9 (2016), Nr. 3, S. 811–833 – Überprüfungsdatum 2018-09-15

[17] DOMBROWSKI, U., WAGNER, T.: Arbeitsbedingungen im Wandel der Industrie 4.0 – Mitarbeiterpartizipation als Erfolgsfaktor zur Akzeptanzbildung und Kompetenzentwicklung. In: ZWF (2014), Nr. 5, S. 351–355

[18] LUDWIG, T.; KOTTHAUS, C.; STEIN, M.; DURT, H.; KURZ, C.; WENZ, J.; DOUBLET, T.; BECKER, M.; PIPEK, V.; WULF, V.: Arbeiten im Mittelstand 4.0: KMU im Spannungsfeld des digitalen Wandels In: HMD: Praxis der Wirtschaftsinformatik 53 (2016), Nr. 307, S. 71–86

[19] FORSCHUNGSUNION WISSENSCHAFT – WIRTSCHAFT: Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0: Abschlußbericht des Arbeitskreises Industrie 4.0. Essen: Forschungsunion Wirtschaft - Wissenschaft, 2013

[20] NYHUIS, P.; KLEMKE, T.; WAGNER, C.: Wandlungsfähigkeit: Ein systemischer Ansatz. In: Wandlungsfähige Produktionssysteme. Berlin : GITO-Verl., 2010, S. 3–21

6 BIOGRAPHY

Cihan Cevirgen (*1990) has been a research associate in the specialist factory planning group at the Institute of Production Systems and Logistics (IFA) at Leibniz University Hanover since 2018. He previously studied industrial engineering (B.Sc., M.Sc.) at the Leibniz University of Hanover (LUH).

David Herberger (*1990) has been a research associate in the specialist factory planning group at the Institute of Production Systems and Logistics (IFA) at Leibniz University Hanover since 2016. He previously studied mechanical engineering (B.Sc., M.Sc.) at the Karlsruhe Institute of Technology (KIT).

Sebastian Schenk (*1986) has been a factory planning consultant at GREAN GmbH since 2016. He previously studied industrial engineering (B.Sc., M.Sc.) at the Braunschweig University of Technology.

Peter Nyhuis (*1957) has been head of the Institute of Production Systems and Logistics (IFA) at Leibniz University Hanover since 2003. Prof. Dr.-Ing. habil. Peter Nyhuis has been a member of the German Science Council and Chairman of the Science Council of the AiF (GAG 3) since 2015. He is also a member of the German Academy of Technical Sciences acatech and the Scientific Society for Production Technology (WGP)

Industry 4.0 and Digitalisation of Production Systems – How Remote Control of Robots and other Mechatronic Systems Can Contribute

G. Gruhler[1], J.Hermann[2]
[1]Robotics and Telematics Laboratory, Reutlingen University, Germany
[2]ESB Business School, Reutlingen University, Germany

Abstract
It is expected that ongoing digitalisation will drive the merger between the manufacturing world and the internet world, possibly leading to a next industrial revolution, currently called "Industry 4.0". The driving forces behind this development are new business opportunities and competition advantages arising from mass production customisation as well as rapid individual product development and manufacturing. Key factors of the development towards Industry 4.0 are discussed. Threats and opportunities arising from these developments for future production are discussed. Actual examples from real-time customized manufacturing of consumer products are given. As mechatronic systems and industrial robots are widely used in manufacturing and in particular in assembly, it is discussed how they can be connected to and used in digitalised industrial systems. Different examples of remote controlled systems are presented, like remote controlled KUKA robot for handling and quality control, PLC-controlled equipment, drive systems, FESTO handling system and others. The architecture of an assembly cell is presented, where industrial robots are set-up for batch-one production or can directly receive control / production information on-line and in real-time over the factory network. Methods for remote maintenance and monitoring of systems over the internet and production operator support over the internet are presented as well.

Keywords
Digitalisation of production, reconfigurable manufacturing systems, remote control

1 INTRODUCTION

The term 'Industry 4.0' was generated with the intention to describe the possible 4th industrial revolution which might be implemented by the 'digitalisation' of a wide range of industrial and production processes. The term digitalisation itself is not a very fortunate wording since e.g. digital control, data processing and even computer controlled manufacturing is well-established since many decades. However, the term has become widespread and is obviously used to describe the introduction and application of advanced digital methods in the industrial context. Networking plays a major role in digitalisation, therefore the merger of the internet world with the industrial world is considered to be one important part of the new developments.

2 KEY ELEMENTS, OPPORTUNITIES AND RISKS ARISING FROM DIGITALISATION OF PRODUCTION

2.1 Digitalisation of production – some key elements

It is still difficult to give an exact definition of Industry 4.0 and digitalisation of production. It rather makes sense to summarize some characteristic elements, which include [1], [2]:

- Merger of the Internet world and the production world.
- New methods of human-system-interaction, including online services more or less directly linked to production.
- Rapid connection of and rapid communication between (embedded) components, systems, users, including in particular physical components and systems.
- Distributed intelligence and, up to a certain extent, autonomous behavior of subsystems.
- Digitalisation throughout the complete supply chain.

Digitalisation of production can make use of Cyber Physical Systems (CPS), software representation of physical products and systems (digital twins), collecting and processing of big amount of data (big data), Internet of Things (IoT) up to Internet of Everything (IoE).

2.2 Opportunities, consequences and risks

One major driver behind digitalisation in industry are many new business opportunities, which are expected and might be sparked by the commercial success of online companies in the consumer sector such as Amazon.

2.2.1 *Mass product customisation / rapid individual product development and manufacturing*

Expected benefits of digitalisation in industry are competition advantages by individualisation of products at short notice and at costs comparable or not tremendously higher compared to mass products. Additionally, some customers may even be prepared to spend extra money if they get individualised products (see example of sports shoe customization below). Individualisation of production is either based on further customisation of mass products (development and manufacturing of those) or will make use of new rapid development methods and new rapid production methods (like e.g. Additive Manufacturing equipment, 3D printers).

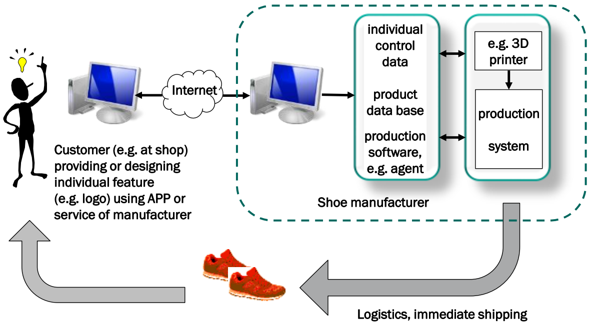


Figure 1 – Application scenario from consumer products: Individualised sports shoe production

Figure 1 shows an example of such a procedure from sport shoe production: Design of individual features, like colour, logo or print on the shoes during the customer's visit and probing of the shoe in the shoe shop (or if one wants to buy with the risk of not-fitting shoes) at the own personal end device (PC, tablet, smart phone) using internet-based software of the shoe manufacturer. Immediate production of the individualised pair of shoes at the production plant and fast shipping by courier to the customer.

2.2.2 *Rapid development and highly reconfigurable manufacturing and assembly equipment*

It becomes obvious that customisation or individual product manufacturing requires rapid development methods as well as a new range of flexibility of production systems.

Rapid development can either be achieved by selection from pre-defined options (basically as it is done during the ordering process of a new car), or by integration of customer's elements or at least customisable elements (like a logo on a product or the size of a garment, or the size of furniture, or whatever other feature).

With respect to production systems, unfortunately the basic relation between flexibility and costs is that flexibility tends to be expensive. One reason is that highly flexible systems are much more complex than mass production systems. There are a lot of examples: mass production systems are highly specialised and tuned to very low cycle time. On the other hand flexible robots, e.g. with vision systems, tend to be more expensive and slower in cycle time. The key to success in future will be to reduce production costs of individualised products. One appropriate measure can be to reduce the time for system reconfiguration or even for new set-up of production systems (see 'factory-in-a-day' project [3]). Looking to the assembly section of production systems, there are a number of components which offer a new range of flexibility. One example for highly reconfigurable equipment is the BOSCH APAS robot and manufacturing system. APAS consists of a mobile robot arm which can easily and rapidly be placed e.g. at an assembly spot and which allows fast and user friendly programming.

Since humans still provide a maximum of flexibility in production, the focus on human-machine interaction and collaboration becomes once more very important. The number of robots designed for human-robot collaborations was highly increasing over the last years. Examples are the Universal Robot types, KUKA iiwaa, ABB YuMi.

Even the previously mentioned APAS robot is covered by a sensor skin which protects both human and robot in case of collisions. The system forms therefore a step also in the direction of flexible human-robot collaboration.

2.2.3 *Risks and crucial factors for success of digitalisation*

It becomes obvious that Industry 4.0 leads to considerably increased complexity. Additionally to complexity, the following issues seem to be major risks and are often hurdles of digitalisation in production:

- Data security,
- physical system safety,
- customer acceptance,
- human qualifcation issues,
- investments and costs.

seem to be major risks and are not seldom hurdles of digitalisation in production.

Consequently, key issues for the (technical **and** commercial) success might be:

- Data and communication protection,
- protection of physical production equipment and systems against damage caused by unauthorized or incorrectly generated procedures,
- standardisation (e.g. of reference models, interfacing etc.),
- reduction of complexity and implementation of convenient functions for users, development of easy-to-use online services.

The implemented examples of digitalisation detailed in the following sections include also some examples how to overcome one or the other of these hurdles.

3 INCREASING AND DRIVING FLEXIBILITY OF PRODUCTION: METHODS FOR AND OF DIGITALISATION

3.1 Production equipment and the Internet

Figure 2 shows the general approach of human-machine interaction over the internet. This was implemented in one of our laboratories. More than 10 physical systems, including industrial robots, handling systems, drives, PLC-controlled and other equipment can be accessed, observed by interactive web-cams and even controlled over the internet [4].

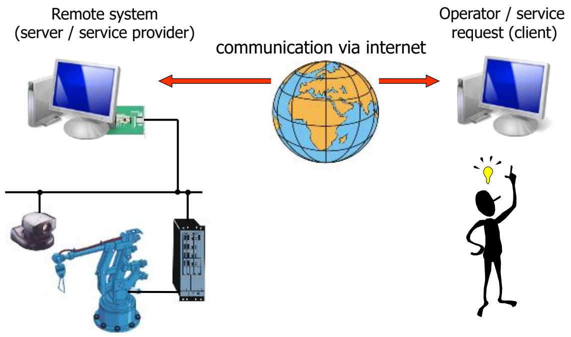


Figure 2 – General scenario of implemented internet access to production equipment

Here, we are in particular interested in some of the described key issues of digitalisation, namely

- the protection of physical production equipment and systems against damage caused by unauthorised or incorrectly generated procedures,

as well as

- the reduction of complexity and implementation of convenient functions for users and development of easy-to-use online services.

Therefore the implemention allows access

- from everywhere,
- at any time,
- by anybody.

In particular, to allow access by anybody is the major challenge with respect to the physical systems' safety, because this includes no protection by password or the like, and allows acces for completely unkown users.

To learn more about system protection, the method we apply are different access levels for the users, depending on the features of the different systems. In one example (handling system) we allow full graphical and text-oriented user programming via the internet because in this case there is no danger of mechanical collision with the environment.

The remote control of a KUKA industrial robot is done by a number of predefined actions for workpeace handling, measurement procedure etc. Sensor and status data generated during operations are transferred to and displayed on the remote client's system.

Closed-loop controlled systems can be tuned by changing control parameters and the influence of user-generated disturbance can be observed.

The technical implementation of the remote access to a few of the mentioned systems and devices is explained in section 5. More implementation details as well as actual remote control access are provided via [4].

3.2 How product configuration and reconfiguration of assembly systems can be supported by digitalisation

Digitalisation can help to manage the mass customisation of products. One of the enablers for mass customisation is 'customer-driven design and manufacturing' [5]. A digital instrument for this are online product configurators. These are web-based software applications for designing products that are precisely tailored to the individual needs of customers. A product configurator makes it possible to choose between product features, product options and technically feasible combinations. In mass customisation, this is the necessary link between customer-specific production and mass production. The configurator allows the customer to design the product to a certain degree by himself.

From the customer's point of view, the following advantages arise [6][7]:

- **Optimum fulfilment of requirements,** because the customer composes the product according to his specifications.
- **Avoiding of misconfigurations,** for products with many variants, a set of rules can prohibit incorrect selection options.
- **Reduced delay time** as no quotes have to be created, the price can be displayed during the configuration process.

The company can also gain advantages such as [6][7]:

- **Greater customer loyalty** - individual configuration increases emotional connection to the product.
- **Cost reduction** as no or fewer salespersons are required.
- **Amount of tied-up capital is reduced** as many combinations do not have to be pre-produced.

Flexible and reconfigurable manufacturing systems (RMS) are used to enable the production of individual products. These are necessary to react to changes on the market [8]. The six core features of a RMS defined by [9] are:

1. Customisation (flexibility limited to part family).
2. Convertibility (design for functionality changes).
3. Scalability (design for capacity changes).
4. Modularity (components are modular).
5. Integrability (interfaces for rapid integration).
6. Diagnosability (design for easy diagnostics).

In order to be able to fulfill these points, networkable systems, which have access to a company-wide interconnected information network, are necessary.

4 IMPLEMENTED EXAMPLE FOR HIGHLY FELXIBLE PRODUCT CONFIGURATION AND ASSEMBLY

4.1 Concept for customising and assembling a multivariant product

4.1.1 Customisable mass product

The product used is a city scooter with some interchangeable parts. E.g. different colored parts can influence the appearance of the scooter. It is also possible to adapt the product to customer requirements by selecting different wheels or add optional accessories like a mobile phone holder. To get an overview of the possible variants, a tree structure of the scooter product family was created as shown in figure 3 using the presentation scheme of [10]. The total number of possible product variants results from the number of feature variants of each functional requirement. With our selection options, the total number of variants is 256.

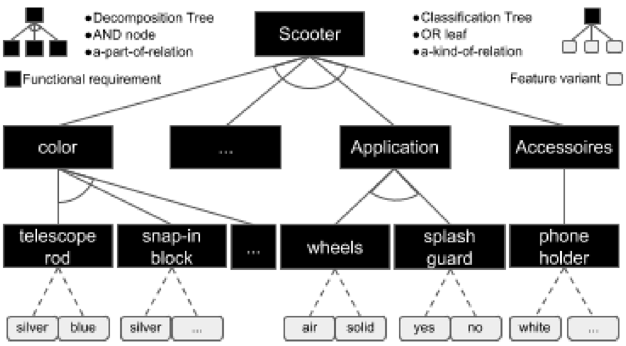


Figure 3 – Excerpt from the Product Family tree of the scooter

4.1.2 Front-end, backend and Database

Because JavaScript (JS) is one of the most popular programming languages [11] of the last years and almost the only one in the Front-end area [12], we decided to use the MEAN stack. This consists of the following components:

- **M**ongoDB - NoSQL Database which stores the data in JSON (JavaScript Object Notation) like format.
- **E**xpress.js - Server-side JavaScript Framework for web applications running on the Node.js platform.
- **A**ngular - TypeScript (superset of JavaScript) based Front-end framework for web applications.
- **N**ode.js - An implementation of a JavaScript engine that interprets the JS-code and converts it into machine-readable code.

This combination offers some advantages regarding the flexibility of the system. The use of NoSQL databases facilitates horizontal scalability, e.g. the enlargement of the product range. By using JavaScript in back- and Front-end, isomorphic and interchangeable code can be written. The JSON data format is transmitted through the entire communication path. Angular provides the platform to create a rudimentary product configurator.

4.1.3 Assembly process

This process is handled simultaneously by two different workstations: A collaborative robot that handles the heavy and bulky parts. And a pick by light system where one worker takes the remaining parts and places them on the assembly jig. Each picking order is triggered by scanning the RFID tag on the assembly jig.

4.1.4 Devices and systems

The collaborative robot is a UR10 type from Universal Robots. It is equipped with an electric 2-finger-gripper. Simple commands can be sent to the robot via the so-called dashboard server (TCP interface). This allows to remotely control the robot and to diagnose the condition.

The Pick by Light system is from the company Wibond and is comprised of a controller to which displays are connected in a line structure. The orders are managed by the Wibond software, which runs on a separate PC.

To assign the orders to RFID-chips we have developed an RFID reader. It is connected to the network via WiFi and it reads the ID and sends it to an MQTT server.

The intermediate layer on the software side for networking between the devices as well as for implementing the interface to the database is handled by Node Red. This is a graphical programming environment to connect e.g. devices with different communication protocols. It is also based on Node.js and exchanges data between nodes with JSON objects.

4.2 Process flow

The description of the process flow refers to figure 4. First, the customer visits the product configurator website provided by the Express app. Then the customer can compose the scooter according to his requirements and send the order back to the server. This is then saved as a JSON document within the Mongo database. The next step is then initiated at the respective workstation. In our scenario, the assembly process is the first workstation, but the concept can also be extended to subsequent work steps. When an empty assembly jig arrives at the workstation, a

RFID chip is scanned. Because this is the first step, the ID of the assembly jig is assigned to the next job in the queue based on a priority. There are several programs on the robot, each program being assigned to a specific part. The programs are triggered by Node Red depending on the required parts and the robot starts to pick up the parts and place them on the assembly jig. A similar situation applies to the Pick by Light system: depending on which parts are required from the supply rack, the display lights up on the corresponding compartment and thus shows the worker which parts and how many he has to take out. Then he puts the parts also on the assembly jig.
The feedback of the systems is received by Node Red and is recorded in the database.

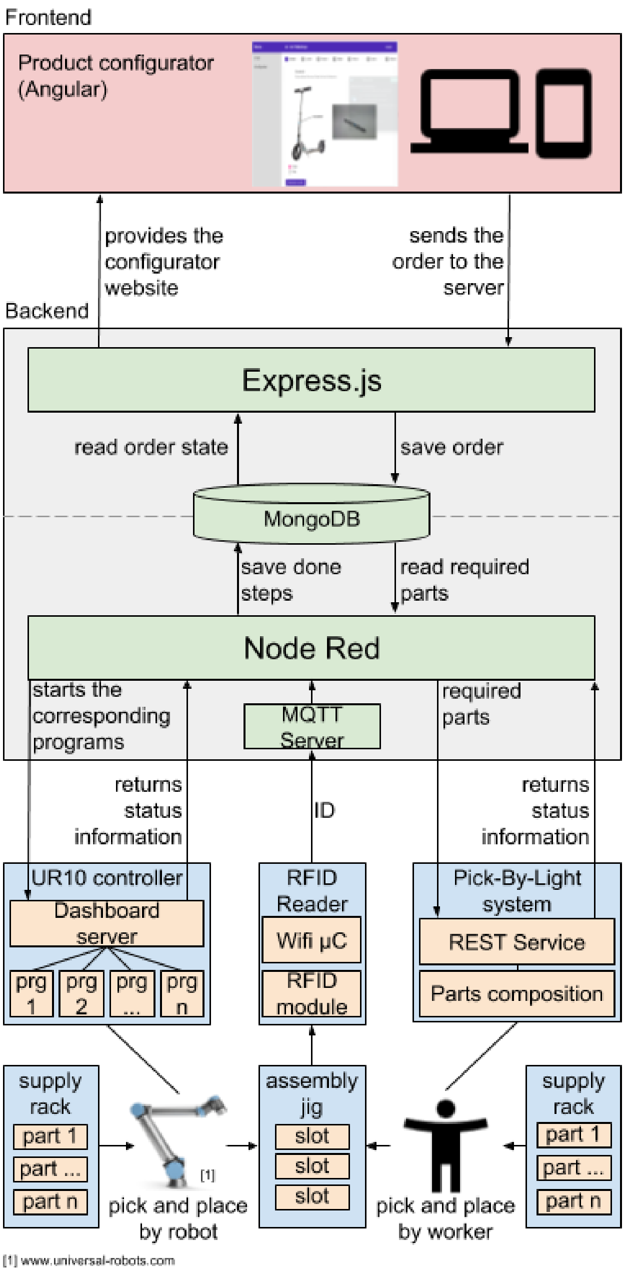


Figure 4 – Information flow of the configuration and assembly process

4.3 Possible extensions

In our prototype implementation the entire backend runs on a local server. It is also an option to run the database and Express app on a cloud and use the local server with Node Red as edge cloud. In this way, several geographically separated locations could also be connected. If additional accessories or variants are added, they can also be integrated, for example by creating a new robot program and linking it to the part number. The principle is not limited to the assembly process and can also be applied to subsequent work steps.

5 IMPLEMENTED EXAMLPES FOR REMOTE CONTROL OF PRODUCTION EQIPUMENT VIA THE INTERNET

Because not all, and in particular older, production systems are networkable from the beginning, the above concepts cannot always be applied. As a solution, various devices and systems were extended so that they can be controlled by a higher-level server. For demonstration purposes, a web interface was created to control them remotely via web browser.

5.1 Remote controlled Industrial Robot

The schematic connection of the individual components can be seen in Figure 5. The existing KUKA robot system provides several digital inputs and outputs. These are connected to a server via CAN-fieldbus (CAN: Controller Area Network) using an I/O-coupler. Various robot programs can be selected and started via a web interface. The corresponding program number is written as binary code to the outputs of the IO-Device. An orchestration program on the robot controller interprets the code and starts the requested program. The system and program status are constantly checked and transmitted to the remote client, as well as measurement values from a distance sensor attached to the robot.

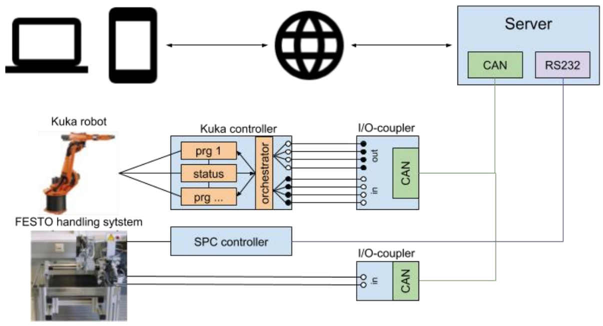


Figure 5 – Scheme for internet controlled devices

5.2 Remote controlled Handling system

This is a 2.5-axis pneumatic handling system which is programmable according to DIN 66025/ISO 6983 (G-Code). A web interface to the system was developed where user programs can be generated and uploaded to the remote system. Subsequently, the commands are transmitted from the server to the controller via serial interface. It is also possible to read the current position of the axes through the serial interface, but only with a low refresh rate. Therefore a CAN module with analogue inputs was installed which captures the values of the analogue measuring system and outputs them in 30 ms cycles. This allows to visualise the real trajectory. Figure 6 shows the visualisation of the web interface. The black lines show the direct connection of the target points and the green dots are the real waypoints executed by the point-to-point controller.

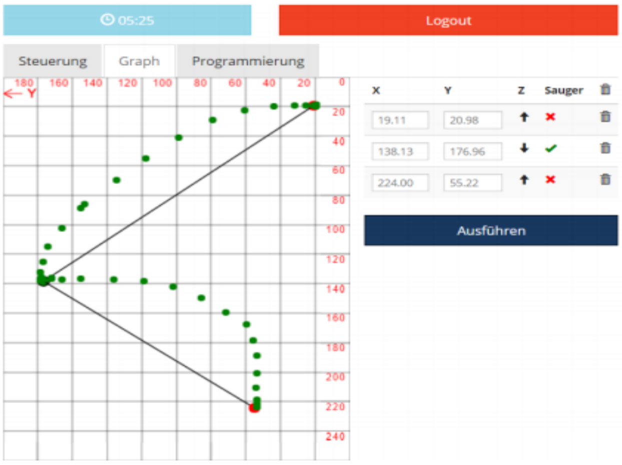


Figure 6 – Control and visualisation web interface of the Handling system

6 CONCLUSIONS

The presented methods and examples show how RMS can be implemented in different ways, using modern technologies. This includes web technologies for Front-end development, building backend solutions with database connection and networking production systems with each other and with higher-level systems.

7 REFERENCES

[1] We push the limits. The journey to Internet of Things and Connected Industry. Robert Bosch GmbH Automotive Electronics, Reutlingen Germany, 2015

[2] Thomas Bauernhansl, Michael ten Hompel, Birgit Vogel-Heuser (eds): Industrie 4.0 in Produktion Automatisierung und Logistik. Anwendung Technologien Migration. 648 pp. Springer Vieweg 2014.

[3] Website of the European research project Factory in a Day, www.factory-in-a-day.eu. Accessed 07-Aug-2018.

[4] Website and open access remote control link of the Virtual Lab and Internet of Things project, http://vvl.fh-reutlingen.de, Accessed 30-Aug-2018.

[5] Giovani Da Silveira, Denis Borenstein, Flávio S Fogliatto, 2001, Mass customization: Literature review and research directions, International Journal of Production Economics, 72/1:1-13

[6] Herrmann A., Hofstetter R., Beck L., 2014, Big Data aus dem Produktkonfigurator: Entscheidungsunterstützung für Kunden und Basis für Customer Insights. In: Brenner W., Hess T. (eds) Wirtschaftsinformatik in Wissenschaft und Praxis. Business Engineering. Springer Gabler, Berlin, Heidelberg

[7] Hvam, Lars & Haug, Anders & Mortensen, Niels Henrik & Thuesen, Christian, 2013, Observed benefits from product configuration systems. The International Journal of Industrial Engineering: Theory, Applications and Practice. 20. 329-338.

[8] S. Jack Hu, 2013, Evolving Paradigms of Manufacturing: From Mass Production to Mass Customization and Personalization, Procedia CIRP, 7:3-8

[9] Yoram Koren, Moshe Shpitalni, 2010, Design of reconfigurable manufacturing systems, Journal of Manufacturing Systems, 29/4:130-141

[10] Jianxin Jiao,Mitchell M. Tseng, A methodology of developing product family architecture for mass customization, 1999, Journal of Intelligent Manufacturing, 10/1:3-20

[11] Report Website of Github, https://octoverse.github.com, Accessed 29-Aug-2018

[12] Usage of client-side programming languages for websites, https://w3techs.com/technologies/overview/client_side_language/all, Accessed 29-Aug-2018

8 BIOGRAPHY

Gerhard Gruhler obtained his Dr.-Ing. degree from the University of Stuttgart. He was appointed Professor for Robotics and Manufacturing automation at Reutlingen University. Currently he is Vice President for Research at Reutlingen University, Germany

Julian Hermann obtained a M.Sc. degree in Mechatronic Engineering from Reutlingen University. He holds a position as researcher at ESB Business School, Reutlingen University, Germany

Energy-Oriented Production Control Using Reinforcement Learning

M. Roesch[1,]; V. Priesack[1];C. Berger[1]; S. Braunreuther[1,2]; G. Reinhart[1,3]

[1] Fraunhofer Research Institution for Casting, Composite and Processing Technology IGCV, Augsburg, Germany

[2] University of Applied Science Augsburg, Augsburg, Germany

[3] Institute for Machine Tools and Industrial Management (*iwb*), Technical University Munich, Germany

Abstract

Going along with the lasting changes on the field of power generation, the prices of electricity are increasing. Hence, manufacturing companies can reduce electricity costs by self-supply, e.g. investing in power plants using wind or solar energy. However, such renewable sources show fluctuating power generation. In order to increase self-supply, companies need to apply measures for flexible energy consumption, e.g. increase energy demand in periods of high energy availability and vice versa. To influence the electricity demand on short-term level, the production control is suitable by adapting the production schedule and thus reacting on unforeseen events. Thereby, on the one hand, trends like future electricity prices need to be considered and thus a predictive decision-making is mandatory. On the other hand, the decisions also need to be taken in real-time. In order to cope with these requirements, a Multi Agent System (MAS) using Reinforcement Learning is proposed in this paper. The decentralized production resources are modelled by autonomous agents which are trained to minimize the electricity costs applying a Q-learning function. An artificial neural network is applied to predict the reward. As the predicted power generation is displayed in the state-action pairs of the model, future electricity availability is considered in decision-making. The presented system decreases the overall electricity costs compared to a benchmark scenario.

Keywords

Energy flexibility; production control; reinforcement learning

1 INTRODUCTION

Climate experts consider the globally increasing number of extreme weather events to be an effect of the anthropologic climate change. [1]. Due to its relation to human carbon dioxide emissions, this leads to increased environmental awareness and puts pressure on politics all over the world to impose profound and sustainable changes. The Paris climat treatment displays so far the temporary highlight in this context, which was signed by in total 175 nations [2]. They all committed to reduce greenhouse gas emissions and reshape the power generation towards renewable sources like wind or photovoltaic (PV). In Germany, the course for sustainable electricity generation was already set in 2010 by establishing the German energy turnaround [3]. Some of the consequences are already visible by now. The costs for electricity sharply increased and compared to the other countries of the European Union, the German industry has to face high prices for electricity [4]. However, electricity prices are an important location factor in global competition and thus companies are looking for ways to reduce costs [5].

As taxes, grid charges and other allocations roughly contribute one half of the industry's electricity costs, energy self-supply using wind, photovoltaic (PV) or cogeneration of heat and power (CHP), is a serious option, as industry can save up to 25 % of energy costs [6,7]. The major drawback of renewable energy sources is the fluctuating power generation, which is strongly dependent on the weather. To increase the power self-consumption, companies can adapt their energy demand on the actual availability [8]. This ability is also called energy flexibility (EF) [9]. EF is enabled by considering energy issues within the production planning and control and thus influence the production program, which determines the energy demand [10]. As in a production environment disturbances and unforeseen events may occur, adequate actions need to be taken in real-time by production control to maintain the production flow and at the same time scoop the potential of EF [11]. Therefore, this paper presents an approach for energy-oriented production control, which considers the energy availability and hence reduces electricity costs.

2 APPROACHES OF PRODUCTION CONTROL

2.1 Strategies

Production planning and control plays a central role within modern manufacturing systems. The production planning appoints, when which orders have to be machined on the given manufacturing ressources. The result is a defined production schedule. In contrast, it is the central duty of production control to implement this given production schedule. Thereby production control has to counteract disruption and unfroseen events, e.g. machine breakdowns. That is why production control is also often called real-time scheduling or rescheduling, as the already given schedule from production planning is adapted [12].

In the literature, two main strategies of production control are found [13–15]:

- **Reactive control**: Thereby, no fixed production schedule is determined. The orders are scheduled based on decentral dispatchin rules, e.g. using priorities. Thus, decision may be taken in real-time, as little computational effort is needed. As a disadvantege, not all effects of the decision may be considered [12]. The decision is all about which order to process next, but there is no deterministic and predictive scheduling of a whole set of orders for a decent time period [12]. One of the most common used diespatching rules is First in First Out (FIFO) or Eraliest Due Date (EDD) [16].
- **Predictive-reactive control:** A deterministic production schedule is calcualted for a given period. The rescheduling can either be triggered periodically or due to unforseen events. Mathematically this process in an optimization problem, which is challenging, as it has to be solved in real-time.

2.2 Techniques

Most of the production control systems are centralized and strictly hirarchically, as these approaches result in a global optimal solutions [14]. Consequently, these systems are applied for predicive-reactive control strategies and calculate a detailed production schedule for a given period in advance. For this task, many diffrent algorithms, such as esxact solvers or heuristcs exist. For a detailed overview see [14]. However, such approaches show some major drawbacks. Scheduling problems are in general NP-hard, this problem even gets worse when processes are disruptable [13,17]. As the required computational time in addition increases exponentially with growing problem complexity (e.g. number of ressources or orders), they may not be suitable for real-time requirements [18]. Some further drawbacks ar vulnerability for errors due to single failors of central nodes and the lack of capability to react quickly on new situations and disturbances [19]. To reduce compuational costs, coordination mechansims for a distributed problem solving can be applied [20]. In this context, Multi Agent Systems (MAS) are widely used for decentral systems in the field of manufacturing [21]. Thereby, various autonomos sofware agents can take decisions locally and communicate with each other and the surrounding environment. Thus, calculation time can be reduced and highly reactive and robust systems can be deployed [22]. However, MAS again have some major disadvantages. The agents do not always succeed in finding a optimal solutions, as they only have a local view of the problem and try to reach their individual targets [23,24].

Though, the application of Artificial Intelligence (AI) can help to overcome these structural problems of MAS [17]. Thereby methods of Reinforcement Learning (RL) are used to create dispatching rules, which are able to flexibly adapt to changing conditions. Thus, the agents are capable to work towards long term goals and to consider more effects in decision making [25]. This displays a big advantage compared to classical reactive control approaches. The Algorithms thereby need to be trained based on a set of training examples, generated by a simulation model. The decentralized agents take some decision and receive a specific reward, depending on the effects of their actions on the fulfillment or non-fulfillment of the global target [26]. Doing so, the agents learn which decisions to take in order to receive the highest reward. In this manner, the agents can compensate their lack of knowlege by learning through trial and error [26]. All in all MAS with RL may be suitable to implement reactive control systems with the ability to act in real-time but also at the same time work for long-term goals und consider various effects in decision making.

2.3 Energy-oriented production control

The classical target functions of production control are delivery reliability or throughput [16]. Considering energy aspects within the control process adds an additional command variable. The orders with high energy demand need to be scheduled in times of high electricity availability and vice versa. Such an energy-oriented schedule can be determined a few day in advance by production control and thereby rely on weather forecasts, which predict the generated power of the installed renewable power plants. As these weather forecasts show a limited accuracy, still some deviations between expected and real power generation may occur. In addition, unexpected events like machine breakdowns in the production site may also cause deviations of the planed schedule. Due to these influencing variables, the production control has to adapt the given production schedule during the day of execution in real-time.

It comes clear, that in order to find the best global solution, also the future power availability has to be considered for decision-making. Reactive control strategies using orthodox dispatching rules are not able to consider these future circumstances. Predictive-reactive systems in contrast are capable to include future aspects like the energy availability, however, the computational expense may increase rapidly like discussed above, thus precluding real-time decision-making. Due to their ability to generate adaptable and flexible dispatching rules, a MAS using RL seems appropriate to cope with the given requirements.

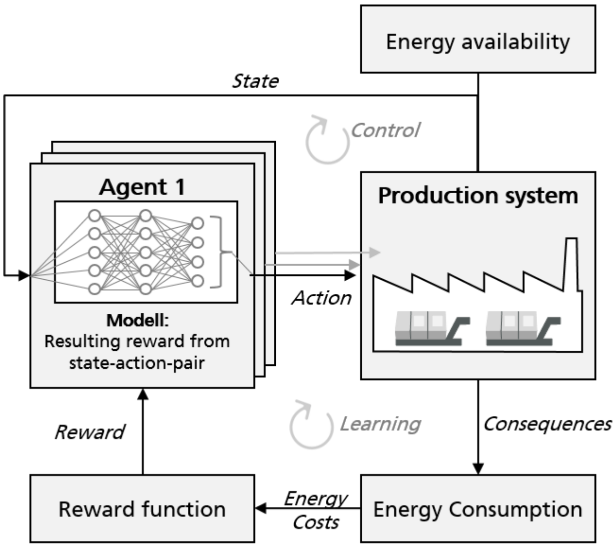


Figure 1 - System concept of production control

3 MAS FOR ENERGY-ORIENTED PRODUCTION CONTROL

3.1 System Concept

In the following, a multi agent approach using RL in order to control short-term energy consumption is presented. The system concept is displayed in figure 1. Every production ressource is represented by a software agent. Besides, the current state of the production system, e.g. number of waiting orders, deadlines and also the future energy availability is regarded for decision making. Based on the current state of the production system, the agents decide for some specific actions, more specifically select which order to process. These actions are executed in the production system and therby entail some consequences, e.g. the order-specific energy consumption and processing duration. Thus, the energy consumption of each agent my be determined and result in sumarized energy costs. All agents receive a global reward, based an the hight of resulted energy costs and train a neural network to predict the received energy costs as a function of the state-action-pair. Thereby the agent-individual model leans to predict the likely resulting global reward depending on the sate-action pair. Thus, alway the one order is selected to be processed, which promises the highest reward. By granting the same global reward to all agents, a cooperative beahviour is encouraged [26].

For each ressource, a decent number of orders may be processed. Every ressource can only process one order at a time and the orders are not disruptable. It is assumed, that the energy demand of ressources when processed is constant and order-specific. In between the processing of different types of orders on the same machine, setup is necessary which consumes energy as well. Therefor the orders are characterized by following three parameteres:

- Order Type
- Required processing time
- Required average energy consumption.

After finishing an order, the agents decide again, which order to process next, considering occuring events like new order arrivals or machine breakdowns.

3.2 Reward Function Learning Approach

The autonomous system agents learn to improve their decisions over time. Their main goal is to maximize the received reward. In the case of this energy-oriented approach, the reward function mainly consists of electricity cost factor. For each production interval t, a decent amount of electricity $E_{AV,t}$ is already available, e.g. supposed to be generated by self-supply. If the total consumed energy $E_{total,t}$ exceeds the available electricity, additional power needs to be purchased from the public grid by a decent price c_t. In contrary, is less energy consumed than actually available, a penalty of c_t has to be payed. However, a deviation of less than 5 % is neglected. Therefore, the energy cost for every time step $C_{E,t}$ is formulated as followed:

$$C_{E,t} = c_t \cdot \delta \qquad (1)$$

$$\delta = \begin{cases} E_{AV,t} - P_{total,t}, & E_{AV,t} \cdot 0{,}975 > P_{total,t} \\ E_{total,t} - E_{AV,t}, & E_{total,t} > E_{AV,t} \cdot 1{,}025 \\ 0, & E_{AV,t} \cdot 0{,}975 < E_{total,t} < E_{AV,t} \cdot 1{,}025 \end{cases}$$

Derived from the total energy costs, the reward function *r* is formulated as followed:

$$r = \sum_{t=t_0}^{t_e} C_{E,t} \qquad (2)$$

The global immediate reward is received only after execution of an order. Thus, the reward is the sum of energy costs between the start-time t_0 and the end-time t_e of the executed order. In this setting the reward consists of positive costs, that are to be minimized. For that reason, the problem at hand is a minimization problem.

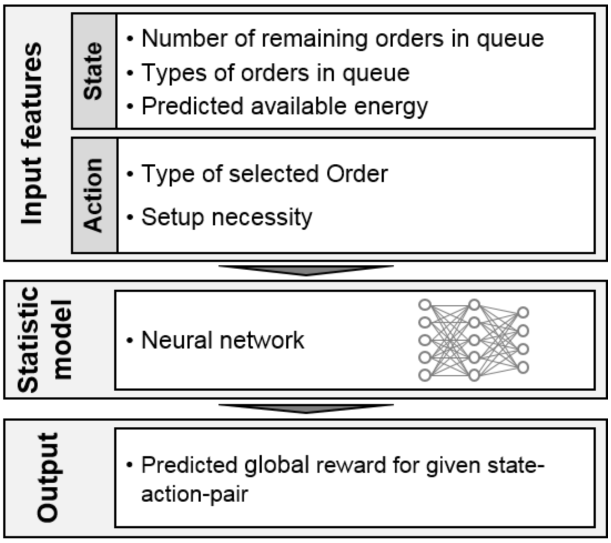


Figure 2 - Overview of learning process

The agents' goal is to obtain an optimal policy π following which means conducting optimal behaviour.

To achieve this, the intelligent agents employ the widely used model-free Q-learning algorithm, which determines the expected reward if action a of the available actions A is chosen in state s. Thereby, policy π is derived. Learning means updating the value-function Q, in case of new experiences were gathered:

$$Q(s,a) = (1-\alpha)Q(s,a) + \alpha(r + \gamma min_{\pi} Q(s',a')) \quad (3)$$

The variable α represents the learning rate, which is set to $\alpha = 0{,}1$ and the variable γ represents the discount factor, which is, due to a finite action-set, defined as $\gamma = 1$. Assuming an optimal value function Q^* the optimal policy can be induced following:

$$\pi^*(s) = \underset{a \in A(s)}{\operatorname{argmin}} Q^*(s,a). \quad (4)$$

The Q-function is utilized as well to predict the power consumption of each agent. Assuming every agent follows the policy and no external events as new orders or breakdowns occur, the orders that will be selected after the execution of the current order are predicted for the next five hours. Thus, the system is extended by a simulation-approach to predict the energy consumption. However, the simulation can still be made in real-time, as little computation time is needed for applying the policy search.

As the state and action sets at hand are very large or continuous, the Q-value function needs to be approximated using a statistic model e.g. a neural network, as it can no longer be represented by a table. This implies, that the optimal policy as well becomes an approximation. Like displayed in figure 2, the input features are the number and type of orders and the predicted available energy of every agent. The algorithm used is the neural fitted Q-Iteration which is implemented using a multi-layer perceptron [27].

The cooperation between the agents is based on the global reward, as the improvement of the local rewards leads to an improvement of the global reward represented by the sum of local rewards. Furthermore, the agents act under an optimistic assumption: every agent assumes that the other agents act optimally. This leads in general to an overestimation of the true value of local state-action pairs and therefore non-optimal local actions may receive too good rewards. The optimistic assumption still leads to better results, because all agents are optimizing themselves during training and the scheduling problem underlies specifications which restrict the size of the state-action space [26].

In order to train the neural network, batches of tuples containing the state, action, reward and successor-state are gathered instead of training using only one tuple. Combined with the resilient backpropagation algorithm (Rprop) this training enhances the convergence towards an optimum [28]. One batch contains 50 four-tuples.

4 PRELIMINARY RESULTS

For proof of concept, a simulation study is applied. The available electricity is a conjunction of at most 25 % wind, 35 % solar and 40 % CHP. The power generation of CHP is constant, whereas solar and wind power are included when available. Therefor, historic data of a wind and solar power plant over one year is applied. The power consumption varies due to the chosen order.

A production system with six production resources with a maximum power consumption of 33 kW is modeled. Every production resource can process between two and four different types of orders. During poduction new orders are arriving and arbitrary breakdowns occur. No more than ten orders are waiting in the queue of one machine. At standby, 30 % of the resource's energy demand is consumed. The setup of a production ressource is represented by an order as well and is always available in the machine queue. The setup-process however requires 50 % of the maximum energy consumption and 25 % of the duration needed to process the respective order. The specifics (e.g. energy consumption or duration) of those orders do not change with exception to their due date, which is generated arbitrarily. The orders take 45 to 120 minutes with a mean duration of 100 minutes and consume between 40 and 100 % of the production resource energy demand. During the learning phase a set S_L is used to generate the environment in which the neural networks are trained.

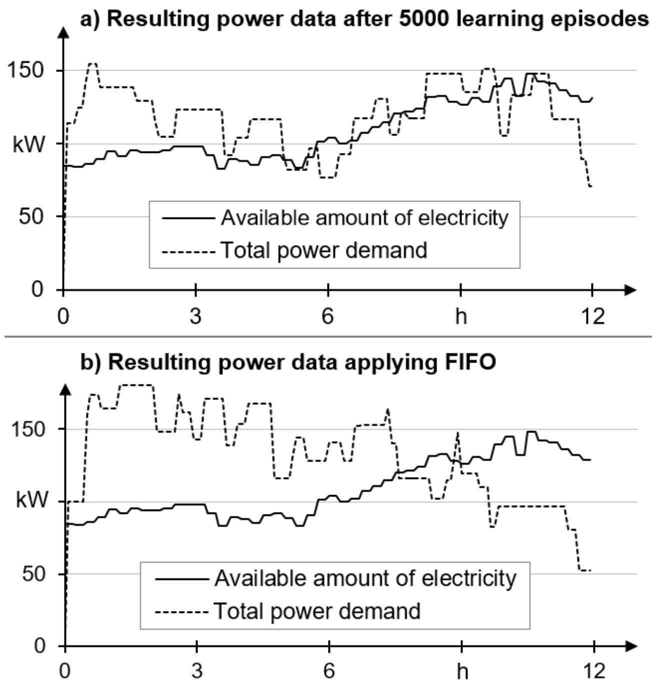


Figure 3 - Adaption of power Consumption

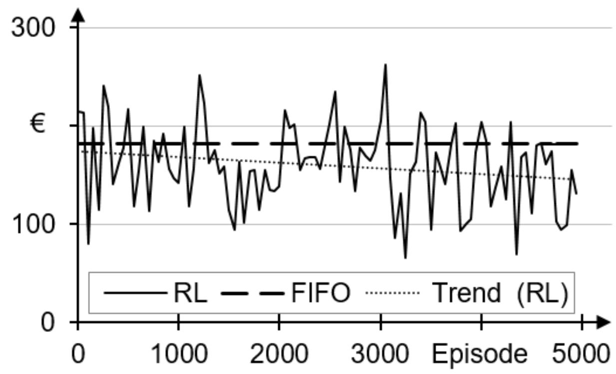


Figure 4: Resulting energy costs as a function of learning episodes

During the application phase a set $S_A \neq S_L$ is used to evaluate the trained neural networks. The application phase is represented by a twelve-hour shift. The system is implemented using C# and Open CV library.

figure 3(a) shows the power data result of the evaluation of the 5000th learning episode. In comparison the power data resulting from the application of FIFO is shown in figure 3(b). It comes clear, that the energy consumption is adapted to the availability after training and thus leads to better results than FIFO.

Considering the costs resulting of the difference between available and consumed energy, training the neural networks shows a trend to reduce these. Compared to the cost resulting from the application of FIFO, RL shows better results. A linear trendline using the least squared method indiciats the optimization trend of the neural networks.

5 OUTLOOK AND CONCLUSION

This paper proposes a real-time production control using a MAS and RL in order to increase electricity self supply of industrial companies. A method for energy oriented order sequencing is thereby presented. A reward function is formulated which consists of energy costs and Q-learning is applied. In order to consider the future energy availability within decision-making, these informations are also included as features in the applied statistic model. Thus, predictive decision can be taken in real-time. For this purpose, additionally the future order selection is simulated 5 hours in advance. A simulation study shows the effectivness of this approach. Within learning episodes a trend towards reduced energy costs can be displayed. The total energy costs are clearly reduced compared to FIFO as benchmark. As major drawback, this system is only able to consider energy costs. In future work, the reward function should be expanded by further relevant cost-factor in production environment. In addition, other training setupts and learning techniques should be implemented in order to enhance convergence.

6 ACKNOWLEDGEMENTS

The authors would like to thank the Bavarian Research Foundation for funding the research. The results presented in this paper have been achieved within the scope of the project “Energy self-sufficiency in manufacturing (German: Energieautarke Produktion)”.

7 REFERENCES

[1] Mann, M.E., Rahmstorf, S., Kornhuber, K., Steinman, B.A., Miller, S.K., Coumou, D., 2017. Influence of Anthropogenic Climate Change on Planetary Wave Resonance and Extreme Weather Events. Scientific Reports 7, 45242 EP -.

[2] UNFCCC, 2015. Paris Agreement. UNFCCC. http://unfccc.int/files/essential_background/convention/application/pdf/english_paris_agreement.pdf.

[3] BMWi, 2010. Energiekonzept für eine umweltschonden, zuverlässige und bezahlbare Energieversorgung, Berlin.

[4] Eurostat, 2018. Electricity prices for non-household consumers - bi-annual data (from 2007onwards). http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=nrg_pc_205&lang=en. Accessed 2 August 2018.

[5] Bankhofer, U. (Ed.), 2001. Industrielles Standortmanagement: Aufgabenbereiche, Entwicklungstendenzen und problemorientierte Lösungsansätze. Deutscher Universitätsverlag, Wiesbaden, s.l., 273 pp.

[6] Ernst & Young GmbH Wirtschaftsprüfungsgesellschaft, 2016. Wirtschaft unter Strom: Wie Unternehmen sich unabhängiger und kostengünstiger selbst mit Energie versorgen.

[7] BMWi, 2016. Fünfter Monitoring-Bericht zur Energiewende, Berlin.

[8] Keller, F., Braunreuther, S., Reinhart, G., 2016. Integration of On-site Energy Generation into Production Planning Systems. Procedia CIRP 48, 254–258.

[9] Graßl, M., 2015. Bewertung der Energieflexibilität in der Produktion. Zugl.: München, Techn. Univ., Diss., 2015. Utz, München, 163 pp.

[10] Keller, F., Schultz, C., Braunreuther, S., Reinhart, G., 2016. Enabling Energy-Flexibility of Manufacturing Systems through New Approaches within Production Planning and Control. Procedia CIRP 57, 752–757.

[11] Schultz, C., Braun, S., Braunreuther, S., Reinhart, G., 2016. Integration of load management into an energy-oriented production control. 14th Global Conference on Sustainable Manufacturing.

[12] Niehues, M.R., 2017. Adaptive Produktions-steuerung für Werkstattfertigungssysteme durch

Dissertation Techni-sche Universität München. Utz, München.
[13]Brackel, T., 2009. Adaptive Steuerung flexibler Werkstattfertigungssysteme: Nutzung moderner Informations- und Kommunikationstechnologien zur effizienten Produktionssteuerung unter Echtzeitbedingungen. Zugl.: Paderborn, Univ., Diss, 2008. Gabler, Wiesbaden, 284 pp.
[14]Ouelhadj, D., Petrovic, S., 2009. A survey of dynamic scheduling in manufacturing systems. J Sched 12 (4), 417–431.
[15]Vieira, G.E., Herrmann, J.W., Lin, E., 2003. Rescheduling Manufacturing Systems: A Framework o Strategies, Policies And Methods. Journal of Scheduling 6 (1), 39–62.
[16]Lödding, H., 2013. Handbook of manufacturing control: Fundamentals, description, configuration. Springer, Heidelberg.
[17]Priore, P., Gómez, A., Pino, R., Rosillo, R., 2014. Dynamic scheduling of manufacturing systems using machine learning: An updated review. AIEDAM 28 (01), 83–97.
[18]Shahzad, A., Mebarki, N., 2012. Data mining based job dispatching using hybrid simulation-optimization approach for shop scheduling problem. Engineering Applications of Artificial Intelligence 25 (6), 1173–1181.
[19]Leitão, P., 2009. Agent-based distributed manufacturing control: A state-of-the-art survey. Engineering Applications of Artificial Intelligence 22 (7), 979–991.
[20]J. Liu and K. Sycara., 1995. Exploiting Problem Structure for Distributed Constraint Optimization, 246–253.
[21]Shen, W., Norrie, D.H., Barthès, J.-P., 2001. Multi-agent systems for concurrent intelligent design and manufacturing. Taylor & Francis, London, 386 pp.
[22]Shen, W., Hao, Q., Yoon, H.J., Norrie, D.H., 2006. Applications of agent-based systems in intelligent manufacturing: An updated review. Advanced Engineering Informatics 20 (4), 415–431.
[23]Caridi, M., Cavalieri, S., 2004. Multi-agent systems in production planning and control: an overview. Production Planning & Control 15 (2), 106–118.
[24]Henseler, H., 1998. Aktive Ablaufplanung mit Multi-Agenten. Zugl.: Oldenburg, Univ., Diss., 1997. Infix, Sankt Augustin, 222 pp.
[25]Jiménez, Y.M., 2012. A Generic Multi-Agent Reinforcement Learning Approach for Scheduling Problems: PhD Thesis.
[26]Gabel, T., 2009. Multi-Agent Reinforcement Learning Approaches for Distributed Job-Shop Scheduling Problems. Ph.D. thesis.
[27]Riedmiller, M., 2005. Neural Fitted Q Iteration – First Experiences with a Data Efficient Neural Reinforcement Learning Method, in: Gama, J. (Ed.), Machine learning: ECML 2005. 16th European Conference on Machine Learning, Porto, Portugal, October 3 - 7, 2005 ; proceedings, vol. 3720. Springer, Berlin, pp. 317–328.
[28]Riedmiller, M., Braun, H., 1993. A direct adaptive method for faster backpropagation learning: the RPROP algorithm, in: IEEE International Conference on Neural Networks 1993. IEEE International Conference on Neural Networks, San Francisco, CA, USA. 28 March-1 April 1993. IEEE, Piscataway, pp. 586-591.

8 BIOGRAPHY

Martin Rösch studied Mechanical Engineering at the University of Applied Science in Regensburg and the Technical University Munich. He is currently research associate at Fraunhofer IGCV with focus on production control using artificial intelligence.

Verena Priesack graduated from Technical University Munich with a B.Sc. in Mechanical Engineering and is currently pursuing her M.Sc. in Mechanical Engineering focusing on production control and software engineering.

Christoph Berger has been a research assistant at Fraunhofer IGCV since 2014. His research interests lie in event-driven production control for cyber-physical production systems. He previously studied energy and automation engineering.

Prof. Dr.-Ing. Stefan Braunreuther is head of the department for production planning and control at the Fraunhofer IGCV in Augsburg. Since 2017 he is Professor at the Augsburg University of Applied Sciences, where he lectures regar-ding topics like factory operations and production management.

Prof. Dr.-Ing. Gunther Reinhart studied Mechanical Engineering at the Technical University of Munich (PhD 1987), where since 2007 he holds the Chair for Industrial Management and Assembly Technologies. Since 2016 he is executive director of Fraunhofer Research Institution for Casting, Composite and Processing Technology IGCV.

Method for the Identification of Energy-Cost-Optimized Machinery Components in Processing Machines.

M. Brugger*[,1], C. Richter[1], C. Härdtlein[1], G. Reinhart[1,2]
[1]Fraunhofer Research Institution for Casting, Composite and Processing Technology IGCV, Augsburg, Germany.
[2]Institute for Machine Tools and Industrial Management (iwb), Technical University Munich, Germany

Abstract
A worldwide rising awareness for green production, increasing energy costs and the change from conventional to renewable energy sources result in new challenges for the manufacturing industry. In South Africa, which has a growing power demand combined with an increment of renewables due to the REI4P (renewable energy independent power producer procurement programme), energy-related topics will play a major role in manufacturing companies. To minimise the energy consumption of machinery, energy efficient design and guidelines are already implemented. Existing machinery in energy-intensive production is optimised during its lifetime to save energy costs. On the other hand, renewable energy sources lead to a fluctuating energy supply. In contrast to energy efficiency the potential of energy-flexibility (the ability of a production system to adapt quickly to the changes at the energy market) on machinery level is often unknown, even for energy-intensive processes. Rules or guidelines for energy flexible design of machinery do not exist. This paper presents an approach to analyse energy-related information of machinery components to identify the energy efficiency and the potential for energy flexibility of machines. Furthermore, first results of an implementation are shown in this paper.

Keywords: Distribution energy; energy flexibility analysis; energy-cost-optimization

1 INTRODUCTION

The increasing number of renewables affect the energy supply system progressively [1]. Especially, solar and wind energy is subject to strong fluctuation, which leads to fluctuating conditions in the grid [2]. Until 2030, the share of renewable energies in gross final energy consumption should reach 30 %. In addition, greenhouse gas emissions are to be reduced by at least 55 % in 2030 compared to 1990 [3]. South Africa´s stagnant economic growth can also be attributed to the ongoing energy crisis in the country. The escalating power prices and shortages at peak times are consequences of massive underinvestment in the electricity sector [4]. In March 2011 South Africa promulgated the Integrated Resource Plan (IRP) 2010-30 and the program REI4P which proposed that 42 % of new electricity capacities initiated between 2010 and 2030 must be generated from renewable energy sources [5]. Figure 1 shows the comparison of the climate goals between Germany and South Africa.

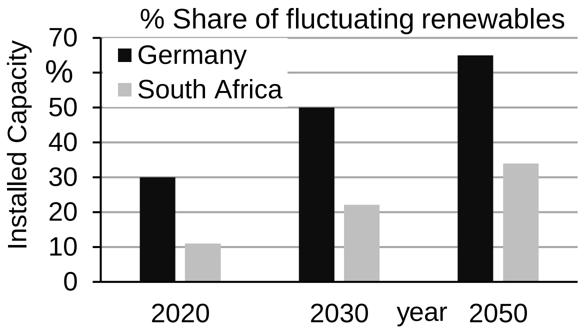


Figure 1 - fluctuating renewables based on [4]

In Figure 1, only the fluctuating renewables are taken into account. Green and non-fluctuating capacity such as green gas is excluded. Figure 1 demonstrates that Germany has already implemented a higher percentage of fluctuating renewables and can be seen as a pioneer country.

To compensate grid fluctuations due to increasing irregular feed-in power, flexible loads are necessary. The energy producer takes actions to adjust the feed-in power or to compensate peak loads. For the regulation, control power by means of conventional power plants is provided. Due to the dismantling of power plants, the consumer will move into focus. Numerous demand-side-management actions can be implemented in manufacturing companies to influence the electricity demand [6].

2 ENERGY FLEXIBILITY IN INDUSTRIAL COMPANIES

The implementation of energy flexibility actions in production leads to targeted influence of the electricity demand. Furthermore, the electricity costs can be reduced by decreasing peak loads or fluctuating electricity prices can be used as an advantage.

Energy flexibility is defined by Graßl as:

"the ability of a production system to adapt quickly and with little financial effort to the changes at the energy market" [7].

This paper proposes to analyse the energy flexibility potential of production machinery.

2.1 Energy-Flexibility Actions

Prerequisite for the identification of energy-flexibility actions is the knowledge of the specific power requirements of the machining process. Five energy-flexibility actions were identified for the machinery level, which are introduced as follows [6]:

Interrupt process
The action includes the temporary suspension of the production process or individual energetic consumers to quickly understand how to achieve a lower performance level of the machine.

Adjust process parameters
The action includes the modification of individual process parameters, such as the adjustment of a temperature level, to operate the machine deviating from the standard.

Change processing order
Changing the sequence of individual process steps within a machine cycle leads to a temporary change of the required energy.

Change energy source
The action refers to the use of different energy sources or their temporary change to the provision of power and influence of the energy demand.

Store energy (short term)
Energy is temporarily stored in a suitable storage medium. For example, electrical energy can be stored as heat energy.

2.2 Energy-Flexibility at Machinery Level

To realize energy flexibility actions sometimes energy has to be converted. The general subdivision of energy is based on its degree of conversion from the extraction to the final use. The common next step is the conversion of final energy into useful energy. Useful energy is referred to as the energy available after the last conversion in consumer´s appliances. An on-machine or in-machine intermediate step of energy conversion between multiple components is not considered [12]. Therefore, the conversion level between final and net energy is defined as distribution energy.

"Distribution energy is a converted final energy level that is available to the end user indirectly and within the process as energy for conversion into net energy."

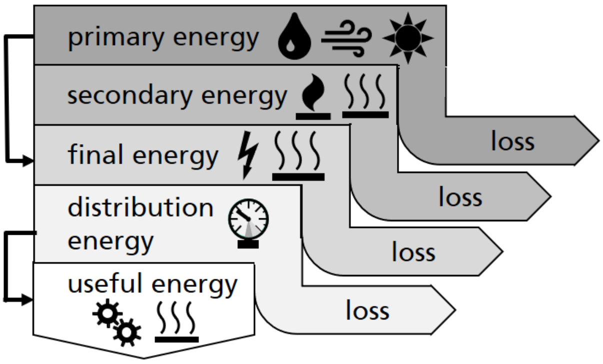


Figure 2 - Distribution energy

As a result, the distribution energy is often tied to the location or consumer (within the machine system boundary) of the last energy conversion. Theoretically, distribution energy can be available for several conversions, with each stage being subject to losses. For example, the electrical final energy can be converted into compressed air energy through a compressor within the equipment. Afterwards, the energy of the compressed air could be transformed to mechanical energy within a machine containing a pneumatic cylinder, which moves an object to its dedicated position.

2.3 Actions at Machinery Level

Various approaches exist for manufacturing companies that address energy flexibility job scheduling, production planning and execution systems. In the field of machine tools, the intelligent coupling of electricity and heat to react to fluctuating prices were investigated through bivalent production components [8]. Another approach shows how the total power requirement of machining centres can be adapted by combining individual consumers or processes and varying the machining energy [9]. Based on a model factory, an evaluation of the economics of machine tools is carried out. A shift of energy demand could reduce the total energy consumption and save costs during simultaneous maintenance of productivity [10]. The approach by Graßl describes how to evaluate energy flexibility based on different machine states [7].

2.4 Need for Reseach

To carry out the full potential of energy flexibility, the energy flow at machine level must be considered too. It is vital to look closely at each machine's components and processes based on energy data [11]. An identification of energy flexibility based on energy data is currently not carried out. Energy data are load profiles of individual components or machines, as well as energy-related machine data sheets. Load profiles are created by time-consuming electrical measurements. To reduce the effort in advance, energy-related machine data sheets can be used to determine the relevance of individual consumers or the energy flexibility potential of machines. To identify the energy flexibility of machinery, thefollowing approach was implented.

3 METHOD FOR A COST-OPTIMIZED ENERGY ANALYSIS

Aim of the method is to combine energy flexibility and efficiency to minimse the total energy costs of machinery as shown in the picture below.

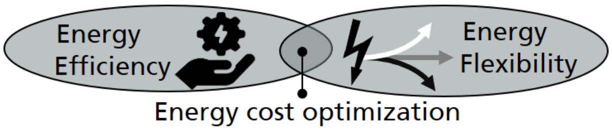


Figure 3 - Field of the energy-cost-optimization

The intersections and contrasts between energy eficeny and energy flexiblitiy are described in [13]. Hence, this paper focuses on the method to identify the energy flexibility of production machinery.

3.1 Hypothesis for an evaluation

To evaluate the energy-flexibilisation potential of individual consumer of production machinery, four hypothesis are introduced and described below.

I. The higher / lower the rated power and peak load of a consumer, the more energy-flexible / energy-inflexible he is.

II. The higher / lower the power output of a consumer in the overall performance of the machine, the more energy-flexible / energy-inflexible he is.

III. The larger / smaller the difference between possible target states of a consumer, the more energy-flexible / energy-inflexible he is.

IV. The faster / slower a change can be made between the states of a consumer, the more energy-flexible / energy-inflexible he is.

The evaluation focus is on the rated power, the peak load and the average power of a consumer. With reference to hypothesis I, it is assumed that a high nominal power and peak load result in a high potential for energy-flexibility. The power share of a consumer in the overall performance reflects its energetic relevance for the overall process. The influence on the energy flexibility of the machine by a conversion increases with rising power share of the consumer (Hypothesis II). Under target conditions of a consumer in the context of Hypothesis III e.g. possible performance points understood in which in the case of a change no restrictions on productivity and product quality arise. A target state has a constant power consumption. With increasing change speed between target states the consumer's level of performance can be adjusted faster and thus it is more energy-flexible (Hypothesis IV). To evaulate these hypotheses, first of all the relavant consumers have to be identified.

3.2 Stucture of the Method

For complex machinery with several consumers first, the most relevant consumers have to be identified. Figure 4 shows the basic procedure for a energy flexibility analysis:

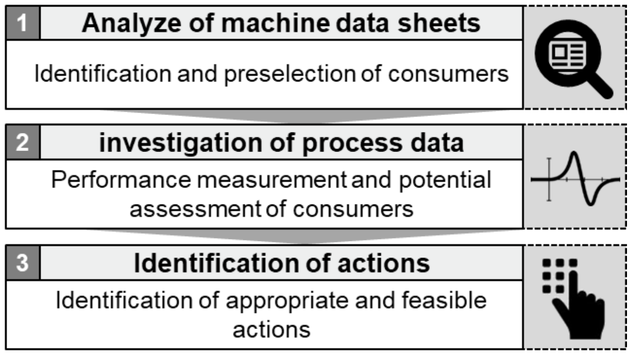


Figure 4 - energy flexibility analysis approach

In the first step an analysis of the machine data sheets is carried out in order to identify relevant consumers with regard to their energy-flexibility potential. As a result, a preselection can be made. The energy consumers are ordered accordingly to their nominal power in relation to Hypothesis I and visualised in a Pareto diagram. The focus is on the most important values, which can achieve the greatest impact [14]. Taking into account the second hypothesis, the cumulative power is displayed by calculating the respective power share of the individual consumers in the total machine power.

The second step of the energy flexibility analysis involves measuring the performance of all relevant consumers during the production process. After measuring the time, the course of the consumed services is presented in a load profile. The aim of the performance measurement is to enable the classification of energetic consumers with regard to their energy-flexibility potential. This will provide a basis to decide on the relevance of consumer-specific energy-flexibility actions.

A decision diagram, which was developed in this context, serves to classify individual consumers as a function of their maximum power and the percentage of average power in the total power [15]. Consumers in priority areas have a high maximum power and a high proportion of the total power. According to hypotheses I and II, they have a high potential for energy-flexibilisation and should be given priority for energy-flexibilisation actions.

In the final step of the energy flexibility analysis, it is important to identify suitable energy flexibility actions (described in chapter 2.3), which can be implemented in terms of machinery and process technology. The time dependencies of individual consumers are used first, which significantly influence the potential for energy-flexibility. In a second step, the energy-flexibility actions existing at the production level are compared with the individual consumers in order to determine the technical feasibility of an implementation.

4 EVALUATION OF THE METHOD

This chapter introduces the machine and measurement technology used for evaluation purposes and applies the method described in the last chapter. Subsequently, the identification, implementation and effectiveness check of derived energy-flexibilisation measures is presented.

4.1 Machinery

A thermoforming packaging machine (TPM) was used for an evaluation. The process stages of a TPM can be divided into the five steps of forming, filling, sealing, seperating and comissioning. These steps are interdependent, although product quality is not affected by the respective previous process step. The machine combines two films (top and bottom film). In the forming station, the lower film is first heated by two preheating plates and softened for the forming. The lower film is then pressed into a mold with compressed air. A water cooling system causes the formed film to solidify. The sealing of the plastic packaging is carried out by a heat-sealing process. In the separation station, the welded moldings are disassembled into individual packages. Beside the production stations also the peripheral equipment e.g cooler is included. Table 1 shows an extraction of the electical components in the thermoforming packaging machine.

Station	Consuming component	Function	Useful Energy
forming	heating plate top	heating top film	thermal
	heating plate bottom	heating top film	thermal
sealing	heating plate	heating sealing	thermal
commis-sioning	trim rewinders	winding of trim	mechanical
	conveyor belt	movement of packages	mechanical
peripheral equipment	chain conveyor	movement of film	mechanical
	cooler	cooling forming plates	thermal

Table 1 - Extract of the electrical components inside the thermoforming packaging machine.

4.2 Step 1 Analysis of the data sheets

Machine data sheets can be procured and evaluated with the operating instructions or the type designations of assembled components. The result is shown in Figure 5.

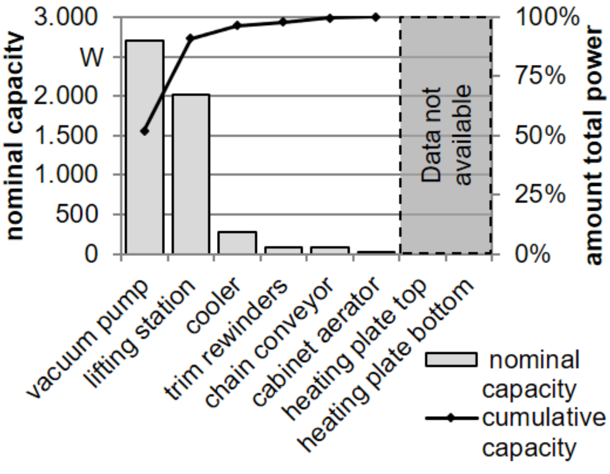


Figure 5 - Pareto diagram power consumer TPM

The vacuum pump and lifting station have the highest rated power and account for almost 90% of the total output. Due to the lack of data, not all individual consumers can be analyzed and ranked according to their nominal output. The nominal capacities of the heating plates and other components were not available. Therfore they have to be measured for further analysis.

4.3 Measuring technology

The power measurement of the electrical loads is carried out using the DS-Net V8 measuring system from DEWESoft. The maximum sampling rate of the system is 10 kHz [16]. The electrical current is measured inductively on the supply lines. The electrical voltage is measured with a measuring adapter type DLA-3L 32 from Voltcraft. [17]. The system software enables real-time display and data storage of the measured values [16]. Compressed air consumption is measured initially by flow meters, which have a measuring range of 0.1 to 1,000 l/min and operate according to the calorimetric measuring principle. [18]. For temperature measurements, Type K thermocouples are used by Omega. The temperature measurement has an accuracy of ± 0.1% [19]. In addition, a thermal imager of the type Fluke Ti25 and the associated image processing software is used. [20].

4.4 Step 2 Investigation of Process Data

The classification is done with the exclusion of the vacuum pump, which exceeds the peak load of the remaining power consumers by approximately 100%. As a first result, the investigation shows, that the vaccum pump is overdesigned for the production process and can be replaced for an energy cost optimization. Hence, the vacuum pump will not be considered in the further course.

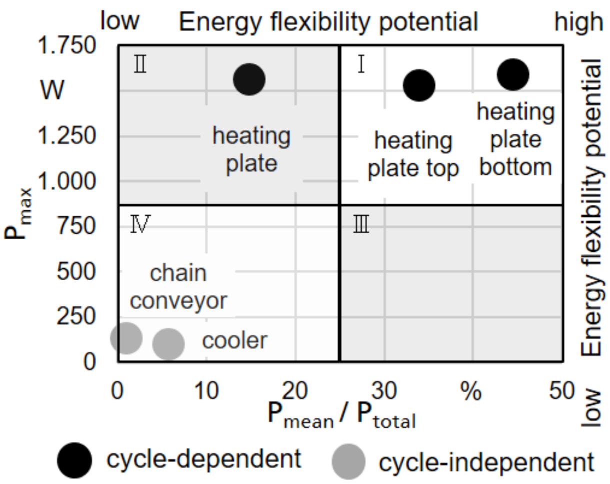


Figure 6 - decision diagram TPM

As shown in Figure 6, the top heating plate and bottom heating plate have the highest energy-flexibility potential. The heating plate is due to a high maximum power and a low share of the total power to the priority area II assigned. The remaining electrical consumers are in priority area IV and therefore have insufficient energy-flexibility potential.

4.5 Step 3 Identification of Actions

In the following, the energy-flexibility actions existing at machinery level compared with the electricity consumers in order to identify suitable actions. The decision on the feasibility of individual actions based on control-technical application options, the procedural process sequence and physical conditions (e.g. theoretical storage capacity). An exchange of individual consumers for the extension of actions is not carried out. The process parameters are adapted via the machine control and the specified operating standard maintained.

The time-independent power consumer heating plate top, heating plate bottom, heating plate and cooling unit allow a process interruption in terms of control technology. An adaptation of the process parameters is not possible with the trim rewind, conveyor belt and chain conveyor. The actions change processing sequence and change energy source are not technically feasible at the TPM. A change of energy source would be associated with a component exchange. Processing order cannot be performed due to the process. The heat plates have their own short-term storage capacity. The following electricity consumers used for actions implemented, due to the highest energy flexibility potential.

- Heating plate top
- Heating plate bottom

● viable ○ Not viable

	consumer	Interrupt process	Adjust process parameters	Change processing order	Change energy source	Store energy (short term)
forming	heating plate top	●	●	○	○	●
	heating plate bottom	●	●	○	○	●
sealing	heating plate	●	●	○	○	●
Commis-sioning	trim rewinders	○	○	○	○	○
	conveyor belt	○	○	○	○	○
peripheral equipment	chain conveyor	○	○	○	○	○
	cooler	●	●	○	○	○

Table 2 – electricity action identification

5 IMPLEMENTATION OF ACTIONS

This paragraph shows the results of the actions that were carried out as well as the energy flexibility potential of the TPM.

5.1 Storage of thermal energy

In the first step, maximum and minimum operating principals were identified to assess the potential of thermal storage. The identification fulfilled by a change of the operating parameter. In addition, a quality analysis of the product took place. The results are presented in Table 3.

temperature	heating plate top	heating plate bottom
ϑ_{min}	95 °C	95 °C
ϑ_{max}	122 °C	130 °C
$\Delta\vartheta$	**27 °C**	**35 °C**

Table 3 - temperature of the heating plates

The temperature differences $\Delta\vartheta$ represent the operating range of the preheating plates. The minimum temperatures ϑ_{min} of the preheating plates are each identical, the maximum temperature ϑ_{max} of the plate top is lower than the plate bottom. With respect to the third hypothesis, ϑ_{min} and ϑ_{max} are the target states for the consumers.

With respect to Hypothesis 4, the storage capacity of the heating plates is analysed. Therefore, it is necessary to determine the time duration between ϑ_{min} and ϑ_{max} and vice versa.

All adjustments of the process parameters are carried out during the production process at a room temperature ϑ_{room} = 23 ° C. Both power consumers are initially set to the minimum temperature ϑ_{min}. The temperature level maintained for a period of about 10 minutes in order to determine the average power P_m representative of this target state.

5.2 Result of the action

The experimental results of the implementation of the measures are shown in Figure 7. The diagram shows, the temperature profile and the average power P_m within the period's t. On the other hand, it points out the electrical energy demand E_{el}. The heating process takes 4.6 min, the cooling process 9.45 min.

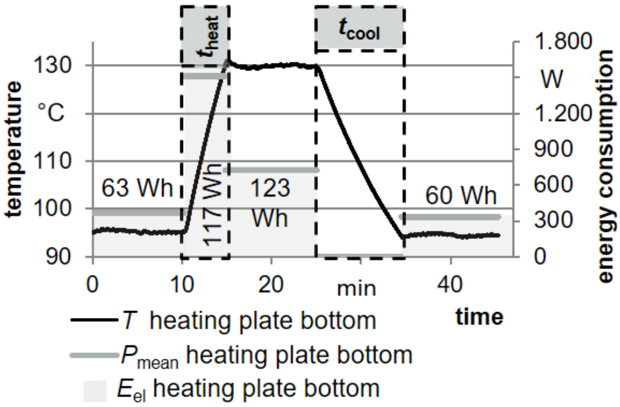


Figure 7 - thermal storage on a TPM

The energy required for the heating process is constant due to the specific heating time per preheating plate.

6 SUMMARY AND OUTLOOK

Initially, distribution energy in the energy transaction as well as an approach for an energy analysis were defined in this paper. The first step of the energy flexibility analysis provides a low-cost option for identifying relevant machine components. The visualization of the nominal power and cumulative power by means of a Pareto-diagram provides a quick overview of the importance of individual consumers. By preselecting in the first step, the measurement effort for the second step is reduced. By means of a decision diagram, the energy flexibility potential is evaluated exclusively based on P_{max} and the contribution of P_{mean} / P_{totoal}. At this point, a differentiation of power and energy must be made. This takes into account the time-share of the services that is relevant for the assessment of the energy-flexibility potential. The feasibility of energy-flexibility actions in the third step depends on the machine and process. The energy flexibility of a machine thus depends decisively on the product to be manufactured or the method used and on the possibilities of intervention in the machine control.

Further research focuses on the generation of a development method regarding an energy-cost-optimization as well as a benefit-cost analysis.

7 ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support of the Kopernikus-project „SynErgie" by the Federal Ministry of Education and Research (BMBF) and the project supervision by the project management organisation Projektträger Jülich (PtJ).

8 REFERENCES

[1] National Academy of Sciences, The Royal Society, Climate Change Evidence & Causes: An overview from the Royal Society and the US National Academy of Sciences (National Academy of Sciences and The Royal Society)

[2] Staffell I., Pfenninger S., 2018, The increasing impact of weather on electricity supply and demand

[3] bmu, 2017, Climate Action in Figures: Emission Targets Germany and EU (bmu)

[4] Pollet B. G., Staffell I., Adamson K.-A., 2016, The Energy Landscape in the Republic of South Africa, 1st edn, Cham

[5] McDaid L., 2014, RENEWABLE ENERGY INDEPENDENT POWER PRODUCER PROGRAMME REVIEW 2014: Electricity Governance Initiative of South Africa

[6] Roesch M., Brugger M., Braunreuther S.; Reinhart, Gunther, 2017, Klassifizierung von Energieflexibilitätsmaßnahmen ZWF 112 567–71

[7] Graßl M., 2015, Bewertung der Energieflexibilität in der Produktion, Technische Universität München

[8] Köse E., Sauer A., Pelzel C., 2017, Energieflexibel durch bivalente Produktionsanlagen wt Werkstattstechnik online 107 366–72

[9] Brecher C., Kehne S., Epple A., 2017, Energieflexible Fertigung mit Werkzeugmaschinen ZWF 112 25–8

[10] Popp R. S.H., Liebl C., Zaeh M. F., 2018, A Multi-level Procedure to Evaluate the Energy Flexibility Potential of Production Machines Procedia CIRP 67 75–80

[11] Brugger M., Götz G., Richter C., Reinhart G., 2017, Study on Energy Efficiency and Flexibility in Machinery and Production

[12] Kaltschmitt M., Streicher W., Wiese A., 2014, Erneuerbare Energien, 5th edn, Berlin

[13] Brugger M., Richter C., Reinhart G., 2018, Intersections and Contrasts between Energy Efficiency and Flexibility in Production Machinery Energy – an Analysis Regarding Energy Optimisation

[14] Winz, G.: Qualitätsmanagement für Wirtschaftsingenieure. Qualitätsmethoden, Projektplanung, Kommunikation. München: Hanser 2016. ISBN: 978-3-446-44639-7.

[15] Brugger M, Selleneit V, Richter C, Reinhart G , 2018 Potenzialanalyse zur Energiekostenoptimierung von Produktionsanlagen ZWF 09/2018

[16] DEWESoft: DS-NET. Technical Reference Manual, 2016.

[17] Voltcraft: DLA-3L Messadapter. Bedienungsanleitung, 2015

[18] Festo AG & Co. KG: Durchflusssensor SFAB. Datenblatt, 2018.

[19] OMEGA: RDXL6SD 6-kanaliger Handheld-Datenlogger für Temperaturen mit Touch-screen. Datenblatt, 2018.

[20] Fluke: Ti10, Ti25, TiR and TiR1. Thermal Imagers, 2007.

9 BIOGRAPHY

M. Sc. Martin Brugger studied Technical Operations Management at the University of Applied Science in Mülheim a. d. Ruhr. He is currently research associate at Fraunhofer IGCV with focus on flexible production - flexible systems engineering for easy automation of production processes

Dipl.-Ing. Christoph Richter is head of the department for Systems and control technology at the Fraunhofer IGCV in Augsburg. He studied Mechanical Engineering at the TU Munich with a focus on "Mechatronics & Information Technology".

M.Sc Christian Härdtlein studied industrial engineering at the University of Applied Science in Kempten. He composed his master thesis at the Fraunhofer IGCV in Augsburg

Prof. Dr.-Ing. Gunther Reinhart studied Mechanical Engineering at the Technical University of Munich (PhD 1987), where since 2007 he holds the Chair for Industrial Management and Assembly Technologies. Since 2016 he is executive director of Fraunhofer Research Institution for Casting, Composite and Processing Technology IGCV.

Approach for the Identification of Energy Flexibility in Production Systems

P. Simon[1], A. Hohmann[1], G. Reinhart[1]
[1]Fraunhofer Research Institution for Casting, Composite and Processing Technology IGCV Provinostr. Augsburg, Germany

Abstract
The nuclear catastrophe of Fukushima in the spring of 2011 spawned a huge debate in Germany about how energy should be produced in the future and whether it is still morally justifiable to continue using nuclear energy. As a consequence, a large number of nuclear power plants were turned off in order to replace nuclear power with renewable energy. Due to the non-dispatchable and non-storable nature of renewables, this will lead to high volatility in the German power grid and therefore fluctuating energy prices. Consequently, companies have started to develop methods for energy flexibility - a new form of demand-response in the industry sector - which helps to alleviate power grid instabilities and offer the chance to lower electricity costs. In essence, the main objective of this strategy is the synchronisation of the production load profile with the energy price profile, e.g. in order to produce the planned programme during times of low energy prices.
The goal of this approach is the development of a method for identifying energy flexibility potential in production systems. Thereby the authors deliver different axioms and tools to rank production systems by their capabilities for implanting energy flexible behavior. Several energy flexibility strategies are developed in order to provide an approach that can be used by companies regardless of their size or energy consumption.

Keywords
energy flexibility, flexible production system, renewable energies

1 INTRODUCTION

The focus of the German energy policy is to ensure a secure, affordable and environmentally stable energy supply while maintaining the competitiveness of Germany as an industrial location. These goals are to be achieved through the implementation of the energy transition. Assuming a largely linear development of the expansion of renewable energies in the years 2010 to 2020, Germany would exceed its set minimum target of 35% by the end of 2018 [1]. The availability of renewable energies, especially wind energy, is associated with high uncertainty due to its dependence on environmental and weather conditions [1]. Here it can be seen that the feed-in into the power grid takes place independently of its needs and thus can lead to an imbalance between electricity supply and demand. To ensure grid stability, the electricity market reacts with incentives for load shifting. Producing companies that adjust their power consumption accordingly can save energy costs. There are therefore advantages for both industrial users and network operators [1] [2] [3]. In this context the question arises for manufacturing companies as to the amount of potential energy cost savings. However, there is often a lack of knowledge about the corresponding available potentials. For this reason, the following approach presents a way to identify energy flexibility in production systems.

2 FUNDAMENTALS OF EVALUATING ENERGY FLEXIBILITY

2.1 Definitions and focus

In addition to electrical energy, manufacturing industries also consume other energy carriers, such as oil, gas, coal and heat [4]. But considering the prices of and the demand for the different energy carriers, electrical energy has the mayor relevance for manufacturing industries by determining more than 50 % of the energy costs. Furthermore, prices for electrical energy are more volatile than prices for other energy carriers, which gives manufacturing industries the opportunity to reduce energy costs by mainly consuming electrical energy when energy prices are low. Therefore the focus of this work is on electrical energy.
While oil, gas, coal and heat are especially used for the generation of room and process heat, electrical energy is used in all areas of a factory. The major share of the consumed electrical energy is for generating mechanical energy, e.g. in a machine tool, and the generation of process heat [4]. Therefore, most of the consumed energy of a factory is used in production, see Figure 1.

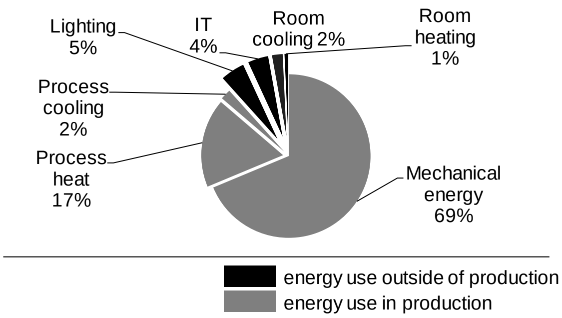


Figure 1 - Consumption of electrical energy in manufacturing industry [4]

2.2 Pricing models for the integration of energy flexible production systems into the energy system

To maintain the balance of the energy grid, the adaptation of the energy demand to the availability in the grid is a promising approach. This approach is known as energy demand response in literature [5]. According to the U.S. Department of Energy, demand response can be defined as follows [5]:

Demand responses are changes in electric usage by end-use customers from their normal consumption patterns in response to changes in the price of electricity over time, or to incentive payments […] when system reliability is jeopardized.

As mentioned in the definition, demand response instruments can be categorized into two basic groups: the price-based demand response and the incentive-based demand response programs, see Figure 2.

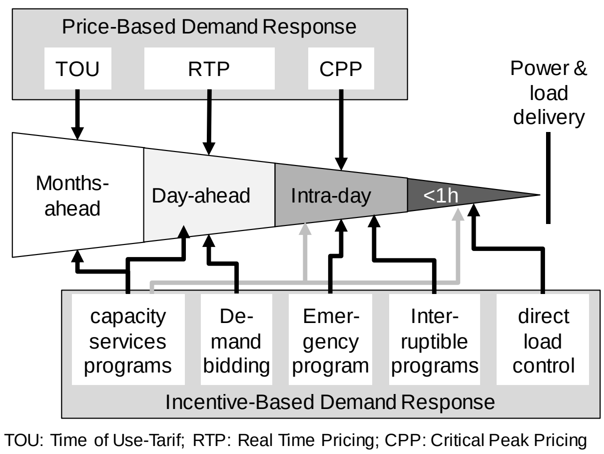


Figure 2 - Categorization of demand response instruments [4]

Price-based demand response gives customers time-varying rates that reflect the value and costs of electricity in different time periods, for example real-time pricing (RTP), i.e. prices varying every hour and reflecting market prices [6].These tariffs also can be more static, such as Time of Use (TOU) and Critical Peak Pricing (CPP) tariffs, where prices are fixed for longer blocks of time [7]. With the information of the actual energy price, customers can decide whether or not they want to use less electricity at times when electricity prices are high.

The second group of demand response instruments is incentive-based. Customers can participate in these programs in addition to normal tariffs. By using incentive-based programs, participants receive a payment when reducing or enhancing its load at times requested by the program sponsor, triggered either by a grid reliability problem or high electricity prices [9], [10]. The different types of incentive-based demand response instruments vary by the time the customer has in advance to commit his willingness to adapt his power demand and how the incentive payment is done.

The presented demand response instruments give production systems the opportunity to achieve energy cost savings as compared to fixed price tariffs, by consuming less energy when energy costs are high and enhance energy demand when energy costs are low. This requires knowing the production systems' energy demand and possibilities to adapt it, i.e. production systems have to be energy flexible.

2.3 Energy flexibility in literature

There are many approaches in literature concerning energy flexibility in production and dwellings. These approaches can be found in three fields: energy engineering, process engineering and manufacturing engineering.

The focus of approaches in the field of energy engineering is mainly on the changing of the energy demand of machines in households, such as washing machines and refrigerators [11], [12]. While they use direct load control, this method cannot be used for production stations in manufacturing industry, as it does not consider costs for adapting the energy demand. Also, a few approaches concerning ventilation facilities in manufacturing industry can be found [13], [14]. Still, approaches for the evaluation of energy flexibility are missing.

The first work regarding energy flexibility in process engineering can be found in the 80's [15]. The potential of saving energy costs by using volatile energy prices was realized early, as processes in process engineering usually are energy intensive. Therefore some approaches exist which take into account all dimensions of flexibility, also the time and cost dimensions [16], [17]. As processes in process engineering have a low complexity compared to processes in manufacturing, these approaches cannot be directly used for this work.

Approaches in manufacturing engineering are not that well developed compared to the state of the art in energy and process engineering. Although there are some approaches using and investigating energy flexibility in production, usually all dimensions of flexibility, especially costs and time dimensions, are not taken into account [18], [19]. However, these are important parameters for the evaluation of energy flexibility. Therefore it can be said that there are no approaches which allow the identification and evaluation of energy flexibility of a production system.

3 CONCEPT FOR THE IDENTIFICATION OF ENERGY FLEXIBILITY

3.1 Overview

The first step is the selection of the relevant production station in order to reduce the work load of the analysis to a level that users will consider taking. The second step describes the modelling of the identified production stations. This is the basis of the evaluation, aided by the axioms of energy flexibility in Step 3. Step 4 describes the analysis of the availability of EFM that are supposed to be used at a parallel time.

3.2 Selection of production stations

The first step is to select the production stations of a production system that have to be considered in the context of energy flexibility. [21] proposes the two characteristics of the power requirement and the share of the total energy demand of a production station in order to evaluate the energy flexibility of this production station. This approach evaluates the energy flexibility at the production system level. Therefore it is necessary to supplement it with the capacity utilization factor of the production station in order to consider the availability of an EFM in a later analysis of the effects of these in the material flow. This is of further relevance since a desired change in performance in the positive (e.g. the connection of a production station) and the negative (e.g. the shutdown of a production station) direction can be included as energy flexibility potential. Thus on the one hand, a production station with high capacity utilization is more often available to carry out a load shedding. On the other hand, a production station with a low utilization is more often available to increase the load. This is the case, for example, in test systems in high-frequency technology, which are used only a few times a month and whose high energy requirement can be used as needed. However, it should also be noted that excessively high utilization is also to be classified as negative, as it is often the achievement of logistical goals in contradiction to the use for energy flexibility. Also, an extremely low utilization indicates that a production station is rarely used and therefore very rarely available for use. The assessment at the production system level naturally also includes an analysis of several production stations. The recommendation of an order in which these are to be examined represents an added value for the user, since the effort can be significantly reduced. This is possible because the procedure can be aborted from any point because the most promising production stations have been investigated. In order to meet the requirements of a selection of production stations on the production system level, a procedure is presented below which takes into account the power requirement and the utilization. The proportion of the total energy requirement will not be considered, since the availability of an EFM cannot be recorded with sufficient accuracy. Starting from this, the maximum value of the power requirements of the production stations, which can often be found in the machine manual, is divided into several corridors. The number of these depends on the number of production stations and should be selected so that they are in the same corridor only in the case of similar performance requirements. For example, it can be checked whether the power requirements of production stations in the same corridor do not differ by more than ten percent. The higher the assigned corridor of a production station, the higher its priority. If several production stations are located in the same corridor, the higher power requirement is not decided by the prioritization, since this is only marginal, but the level of utilization of the production station. Figure 3 provides an example of the process.

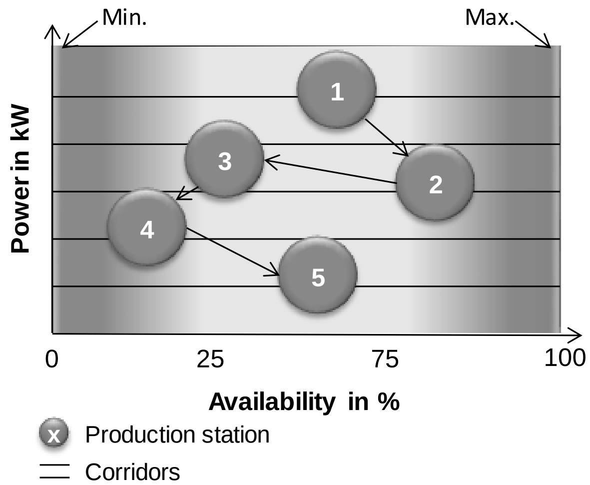


Figure 3 - Example of the result of selection of production stations

The next step provides the detailed analysis of the previously identified production stations.

3.3 Modelling the behavior of the production stations

The power and energy consumption of a production station depends on the different machine states the station can engage, see Figure 4. Therefore, an adaption of the energy demand of a production station can only take place by changing the actual engaged machine state.

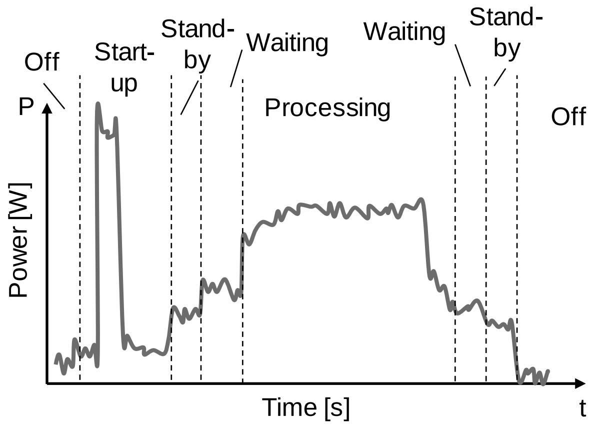


Figure 4 - Power consumption of a production station depending on machine states [20]

In this context a machine state can be defined as a stadium of a production station in which the station is conducting a defined function, e.g. producing or waiting. A machine state has one or more triggering events and one terminating event, which define the function of the state. Every machine state has a constant power demand. If the power demand of a machine state does not deviate from another one and is performing similar tasks, they can be grouped together [21]. A production station can only adopt one state at any given moment.

As adaptions of the energy demand of a production station can only take place by changing the actual engaged machine state, energy flexibility measures (EFM) have to be found. In this work an EFM is defined as an intentional action for conducting a change of the engaged machine state. The next two steps deal with the description of the characteristics of the identified EFM.

3.4 Axioms of energy flexibility

Figure 5 shows three different outgoing machine states which allow adaptions to other target machine states. All of these three outgoing machine states have one or two adoptable target machine states with the same maximal power demand. Since flexibility is a function of variability, a machine state displaying more adoptable target states should be more energy flexible than a machine state with fewer adoptable states. Therefore, machine state 1 has to be less energy flexible than machine states 2 and 3 as it has only one adoptable state. Yet the difference between the power demands of the different machine states also influences energy flexibility. While the power demand of the machine state 1 of the outgoing machine state 2 is very similar to the one of machine state 2, machine state 1 does not give much additional contribution to the energy flexibility of machine state 2. This leads to the conclusion that a more even distribution of the machine states should increase energy flexibility since it increases variability. Therefore machine state 3 is more energy flexible than machine state 2 which again is more energy flexible than machine state 1.

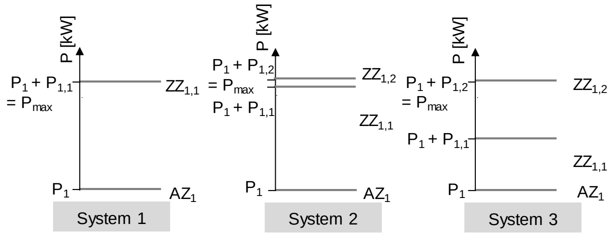


Figure 5 - Systems with different adoptable states

The next property under investigation is the influence of the time and costs needed for changes of states and how this affects energy flexibility of a production machine. As costs are incurred when changing machine states, e.g. additional energy costs for a machine restart after a shutdown, a company will only adapt its energy demand to demand response instruments when energy cost savings increase beyond the costs of EFM for adapting to the energy demand.

While EFM need to be activated for a change of the energy demand of a production station, high activation times bring inertia into the system and can make momentary state changes impossible due to energy price changes. Therefore a system is more energy flexible when the required time for a change of state is low. As shown in Figure 6, system 2 needs 5 periods of time for a change of states while system 1 only needs 3 periods and can therefore adapt its energy demand within the required activation time of the pricing model.

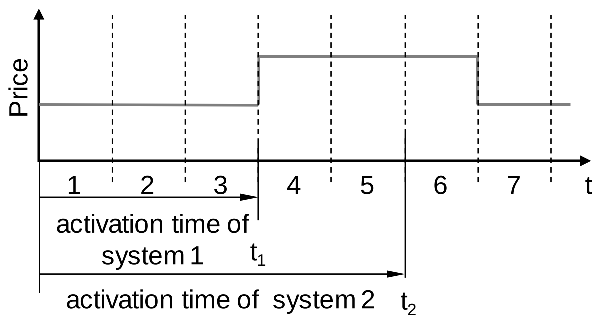


Figure 6 - Influence of activation time on energy flexibility

Also time boundaries of a machine state, which describe how long a machine state is adoptable, affect energy flexibility. A low positioned lower time boundary ensures fast machine state change, and is therefore positive for the energy flexibility of a production station. A high positioned upper time boundary also has a positive influence on the energy flexibility, since it allows long stays in machine states and therefore allows the profit of low energy prices. Referring to the conclusions made in this section, six essential axioms to enable evaluation of energy flexibility are made [21]:

1. The energy flexibility of a system increases/ decreases with an increase/decrease of adoptable states.
2. The energy flexibility of a system increases when the distribution of the power demand of the adoptable states is more even.
3. The energy flexibility of a system increases/ decreases when the required time for a change of state decreases/increases.
4. The energy flexibility of a system increases/ decreases when the lower time boundary of an adoptable state decreases/increases.
5. The energy flexibility of a system increases/ decreases when the upper time boundary of an adoptable state increases/decreases
6. The energy flexibility of a system increases/ decreases when costs for a change of state decreases/increases.

3.5 Availability of production stations

In order to fulfil the practical requirement of the highest possible availability of several EFM, an internal limit value is also introduced, which can be chosen freely by the user. This checks whether the viewing production station is in the desired initial state for a sufficiently long time. The higher the value that is chosen, the more EFM will be included in the rating, as any EFM with a value lower than the threshold will be excluded before bundling. The calculation of the value of combined EFM succeeds on the basis of the already described consideration of states of the respective production station over time. For this purpose, the entire measurement period is divided into intervals of the same size, e.g. 15 minutes. This creates a value and time discrete performance profile. In the next step, the affected output states of the EFM are selected, which are then checked with regard to their number of time coincidences. Finally, the percentage is calculated.

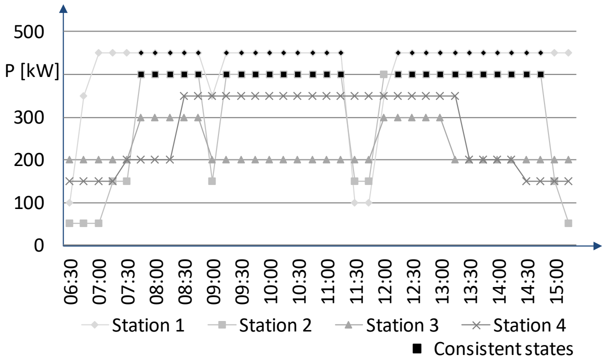


Figure 7 - States of 4 stations as a basis for the analysis of the availability of a combination of EFM

Figure 7 illustrates this relationship in a simplified example of four stations, each of which can take several states. A combination of EFM consisting of an EFM of the production station 1 with the initial state of 450 kW and an EFM of the production station 2 with the initial state of 400 kW, would thus be assigned an availability of 69.44%. The associated states are marked in black.

4 CONCLUSION

Due to the increasing use of renewable energy sources – especially wind power – prices are getting more volatile on energy markets. Based on these uncertainties, different demand response instruments try to achieve changes in electric usage by end-use customers from their normal consumption patterns in response to changes in the price of electricity when system reliability is jeopardized. Therefore, production systems have to be energy flexible.

This work shows how the energy flexibility of a production station can be evaluated. The KPI provided in this work for the evaluation of the energy flexibility is based on the amount and the distribution of energy levels of different machine states. Furthermore, time and cost dimensions of energy flexibility are taken into consideration. Also the application to a production station is shown.

Based on this work future research has to investigate how energy flexibility of a production system can be evaluated. Especially the dependencies between different machines of a production system have to be considered in the evaluation of energy flexibility.

5 ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support of the Federal Ministry of Education and Research (BMBF) and the project supervision of the Project Management Jülich (PtJ) for the project “SynErgie”.The authors are responsible for the contents of this publication.

6 REFERENCES

[1] Von Roon, S.; Grobmaier, T.: Demand Response in der Industrie – Status und Potenziale in Deutschland. München: Forschungsstelle für Energiewirt-schaft e. V. Dezember 2010.

[2] Moura, P. S.; De Almeida, A. T.: The role of demand-side management in the grid integration of wind power. In: Applied Energy 87 (2010), S. 2581-2588.

[3] Torriti, J.; Hassan, M. G.; Leach, M.: Demand response experience in Europe: Policies, programmes and implementation. In: Energy 35 (2010), S. 1575-1583.

[4] Schlomann, B.; Rohde, C.; Eichhammer, W.: Erstellung von Arbeitsbilanzen für das verarbeitende Gewerbe. Karlsruhe: Fraunhofer ISI November 2010

[5] Department of Energy: Benefits of demand response in electricity markets and recommendations for achieving them - A report to the United States Congress pursuant to section 1252 of the energy policy act of 2005. Washington: U.S. Department of Energy 2006.

[6] Faruqui, A.; Sergici, S.: Arcturus: International Evidence on Dynamic Pricing. In: The Electricity Journal 2013.

[7] Klobasa, M.: Dynamische Simulation eines Lastmanagements und Integration von Windenergie in ein Elektrizitätsnetz. Stuttgart: Fraunhofer IRB Verlag 2009.

[8] Faruqui, A.; Hledik, R.; Tsoukalis, J.: The Power of Dynamic Pricing. In: The Electricity Journal 22 (2009) Nr. 3, S. 42-56.

[9] Moghaddam, M.; Abdollahi, A.; Rashidinejad, M.: Flexible demand response programs modeling in competitive electricity markets. In: Applied Energy 88 (2011), S. 3257-3269.

[10] Strbac, G.: Demand side management: Benefits and challenges. In: Energy Policy 36 (2008), S. 4419-4426.

[11] Zehir, M.; Bagriyanik, M.: Demand Side Management by controlling refrigerators and its effects on consumers. In. Energy Convension and Management 64 (2012), S. 238-244.

[12] Ha, D. L.; Joumaa, H.; Ploix, S.; Jacomino, M: An optimal approach for electrical management problem in dwellings. In: Energy and Buildings 45 (2012), S. 1-14.

[13] Alcázar-Ortega, M.; Álvarez-Bel, C.; Domijan, A.; Escrivá- Escrivá, G.: Economic and environmental evaluation of customers' flexibility participating in operation markets: Application to the meat industry. In: Energy 41 (2012), S. 368-379.

[14] Alcázar-Ortega, M.; Álvarez-Bel, C.; Escrivá-Escrivá, G.; Domijan, A.: Evaluation and assessment of demand response potential applied to the meat industry. In: Applied Energy 92 (2012) 84-91.

[15] Daryanian, B.; Bohn, R. E.; Tabors, R. D.: Optimal Demand-side response to electricity spot prices for storage-type customers. In: IEEE Transactions on Power Systems 4 (1989), S. 897-902.

[16] Karwan, M.; Keblis, M.: Operations planning with real time pricing of a primary input. In: Computer & Operations Research 34 (2007), S. 848-867.

[17] Mitra, S.; Grossmannm, I.; Pinto, J. M.; Arora, N.: Optimal production planning under time-sensitive electricity prices for continuous power-intensive processes. In: Computers and Chemical Engineering 38 (2012), S. 171-184.

[18] Yusat, J. M.; Torres, F.; Khodr, H. M.: Optimal methodology for a machining process scheduling in spot electricity markets. In: Energy Conversion and Management 51 (2010), S. 2647-2654.

[19] Emec, S.; Kuschke, M.; Huber, F. W.; Stephan, R.; Strunz, K.; Seliger, G.: Stochastic Optimization Method to Schedule Production Steps According to Volatile Energy Price. In: Seliger, G. (Hrsg.): Innovative Solutions: Proceedings of 11th Global Conference on Sustainable Manufacturing. Berlin, 23.-25. September 2013. Berlin: Universitätsverlag 2013, S. 691-696. ISBN: 978-3-7983-2609-5.

[20] Pechmann, A.; Schöler, I.: Optimizing Energy Costs by Intelligent Production Scheduling. In: Hesselbach, J.; Herrmann, C. (Hrsg.): Localized Solutions for Sustainability in Manufacturing: Proceedings of 18th CIRP International Conference of Life Cycle Engineering. Braunschweig, 2.-4. Mai 2011. Berlin: Springer 2011, S. 293-298. ISBN: 978-3-642-19692-8.

[21] Graßl, M.; Reinhart, G.: Evaluating measures for adapting the energy demand of a production system to volatile energy prices. In: Proceeding of the 21st CIRP Conference on Life Cycle Engineering, Trondheim, 18.-20. Juni 2014, Seite 129-134.

7 BIOGRAPHY

Peter Simon studied Technically-Oriented Business Administration at the University of Stuttgart (B. Sc. 2011) and Technology and Management at the Technical University Munich (M. Sc. 2014). Since 2015 he is a research fellow at the Fraunhofer IGCV in the department of Factory planning and evaluation.

Andrea Hohmann studied aerospace engineering at the University of Stuttgart (graduate engineer 2009). Since 2009 she is a research fellow at the Fraunhofer IGCV. Since 2012 she is

responsible for the life cycle assessment department and since 2018 also for factory planning and evaluation at Fraunhofer IGCV.

Prof. Dr.-Ing. Gunther Reinhart studied Mechanical Engineering at the Technical University of Munich (diploma 1982, PhD 1987), where since 2007 he holds the Chair for Industrial Management and Assembly Technologies at Institute for Machine Tools and Industrial Management iwb. Since 2016 he is executive director of Fraunhofer Research Institution for Casting, Composite and Processing Technology IGCV. A member of supervisory and advisory boards of various companies, he is also chairman of the Bavarian Cluster for Mechatronics and Automation e.V. Gunther Reinhart is author and editor of ten books and 300 journal articles regarding innovations in production technologies, the future of manufacturing systems and intelligent processes.

Analysing the Current Energy Storage Development in South Africa

M. Schloesser[1], J. Niemann[2], C. Fussenecker[3], G. Aschmann[4], R.A.M. Pais[5],
S. Pietrangeli[6], J. Hauser[7]
[1,2,3,4,5]FLiX Research Centre Life Cycle Excellence, Düsseldorf University of Applied Sciences, Düsseldorf, Germany
[6]South African Renewable Energy Technology Centre (SARETEC), Cape Peninsula University of Technology, Cape Town, South Africa
[7]Southern African – German Chamber of Commerce and Industry NPC
Cape Town, South Africa

Abstract
In 2007, Eskom, South Africa's largest producer of electricity, implemented the emergency load shedding for the first time. To avoid future blackouts and negative impacts on South Africa's economic growth, the Independent Power Producer Procurement Programme was established. This Programme promotes the private sector generation, which leads to the diversification of the energy mix and increase of the national electricity capacity. South Africa shows a high renewable energy potential, which is used to support the electricity supply and to comply with the national commitment towards a low carbon economy. The integration of renewable energy presupposes the operation of energy storage technologies. Large scale energy storage solutions are applied to run today's electrical system more efficiently resulting in lower prices, less emissions and more reliable power. To meet the worldwide increasing demand for battery storage technologies, especially used for electrical vehicles, South Africa's mineral resources need to be exploited. This paper assesses the business environment for energy storage production and application in South Africa by applying the PESTLE analysis approach. This method allows to get an overview of the external influences affecting the business environment. The energy storage market is still nascent, as well is an adoption in the regulatory framework needed. If the regulatory framework is altered, South Africa might become a leading country in the production of battery storage systems.

Keywords
Energy, Energy storage, PESTLE analysis

1 INTRODUCTION

South Africa's electricity is supplied by Eskom - the state-owned power company of South Africa (SA) - with approximately 96% of its power. Eskom's installed coal plant capacity of 39 342MW shares 83.3% of Eskom's total installed power capacity of 47 201MW and is significant in the electricity generation [1][2].

The transmission and plant locations of Eskom are covering a minor area and are concentrated in the north-east of South Africa, which leads to a relatively low rural electrification rate of approximately 80% [3] and national electrification rate of 90% [4]. Energy storage systems (ESSs) have the potential to enable the electrification of the leftover rural areas. The Department of Energy's (DoE) Integrated Resource Plan 2010-2030 (IRP) calls for 21,500MW of new renewable energy power capacity to be in place by 2030 to reduce greenhouse gases (GHG), which are mostly contributed by national power generation [5]. Energy storage is defined as a time delay system for energy rather than energy consumer, producer or transmission. It is defined as the 4th pillar of energy systems [6]. ESSs store the excess energy produced when there is low demand for it and release that energy later. With the implementation of renewable energy generation, which may vary throughout the day, it is inevitable to adopt ESSs.

2 FRAMEWORK FOR ENERGY STORAGE IN SOUTH AFRICA

The section assesses the business environment for conventional energy storage production and application in South Africa. In addition, the technical potential for energy storage is elaborated. The assessment is conducted by applying the PESTLE analysis approach. PESTLE is an acronym which stands for the following factors that influences the business environment:

- P – Political
- E – Economic
- S – Social
- T – Technological

- L – Legal
- E – Environment

The analysis is going to take a closer look at the trends and issues relating to these factors, including the likely impact on the further development of the energy storage sector in South Africa.

2.1 PESTLE–Analysis: Political

South Africa is shifting its policies towards sustainable development, including the transition to a greener and more environmentally sustainable economy. The South African National Development Plan (NDP) recognises sustainable development as a critical action to reduce inequality and income poverty and bases its targets around the Sustainable Development Goals. Two out of ten critical actions are investments focusing on energy and interventions to ensure environmental sustainability and resilience to future shocks [7]. The NDP envisages that, by 2030, South Africa's power sector will provide reliable, affordable and sustainable energy by a proportionately less contribution of coal and much larger penetration of gas and renewable energy – especially wind, solar and hydro power. South Africa will need to build 40,000MW of new power capacity to meet the rising electricity consumption and Eskom's plant retirement. Eskom's current committed capacity expansion programme is only adding approximately 10,000MW of new power capacity.

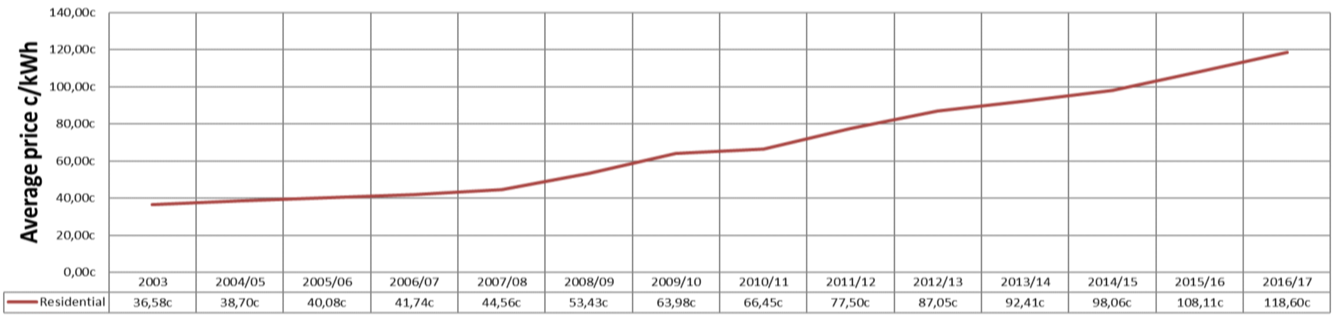


Figure 1 - Eskom's price adjustment from 2003 – 2017 [15]

Currently, South Africa has little initiative to encourage the adoption of energy storage for solar and wind energy because of expensive storage and transmission systems. Therefore, competitively priced energy is needed to exploit mineral resources, which generates the earnings to fund the transformation [8]. In the next 20 years, South Africa will see smarter management of electricity grids, which will open opportunities for more distributed generation systems like renewable energy storage [9]. In the same period, it is predicted that battery storage systems for electrical vehicles (EVs) will grow faster by the promotion of the low-carbon economy, resulting in transport alternatives that minimize environmental harm [10].

Also, the Renewable Energy Independent Power Producer Procurement Programme (REIPPPP) is not specifying a procurement goal or a plan for energy storage, even it has the most proactive plan for SA to integrate renewables. Only Pumped Energy Storage and thermal energy storage to Concentrated Solar Plant is addressed in the DoE's IRP 2010-2030 and it mentions the need for research on energy storage options [11].

2.2 PESTLE–Analysis: Economic

In October 2007, Eskom implemented the emergency load shedding for the first time. Load shedding is used to prevent the electricity power system from a total blackout. Early 2008 Eskom was not able to supply the mines with electricity, so much that they had to be shut down for a week [12]. In 2015, the finance minister, Nhlanhla Nene, stated in an interview that the power crisis slows down SA's economic growth, keeping it under 4% annual growth [13,14].

Energy prices in South Africa have increased sharply in the past ten years and are still growing annually. Between 2003 and 2017, the average price for a kWh increased more than 300% (Figure1).

Eskom had to take this step to secure its financial sustainability and to meet the utility's rising marginal costs for power generation and transmission.

South Africa is rich in mineral reserves, which could be processed into battery electricity storages. To produce Li-Ion batteries, manganese is an essential element of which "South Africa accounts for about 78% of the world's identified manganese resources" [16]. It is forecasted that, with the battery energy storage (BES) "Boom", the demand will increase while the price will decrease by 50-61% by 2030 in stationary applications. The total costs for installed Li-ion batteries for stationary applications will drop between 145 $/kWh and 480 $/kWh, depending on battery chemistry [17].

Besides the application of Li-ion batteries, South Africa has a high potential to produce nickel manganese cobalt oxide batteries with their own mineral reserves and partly nickel cobalt aluminium, lithium manganese oxide, lithium phosphate and lithium titanite batteries. Followed by the flow batteries and especially vanadium redox flow batteries, which have a high potential in SA because of the country's significant vanadium reserves.

2.3 PESTLE-Analysis: Social

The main societal and cultural aspects affecting the utilisation of ESSs are employment creation, acceptance of renewable energy and business practices. Permanent jobs will be needed in areas like: operation and maintenance of energy storage facilities, fabrication and manufacturing of domestically produced materials and equipment. This employment potential has a significant impact in the creation of indirect jobs. Those indirect jobs are necessary to the energy storage value chain and will be created in: construction and installation of energy storage facilities, construction of energy storage manufacturing facilities or subordinate facilities to support mining, as well as refinement and production of materials and equipment. Further jobs will be created in training and skills development. This knowledge transfer is related to the following topics: maintenance of electrical, mechanical or chemical ESSs, installations of control systems, software development and programming. Besides the employment creation energy storage is affecting acceptance of renewables and business practices [18].

The Navigant Research estimates that approximately 225MW of new energy storage power capacity was deployed to the United States during 2016. In the U.S. Energy and Employment Report from 2016, there was a total of 90,831 individuals directly employed in the US energy storage industry. Regarding 2016 as starting point the employment potential of energy storage is estimated at 403.7 jobs per MW of new installed energy storage capacity. The number of jobs per unit of capacity will decrease, as the energy storage industry will grow [19]. This scenario may be applicable to the South African context.

2.4 PESTLE-Analysis: Technology

ESSs are divided into five technology types which are then divided into sub technology types (Table 1). As seen into some research sources, there is a misunderstanding between energy storage and electricity storage, where both terms are used the same but aren't sharing the same meaning. That's why the term energy storage is only used in the next lines.

Technology Type	***Subtechnology Type***
Electro-chemical	Electro-chemical capacitor, lithium-ion battery, flow battery, vanadium redox flow battery, lead-acid battery, metal air battery, sodium-ion battery
Chemical	Hydrogen storage, liquid air energy storage
Pumped hydro storage	Closed-loop pumped hydroelectricity storage, open-loop pumped hydroelectricity storage
Thermal storage	Chilled water thermal storage, concrete thermal storage, heat thermal storage, ice thermal storage, molten salt thermal storage

Table 1 - Energy storage family nomenclature in the United States Department of Energy Storage Database, mid-2017 [20]

Regardless of the debate of when energy storage will be necessary, generally heat storage and pumped hydro storage has already been used for a long time and the deployment of BES is increasing. BES is primarily used as a control and management system of the energy market and capacity reserve to increase self-consumption. Within the next years, there will be a "Boom" for BES because of constantly decreasing costs. Pumped hydro storage (PHS) is mainly used for electric energy time shifts, by storing cheap off-peak electricity and feeding electricity back into the grid during peak in the electricity consumption period.

According to the worldwide BES "Boom", which is expected to take place in the next years, the International Renewable Energy Agency (IRENA) predicted energy storage capacity to triple by 2030, as renewable energy capacity sets to double by 2030. IRENA estimated a total growth from 4.67 TWh of energy storage in 2017 to 11.89-15.72 TWh in 2030. PHS dominates energy storage capacity with approximately 96% in 2017 but the share will fall to 41%-51% by 2030, though the PHS capacity is said to increase by 1,560-2,340 GWh by 2030. While non-pumped hydro energy storage will grow from an estimated 162 GWh in 2017 to 5,821-8,426 GWh in 2030. Because of the small share and capacity of BES in stationary applications in 2017, it will increase from currently estimated 11 GWh to 181-421 GWh, which is growth by a factor of at least 17 by 2030 (Figure 2). This "Boom" in electricity storage will be driven by the rapid growth of utility-scale and behind-the-meter applications like BES market by 2030 [21].

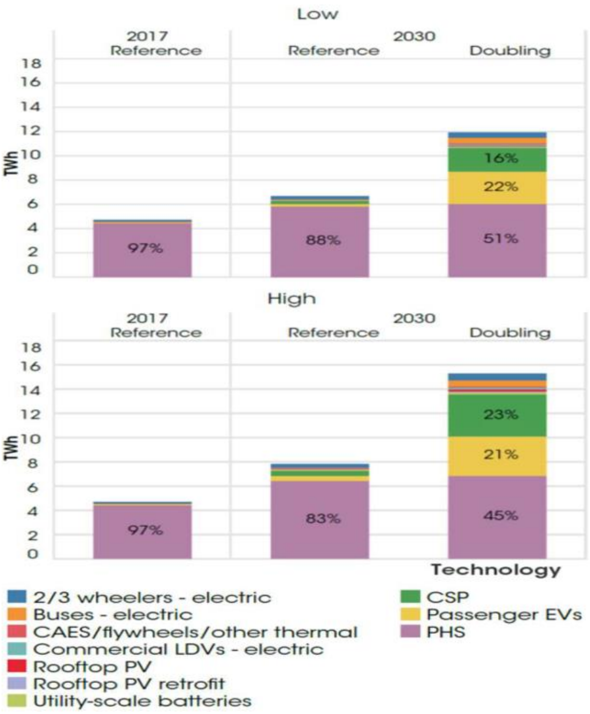


Figure 2 - Energy storage energy capacity growth by source, 2017-2030 [21]

According to Bloomberg's forecasted demand of 1,293 GWh of lithium-ion batteries from EVs in 2030, South Africa and the whole of Africa plays a tiny role. However, South Africa could be leading in battery production because of significant minerals reserves and the growth of the lithium-ion batteries of EVs is increasing exponentially until 2030 (Figure 3).

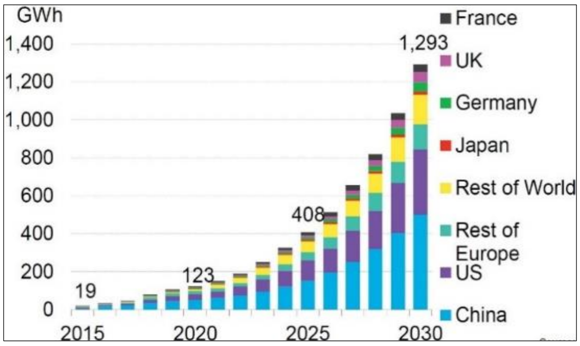


Figure 3 - Forecasted demand for lithium-ion batteries from EVs, 2010-2030 (GWh) [22]

Vanadium redox flow batteries are currently tested as energy storage in South Africa, because of its significant scale-up of vanadium module to the megawatt (MW) scale and discharge durations of 4 to 12 hours and more. Both, Li-ion and flow batteries are the most promising BES in South Africa for the future [23]. PHS and Heat Storage is not evaluated because SA already has significant PHS and Heat Storage resources and potential for the future. South Africa's Li-ion technology research and development pool is built up by the University of the Western Cape, which is the national Li-ion Battery validation facility and is furthermore cooperating with the South African Nuclear Energy Corporation, the University of Limpopo, The Council for Scientific and Industrial Research and the Nelson Mandela University. Other companies and universities are doing research and development as well, those sponsored and supported the SA Energy Storage 2017 conference.

2.5 PESTLE-Analysis: Environmental

This section mentions potential beneficial or harmful environmental impacts in South Africa due to adopting energy storage systems.

Section 24 in the Constitution of the Republic of South Africa is the legal reference source for environmental laws. It ensures a sustainable development of the social and biophysical environment. Depending on the planned energy storage technology, its characteristics vary in terms of the material used, use, recyclability as well as disposal. With the different characteristics between the technologies, it's necessary to consider that the environmental impact and legislative requirements must be proved for each technology. Applicable legislation can, for example, be found in the National Environmental Management Act (No. 107 of 1998) (NEMA), the National Impact Assessment Regulations, 2014, National Environmental Management: Waste Act, 2008 and others.

The rapidly increasing demand of BES and the mining of the involved resources leads to a change in the environment. Crucial legislation is determined in the Mineral and Petroleum Resources Development Act (No. 28 of 2002). Mining has a direct impact on the biodiversity, which is estimated to contribute 7% of South Africa's GDP and supports over a million jobs [24]. It is essential to prove the impact on biodiversity in the life cycle of a mining project. Additionally, BESs have a high risk to harm the environment if they leak their hazardous substances.

South Africa´s annual CO2 emission was at 468 Mt in 2016 and was therefore placed 13th in the world ranking. 393 Mt of CO2 is produced due to the coal consumption [25]. According to BP Statistical Review of World Energy 2017, 70% of South Africa's primary energy consumption came from coal [26]. The IRP was developed by the DoE and describes strategies as well as actions concerning the electricity supply until 2030. The IRP stipulates to integrate a capacity of 17,8 GW using solar photovoltaic (8.4 GW), wind energy (8.4 GW) and concentrated solar power (1 GW) by 2030 [27].

ESSs improve the overall energy management by relieving the base load generation. Additionally, ESSs can be used to cover peak generation, instead of using generators running on fossil fuel. The improved energy management leads to more efficient electricity usage, which consequently reduces greenhouse gas emission.

2.6 PESTLE-Analysis: Legal

The IRP 2010-2030 is the main guiding policy in terms of policies and regulations of the electricity sector in South Africa. Guiding government departments are the DoE, the National Energy

Regulator of South Africa and local municipalities [28]. In addition to that, ESKOM plays a significant role as the main power supplier in changing legislation in favour of ESS. The DoE published the White Paper on Renewable Energy in 2003, mentioning that "cost-effective energy storage mechanisms [have been] investigated" [29]. This Whitepaper encouraged the development of renewable energies. Furthermore, "the Renewable Energy Independent Power Producer Procurement Programme has the most proactive plan for SA to integrate renewables but with no specific procurement goal or plan for energy storage" [30].

Neither the IRP nor the REIPPPP mention energy storage in relation to the installation of renewable energies. In Germany, for example, it is being discussed that energy storage needs to be considered independent from generation, consumption or transmission [31]. The Department of Science and Technology announced to develop an energy storage roadmap, which "will support the alignment of activities at individual science institutions across South Africa" [32].

3 CONCLUSION

For a decade now, Eskom is showing huge deficits in supplying reliable electricity to its consumers. This circumstances led to a negative impact on South Africa's economic growth, pushing the government to implement different legal actions to diversify its electricity market. In addition, South Africa strives for a lower GHG emission. The change from conventional fuels to renewable energies takes ESS for granted, because of its capability of time shifting energy. At the same time, ESS could be used to become independent from the conventional grid or even electrify rural areas, which can be cheaper than extending the electricity network to cover those areas.

South Africa has a high potential to be the leading producer of Li-ion batteries, because of its mineral resources. According to the BES "Boom" and the rapid cost decrease, Li-ion-based energy storage systems are expected to be the leading energy storage technology for commercial, industrial and home consumer utility-scale applications with cycle durations up to 4 hours. It can be assumed that South Africa's economy would increase due to the increasing energy industry as well as mining industry. On the one hand, GreenCape states that even with the dropping battery and generation technology costs, the combination of these two technologies is still more expensive than what customers pay for the power from the conventional grid [33]. On the other hand, the International Finance Corporation mentions that "South Africa is expected to be the largest market in the [Sub-Saharan] region for energy storage" [34].

The advantages of implementing ESSs are obvious, but South Africa's legislative framework shows huge deficits regarding the use and application of this technology. It is recommended to check the best practices overseas and adopt it to the South African context.

4 REFERENCES

[1] Eskom Generation and Corporate Affairs. 2017. Generation Plant Mix [Online]. Available: http://www.eskom.co.za/AboutElectricity/FactsFigures/Documents/GX0001GenPlantMixRev19.pdf [Accessed: 20 March 2018]

[2] Eskom Generation Communication Department. 2013. Eskom Power Stations in South Africa [Online]. Available: http://www.eskom.co.za/Whatweredoing/ElectricityGeneration/PowerStations/Pages/Map_Of_Eskom_Power_Stations.aspx [Accessed: 13 April 2018].

[3] U.S. Department of Commerce. 2017. South Africa - Electrical Power Systems. South Africa Commercial Guide [Online]. Available: https://www.export.gov/article?id=South-Africa-electrical-power [Accessed:22 March 2018].

[4] DoE. 2017. Annual Report 2016/17, p.2.

[5] National Planning Comission. 2012. National Develoment Plan 2030: Our Future – make it work. Department: The Presidency Republic of South Africa, p. 163.

[6] BVES. 2017. Faktenpapier Energiespeicher. Bundesverband Energiespeicher e.V. Berlin p. 1.

[7] National Planning Comission. 2012. National Develoment Plan 2030: Our Future – make it work. Department: The Presidency Republic of South Africa, p. 34.

[8] National Planning Comission. 2012. National Develoment Plan 2030: Our Future – make it work. Department: The Presidency Republic of South Africa, p. 198.

[9] National Planning Comission. 2012. National Develoment Plan 2030: Our Future – make it work. Department: The Presidency Republic of South Africa, p. 170.

[10] National Planning Comission. 2012. National Develoment Plan 2030: Our Future – make it work. Department: The Presidency Republic of South Africa, p. 183.

[11] Parsons. 2017. South Africa Energy Storage Technology and Market Assessment. U.S: Trade and Development Agency, p. 452.

[12] Eskom. 2018. ESKOM 2000 - 2008 - Our Recent Past -"Shift performance and grow sustainably" [Online]. Available: www.eskom.co.za/sites/heritage/Pages/2000.aspx [Accessed: 17 April 2018].

[13] Financial Times. 2015. Power crisis knocks South Africa's economy [Online]. Available: https://www.ft.com/content/e853a048-bce7-11e4-9902-00144feab7de [Accessed: 18 April 2018].

[14] The World Bank. 2018. GDP growth (annual %) adapted by Google [Online]. Available: https://www.google.com/publicdata/explore?ds=d5b

ncppjof8f9_&met_y=ny_gdp_mktp_kd_zg&hl=de&dl=de#!ctype=l&strail=false&bcs=d&nselm=h&met_y=ny_gdp_mktp_kd_zg&scale_y=lin&ind_y=false&rdim=region&idim=country:ZAF&ifdim=region&tstart=975884400000&tend=1512342000000&hl=de&dl=de&ind=false [Accessed 04 December 2018].

[15] Eskom. 2018. Tariff history [Online]. Available: www.eskom.co.za/CustomerCare/TariffsAndCharges/Pages/Tariff_History.aspx [Accessed 26 April 2018].

[16] U.S. Geological Survey. 2018. Mineral Commodity Summaries 2018. U.S. Geological Survey, p. 105.

[17] IRENA. 2017. Electricity Storage and Renewables: Costs and Markets to 2030. International Renewable Energy Agency, p. 18.

[18] Parsons. 2017. South Africa Energy Storage Technology and Market Assessment. U.S. Trade and Development Agency, p. 340.

[19] Dehamna, A. 2017. Energy Stroage Industry Jobs Linked to Energy Storage Capacity [Online]. Available: https://www.navigantresearch.com/blog/energy-storage-industry-jobs-linked-to-energy-storage-capacity [Accessed 23 March 2018]

[20] IRENA. 2017. Electricity Storage and Renewables: Costs and Markets to 2030. International Renewable Energy Agency, p. 29.

[21] IRENA. 2017. Electricity Storage and Renewables: Costs and Markets to 2030. International Renewable Energy Agency, p. 13 - 16.

[22] Curry, C. 2017. Lithium-ion Battery Costs and Market. Bloomberg New Energy Finance, p. 4.

[23] Parsons. 2017. South Africa Energy Storage Technology and Market Assessment. U.S. Trade and Development Agency, p. 36.

[24] Department of Environmental Affairs. 2013. Mining and Biodiversity Guideline: Mainstreaming biodiversity into the mining sector, p. 7.

[25] Global Carbon Atlas. 2017. CO2 Emissions [Online]. Available: www.globalcarbonatlas.org/en/CO2-emissions [Accessed: 2 Mai 2018].

[26] British Petroleum Company. 2017. BP Statical Review of World Energy June 2017. London: British Petroleum Co.

[27] DoE. 2013. Integrated Resource Plan for Electricity (IRP) 2010-2030 [Online]. Available: www.doe-irp.co.za/content/IRP2010_updatea.pdf [Accessed: 3 Mai 2018].

[28] GreenCape. 2017. Energy Service: Market Intelligence Report 2017. Cape Town: GreenCape, p. 21-23.

[29] Department of Minerals and Energy. 2003.White Paper on Renewable Energy, p.34.

[30] Parsons. 2017. South Africa Energy Storage Technology and Market Assessment. U.S. Trade and Development Agency, p. 451.

[31] BVES. 2017. Faktenpapier Energiespeicher. Bundesverband Energiespeicher e.V. Berlin p. 28.

[32] Department of Science and Technology. 2017. Annual Report 2016-2017, p. 20.

[33] GreenCape. 2017. Energy Service: Market Intelligence Report 2017. Cape Town: GreenCape, p. 34.

[34] Navigant Research. 2017. Energy Storage Trend and Opportunities in Emerging Markets. International Finance Corporation, p. 32.

5 BIOGRAPHY

Martin Schloesser holds a B.Eng. with specialization in Product Development and Production from the University of Applied Sciences Duesseldorf. He is currently a scientific assistant in the Department of Mechanical and Process Engineering, focusing on production simulation and Life Cycle Management.

Jörg Niemann obtained his PhD degree in Manufacturing Engineering from the University of Stuttgart. In 2012 he was appointed Professor in Business Management and Mechanical Engineering at the Düsseldorf University of Applied Sciences, Germany.

Claudia Fussenecker holds a M.A. in European Management from the FH Bund in Brühl and a Diploma in Industrial Management from EUFH in Brühl. She is currently a scientific assistant in the Department of Mechanical and Process Engineering, focusing on Marketing and Services and Life Cycle Management.

Effect of Stress Relieving Heat Treatment on High Strain Rate Tensile Properties of Direct Metal Laser Sintered Ti6Al4V (ELI) Parts

A.M. Muiruri[1], M. Maringa[2], W.B. du Preez[3] & L.M. Masu[4]

[1 2 3] Department of Mechanical Engineering and Mechatronics, Central University of Technology, Free State, Bloemfontein, South Africa.

[4] Department of Mechanical Engineering, Vaal University of Technology, Vanderbijlpark, South Africa.

Abstract

One of the major problems associated with additive manufacturing processes particularly – Laser Powder Bed Fusion (LPBF) such as the Direct Metal Laser Sintered (DMLS) process, arises due to the short interaction time of the laser beam with the powder and the accompanying high localised heating and subsequent rapid cooling. The result of this is a build-up of residual stresses in manufactured components, which necessitates subsequent heat treatment. This paper documents an investigation of the high strain rate tensile properties of as-built (AB) and stress relieved (SR) DMLS Ti6Al4V (ELI). The high strain rate tensile tests were carried out using a Split Hopkinson Pressure Bar (SHPB) system at ambient temperatures at average strain rates of 250s^{-1} and 360s^{-1}. Values of Vickers micro hardness were obtained before and after the tensile tests. In addition, SEM investigations were carried out to study the prevailing failure mechanism and to identify any distinct features on the resulting fracture surfaces. The results of the high strain rate tensile testing showed rate sensitivity of the yield stress, flow stresses and true fracture strain for both forms of the alloy. Sensitivity to strain rate for the SR samples was seen to be higher than for the AB samples at all strains. Surface defects were found to be related to the tensile fracture of the alloy.

Keywords

DMLS, Ti6Al4V (ELI), High strain rate.

1. INTRODUCTION

Additive Manufacturing (AM) first emerged commercially in 1987, with stereolithography (SL) from 3D Systems, a process that solidifies thin layers of ultraviolet (UV) light-sensitive liquid polymer using a laser. It was originally known as rapid prototyping using polymers but became a manufacturing technology with time [1]. In AM, material is added, one cross-sectional layer at a time, to create a three dimensional object [2]. AM includes technologies such as fused deposition modelling, laser micro-sintering, Direct Metal Laser Sintering (DMLS), 3D laser cladding, 3D laser deposition, Electron Beam Melting (EBM), and Electron Beam Sintering (EBS). These AM processes are stimulating innovation in engineering component design, enabling the manufacturing of complex parts that are difficult to produce through the traditional methods such as forging, casting, machining and joining. This flexibility in design has led to the technology being used to manufacture topologically optimized parts with improved strength to weight ratios. AM also offers near-net shape fabrication of parts directly from 3D CAD models with minimal machining, resulting in a reduction in lead time, waste and cost. The two key advantages of AM that are very fundamental in the design of parts for the aerospace and automobile sectors include reduction of the overall weight, of the vehicles and aircraft as well as primarily reduction of the "Buy to Fly" ratio in aircraft. The "Buy to Fly" ratio is the ratio of the weight of materials procured for a particular product to the weight of the finished product [3].

Ti6Al4V alloy components are designed for good balance of mechanical characteristics such as high specific strength, ductility and resistance to corrosion, making it a standard alloy for use in the aerospace and medical industries. Parts in this alloy for use in these two industries have of late being processed increasingly using the DMLS process. This is because of the associated benefits of the process such as those mentioned in the preceding paragraph. The major challenges of this process that have been reported include: thermal residual stresses developed during the build process; high surface roughness of the manufactured parts; and finally generation of martensitic microstructures due to rapid cooling [4, 5]. Hence, the material properties of DMLS Ti6Al4V parts are often very different from those of wrought or cast Ti6Al4V parts that are produced at lower cooling rates. For instance, the yield strength of wrought and cast Ti6Al4V are 945 MPa and 885 MPa respectively, while that of DMLS Ti6Al4V has been reported as 1075 MPa [6-8].

The high strain rate (dynamic) properties of the conventional (wrought and cast) Ti6Al4V have been a subject of various studies [9-11]. Generally these studies have demonstrated positive strain rate sensitivity of Ti6Al4V. More recent studies have attempted to demonstrate high strain rate performance of AM Ti6Al4V [12-14]. The work of Peng.*et.al* [12], paid special attention to the coupled effects of the high strain rate and high temperatures on compressive and tensile plastic flow behaviour and fracture characteristics of Ti6Al4V alloys prepared by the 3D laser deposition technology. The study observed that the flow stress under compressive loading increased with increase in the strain rate, whereas the flow stress under tensile loading decreased with increase in the strain rate. These researchers proposed this anomalous strain rate effect on tensile flow stress when compared to compressive flow stress to be a result of fusion pores that were observable in the undeformed samples when examined in a scanning electron microscope (SEM). Peng *et.al* [12], observed that pores arising from lack of fusion that were present in the alloy expanded upon dynamic tensile loading but tended to close under dynamic compressive loading. However, the material exhibited negative temperature rate sensitivity under both loading conditions. Elsewhere, Masood *et.al* [13], studied the dynamic compressive properties of Ti6A4V alloys processed by the EBM method. The study found that the fracture strain was lower in dynamic compression in comparison to the static compression deformation. During subsequent investigation of the microstructure, adiabatic shear bands (ASBs) were found in the samples that were tested at high strain rates. Biswas *et al.* [14] studied the deformation and fracture behaviour of Ti6Al4V produced by the laser-engineered–net shaping (LENS) technique under dynamic and static compression. Both the flow and fracture stresses in this work exhibited appreciable rate sensitivity.

Research on the behaviour of DMLS Ti6Al4V at high strain rates is scant. Moreover, the available literature on the behaviour of AM Ti6Al4V at high strain rate is predominantly focused on compressive loading. This paper presents a study of the deformation and fracture of as-built (hereinafter referred to as AB) and stress relieved (hereinafter referred to as SR) DMLS fabricated Ti6Al4V (ELI) subjected to high strain rate tensile loading using the Split Hopkinson Pressure Bar (SHPB) test. The relative strain rate sensitivity of the flow stress for the two forms of alloy (AB and SR) are compared. In addition, the variation in micro hardness before and after high strain testing are measured and discussed. The resulting deformed and fractured surfaces are analysed using SEM and then discussed.

2. MATERIALS AND EXPERIMENTAL PROCEDURES

2.1 Materials

Ti6Al4V (ELI) powder complying with the ASTM F3001-14 standard, supplied by TLS Technik GmbH, was used to build the high tensile strain rate test specimens using the DMLS. The test specimens were fabricated using an EOSINT M280 DMLS machine which is equipped with a 200 W ytterbium fibre laser that has a beam diameter of 80 μm. The samples were fabricated with a laser power setting of 175 W, a hatch distance of 100 μm and a layer thickness of 30 μm. The Ti6Al4V (ELI) powder particles were of an average diameter of <40μm. Layers were scanned following a back-and –forth raster pattern which was rotated 67º between successive layers. The gauge sections of the tensile specimens were rectangular in shape with values of length, width and height of (10x4x4) mm, respectively. The shoulders of the specimens were printed with threads to mesh with the grips used in the SHPB apparatus at the Blast Impact and Survivability Research Unit (BISRU) of the University of Cape Town. The dimensions and profile of the shoulders as well as the build orientation of the specimens are shown in Figure 1.

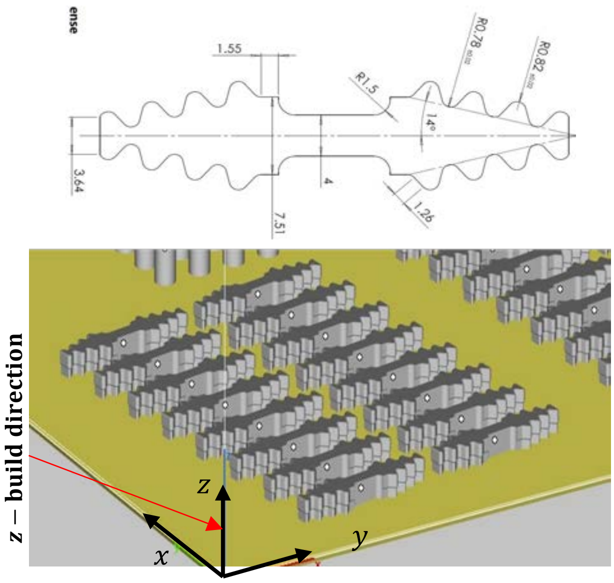


Figure 1 - The dimensions and the build orientation of the high strain rate tensile specimens.

A total of 32 specimens were fabricated, half of which were heat treated to reduce the residual stress created during fabrication. The alloy was heat treated (stress relieved) in argon atmosphere at a temperature of 650 ºC for a soaking period of 3 h and then air cooled.

2.2 Experimental procedures

The dynamic tensile tests were performed using a SHPB test facility with striker, incident and transmitter bars that were made of 7075-T651 aluminium. The shoulders of the specimens were connected through the mechanical joints to one end of the input bar, while the other shoulder was connected to the nearer end of the output bar. The strain rates were controlled by changing the velocity of the striker bar. The SHPB facility used was fairly limited in terms of the velocity range in which it could be operated safely. Nonetheless, two strain rates of 250 s^{-1} and 360 s^{-1} were attained. During the experiment, the test under each strain rate was repeated seven times to enhance credibility and statistical confidence. An extra specimen for each form of the alloy was used to calibrate the equipment prior to the commencement of testing. The reflected signal $\varepsilon_r(t)$ identified from the strain wave of the incident bar and transmitted strain signal $\varepsilon_t(t)$ identified from the strain wave of transmitter bar were used to compute stresses, strains and strain rates using the following equations:

$$\sigma_s = \frac{E A_0 \varepsilon_t}{A_s} \tag{1}$$

$$\varepsilon_s = 2\frac{C_0}{l_s}\int_0^t \varepsilon_r \, dt \tag{2}$$

$$\dot{\varepsilon_s} = \frac{2\varepsilon_r C_0}{l_s} = \frac{v_{s1}-v_{s2}}{l_s} \tag{3}$$

Where the symbols σ_s, ε_s and $\dot{\varepsilon_s}$ represent engineering stress, strain and strain rate of the specimen, respectively. The symbols A_s and A_0 stand for the cross sectional areas of the specimen and the bars, respectively, v_{s1} and v_{s2} represent the input and output velocities at the interface of the specimen-transmitter bars, respectively, l_s the length of the specimen and C_o the longitudinal wave velocity. The curves of stress-strain that were obtained in all instance were averaged using a MatLab code that was developed in the present work and then converted to curves of true stress-strain.

The values of micro hardness for the samples that were not loaded and those that were exposed to tensile high strain rate loading without experiencing fracture were obtained using a Future Tech Vickers hardness tester. A 200 g load, with a dwell time of 10s was used and a minimum of 30 indentations were made on each sample. The indentations were done approximately 0.5 mm from the edges of the cut surfaces of the gauge sections in both the transverse and longitudinal directions and another set of indentations was done along the middle of the specimen, in the loading direction. The deformed longitudinal cut surfaces to be studied using SEM, were then mounted in multfast resin (Bakelite resin) using a Struers Citopress-1 mounting machine. Grinding and polishing of the specimens then followed. Thereafter, the polished surfaces were cleaned individually under tap water, and then dried with a strong stream of compressed air. The samples were then etched using a solution of Kroll's reagent. The fracture surfaces to be studied were subsequently cleaned for 5 min in an ultrasonic cleaner using ethanol as the cleaning solvent.

3. RESULTS AND DISCUSSION

3.1 Stress- strain behaviour

At the average strain rate of 360 s^{-1}, 100% of both forms of the alloy deformed to fracture, while at the strain rate of 250 s^{-1} only 71% and 84% of the tested samples for the AB and SR specimens, respectively, deformed to fracture. Figure 2 shows a typical stress, strain and strain rate against time graph of the signals that were obtained in these tests. The graph shows the tensile stress to be initially almost constant up to approximately 40 μs (which is taken as the 'ring up' period), followed by a rapid increase in value up to the yield point, and then a plateau before sharply dropping to zero. The 'ring up' period in a SHPB test is considered as the period of time taken to attain dynamic equilibrium between the forces at the input bar and the output bar of the SHPB test. It is dependent on the gauge length of the specimen such that shorter gauge lengths reduce the period and vice versa. The tensile strain is also initially almost constant during the 'ring up' period and then increases rapidly in magnitude, drops in magnitude and then attains a more or less constant slope at approximately 150 μs.

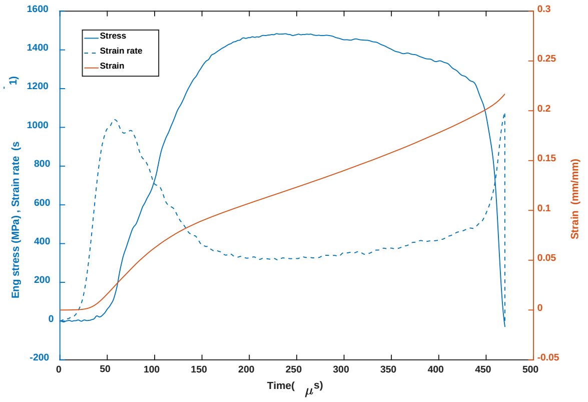


Figure 2 - The tensile stress, strain rate and strain against time curves for the tested SR DMLS Ti6Al4V (ELI) specimens.

That the tensile strain rate increases slowly during the ring up period, is evident from the nearly horizontal curve at the onset of loading. The strain rate then increases sharply before dropping sharply in magnitude, then levels off in the zone of plastic deformation before picking sharply and then finally terminating in a rapid reduction of strain rate.

Figure 3 shows the engineering stress–strain curves for the seven AB specimens of DMLS Ti6Al4V (ELI) that were tested at the same strain rate.

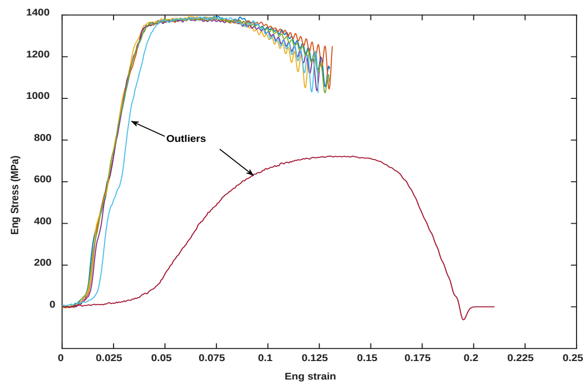


Figure 3 - Stress-strain curves for the seven AB specimens that were tested at an average strain rate of 250 s^{-1}.

The two curves in light blue and red colours shown in Figure 3 were considered as outliers, as they clearly lacked conformity of magnitude for both curves and profile for the red curve, with the other curves and were not therefore taken into account when averaging the curves. Moreover, though the blue curve exhibits a shape profile very similar to the clustered curves, its ring up period was significantly higher (20% higher) than for the clustered curves. Therefore, the two curves in this case were not taken into account when averaging the curves. The averaged curves were converted to true stresses and strains using the following relationship:

$$\sigma_{true}(t) = \sigma_{eng}(t)(1 - \varepsilon_{eng}(t)) \quad (4)$$

$$\varepsilon_{true}(t) = In(1 - \varepsilon_{eng}(t)) \quad (5)$$

The resultant curves of average true stress-strain and the corresponding curves for the maximum and minimum values (dotted) in the plastic range are shown in Figures 4 and 5.

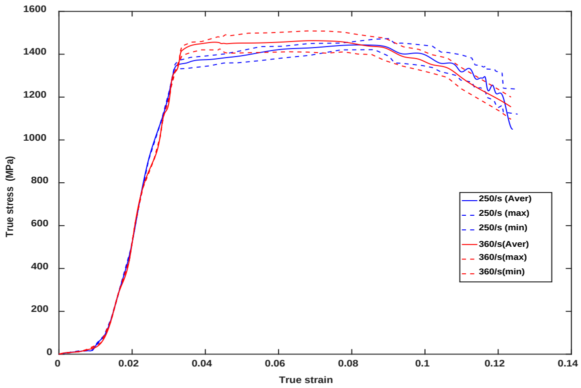


Figure 4 - The resultant curves of average true stress-strain for the AB samples at two different strain rates.

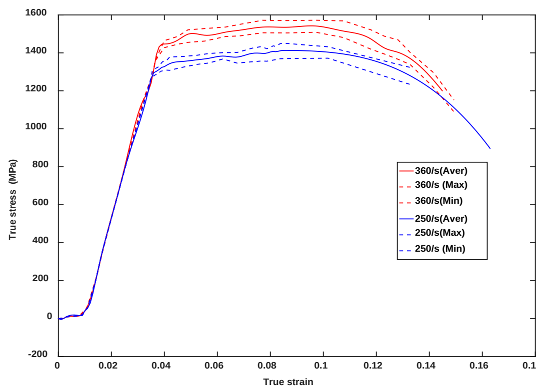


Figure 5 - The resultant curves of average true stress-strain for the SR samples at two different strain rates.

The sudden impact on the specimens during the tensile SHPB tests is suspected to have led to instantaneous strain hardening (due to piling up of dislocations). Therefore, yielding at stresses that were higher than the known tensile yield stress of 1098 MPa were found for the two forms of the alloy as is evident in Figures 4 and 5. The yield stress for the AB samples is seen in Figure 4 to be 1443 ± 17MPa MPa and 1360±19 MPa, while that of SR specimens in Figure 5 is 1446 ± 22MPa and 1340 ± 21Mpa for strain rates of 360 s^{-1} and 250 s^{-1}, respectively.

The effect of the strain rate on the flow stress (σ) at a fixed strain and temperature is described by the power law expressed as follows:-

$$\sigma = C\dot{\varepsilon}^{m} \quad (6)$$

Where the symbols *m* stands for the strain rate sensitivity, $\dot{\varepsilon}$ the strain rate and *C* a material constant. The relative flow stresses at the two strain rates measured at the same strain is given by [15]:

$$\sigma_2/\sigma_1 = (\dot{\varepsilon}_2/\dot{\varepsilon}_1)^{m} \quad (7)$$

The evaluated values of relative strain rate sensitivity *m* at selected strains for the AB and SR samples at the two strain rates are summarised in Table 1.

Strain	**0.04**	**0.06**	**0.08**	**0.1**	**0.12**
	Tensile strain rate sensitivity				
AB	0.14	0.03	0.015	-0.034	-0.06
SR	0.19	0.16	0.21	0.22	0.19

Table 1 - Tensile strain rate sensitivity

The values in Table 1 show the SR specimens to have higher values of strain rate sensitivity of flow stress than the AB specimens. The post yielding curves for the AB specimens in Figure 4, at the strain rate of 250 s^{-1} are steeper than at the higher strain rate of 360 s^{-1}. This was suspected to be a result of the low rate of strain hardening (flatter curve) expected after yielding at higher strain rates than at lower strain rates. The two curves in the figure are then seen to cross over each other at a strain of 0.084 (indicating a higher rate of unloading at the higher strain rate), after which the material shows negative strain rate sensitivity at both strain rates. The two forms of the alloy also show true fracture strain rate sensitivity of the fracture strain as is evident in Figures 4 and 5. Figure 4 shows true fracture strains for the AB samples to be very close to values of 0.1210 and 0.1236 at the strain rates of 250s^{-1} and 360s^{-1}, respectively. The fracture strain for the SR samples that tested at the strain rate of 250 s^{-1}, is seen in Figure 5 to be equal to 0.1633, while at 360s^{-1} it is slightly lower at 0.1401. The fracture strains for the SR specimens are therefore seen to be the higher of the two forms of specimens, which ties in with the fact that the SR specimens were stress relieve annealed and should therefore be more ductile.

3.2 Micro-hardness

The test was conducted in order to study, not only the relationship of the form of material (AB or SR) with hardness but also to understand the changes of hardness in the test samples as a result of the imposition of high tensile strain rates on them. Figure 6 summarizes the micro hardness values of mean and standard deviation before and after the application of high tensile strain rates. Stress relieving resulted in a reduction in of micro hardness of the AB samples as is seen in Figure 6.

A hypothesis of the reduction by stress relieving, of the defects of higher density and a large amount of residual stress created in DMLS process due to rapid solidification can be proposed to have resulted in this observation. Even though the values of standard deviation in Figure 6 show the mean values to overlap in each form of the samples, the difference in the mean values are much higher for the SR than AB samples. The AB and SR specimens recorded a 2.5% and 4.9% increase in hardness, respectively, at the same strain rate of 250 s^{-1} with reference to the recorded mean values for samples that were not loaded. Two factors could be surmised to have contributed to this increase in hardness.

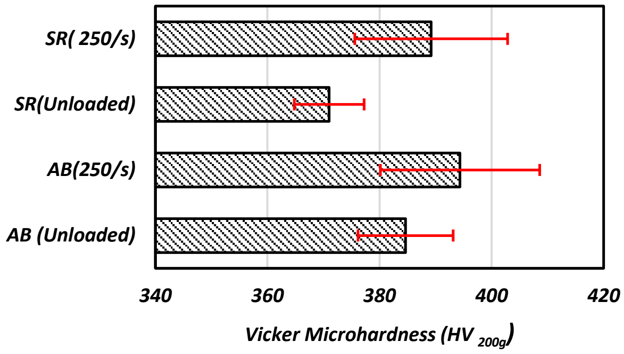


Figure 6 - Comparison of mean values of Vickers micro hardness for the unloaded and loaded DMLS Ti6Al4V (ELI) at a strain rate of 250 s^{-1}.

Firstly, it is suspected that the specimens experienced almost instantaneous strain hardening upon imposition of high strain rates. Secondly, upon yielding and before necking the specimens experienced further strain hardening.

3.3 Microstructural investigation and analysis of failure surfaces

Optical images of the microstructure of the specimens in the AB and SR forms, prior to loading, are shown in Figure 7.

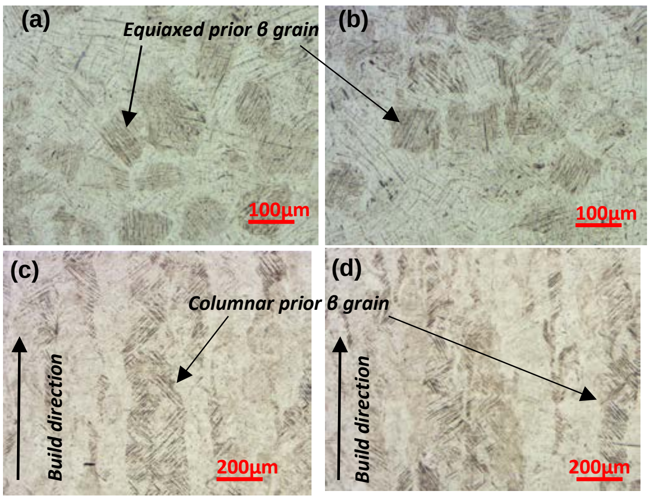


Figure 7 - Optical micrographs of (a) and (c) the AB and (b) and (d) SR samples for the transverse and longitudinal sections with reference to the build direction, respectively.

The images reveal the presence of prior β grains elongated approximately parallel to the build direction, with an almost equiaxed grain morphology in the transverse sections, both before and after heat treatment. The internal columnar/equiaxed grains seen in the micrographs consist of a fine acicular -type structure which is referred to as α'-martensite. There is no evidence of a change in microstructure after the stress relieving heat treatment, which is consistent with the work of Moletsane *et al.* [7]. Moletsane *et al.* further found that, even though the microstructure of the AB and SR Ti6Al4V (ELI) alloy was similar, the UTS, % fracture strain and yield stress of the former decreased, increased, and remained unchanged, respectively, upon stress relieving.

The study and analysis of the deformed surfaces at lower strain rate that were not loaded till fracture, revealed cracks propagating at an angle of inclination, from the necked surface towards the centre of the cut surfaces as shown in Figure 8. Two proposals can be made to explain this phenomenon: the first is due to stress inhomogeneity in the material after necking and the second one is due to surface roughness of the DMLS parts. These two factors combined could increase the strain localization near the surface, thereby triggering the initiation and propagation of cracks.

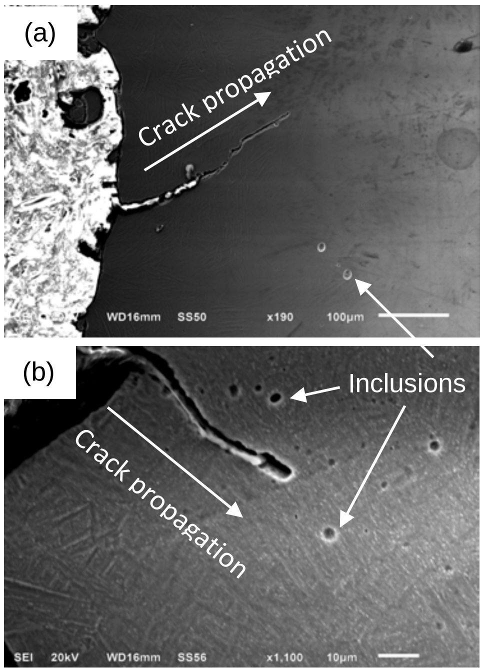


Figure 8 - Typical cracks that have propagated at inclinations to the loading axis on the deformed surface of (a) AB and (b) SR specimens.

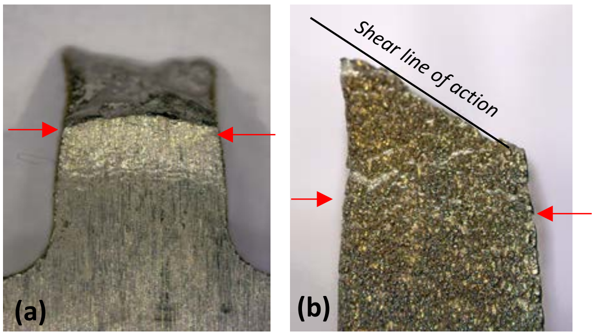


Figure 9 - Macroscale views of the typical DMLS Ti6Al4V (ELI) fractured samples under high strain rate in tension (a) width and (b) thickness.

The loaded specimens showed necking in both the directions of width and thickness of the AB and SR specimens as shown with the red arrows in Figure 9. The side views shown in this figure demonstrate that the samples failed at an angle of approximately 45º to the loading axis, which is the angle of maximum shear in this case. This further explains the inclination of propagated cracks shown in Figure 8.

The tested AB and SR samples showed similar characteristic features on the fracture surfaces. The surfaces consisted of smooth/shear and fibrous regions, which are characterised by elongated dimples as shown in Figure 10.

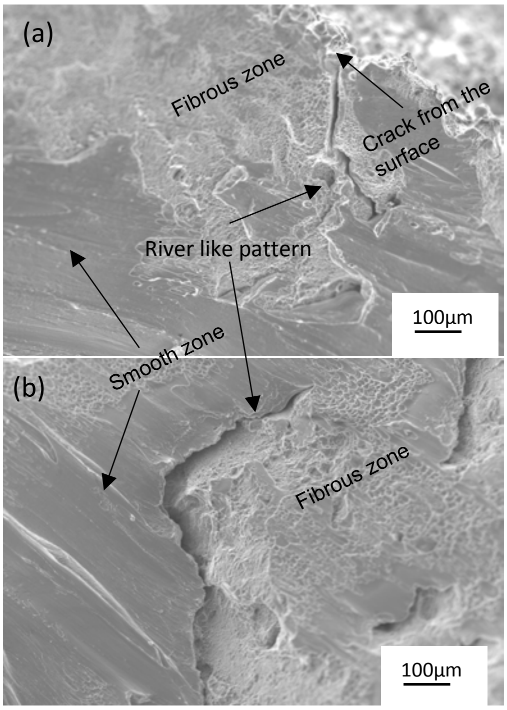


Figure 10 - Characteristic features of the (a) AB and (b) SR fracture surfaces.

In a typical case of uniaxial high strain rate loading there may exist multiple planes of shear stresses, however fracture will occur along the plane experiencing the maximum shear stress. Hence failure on planes other than the plane of maximum shear may be evident as secondary cracks. The presence of such secondary cracks running internally from the edges of the specimens are evident in the micrographs in Figure 10. Often, the subsequent coalescing of multiple cracks and voids as the crack propagates, creates characteristic "river like patterns" on the fracture surface. The ductile fracture mechanism along these "river like patterns" is further revealed by the presence of dimples, as result of the coalescences of micro voids, which further explains the existence of fibrous regions on the primary fracture surface. The shearing action between the two opposite fracture surfaces of the test specimens resulted in smoothening of the surfaces.

4. CONCLUSIONS

The behaviour of DMLS Ti6Al4V (ELI) under high strain rate loading was studied using the SHPB test with the following outcomes.

- The flow stresses for the SR specimens were observed to increase at higher strain rates, while the true strain decreased. The flow stresses for the AB samples increased with strain rate initially, but then decreased at a strain of 0.084. The AB specimens however, did not show sensitivity of fracture strain to strain rate.
- The relative strain rate sensitivities of the SR specimens were higher than those of the AB specimens at the two strain rates used.
- Stress relieving resulted in a reduction in Vickers micro hardness of the AB parts. The micro-hardness was seen to increase after deformation at high strain rates.
- The fracture surfaces were characterised by smooth and fibrous regions and 'river like' patterns of crack propagation.
- The secondary cracks that were observed to have initiated from the surfaces are suspected to have been due to surface roughness of the test specimens that is associated with the DMLS process.

Future works should aim at testing both AB and SR specimens over a wider range of high tensile strain rates.

5. ACKNOWLEDGMENT

The active support and funding of the South African Department of Science and Technology through the Collaborative Program in Additive Manufacturing, is gratefully acknowledged. The authors also acknowledge the Centre for Rapid Prototyping and Manufacturing at the Central University of Technology, Free State for building the DMLS test specimens. The Blast Impact and Survivability Research Unit (BISRU) of the University of Cape Town is also acknowledged for providing the facilities that were used for testing.

6. REFERENCES

1. Wohlers Report, 2014, '3D Printing and Additive Manufacturing State of the Industry', *Annual 666*

Worldwide Progress Report, Wohlers Associates.

2. Ivanova, O., Williams, C., Campbell, T., 2013, Additive manufacturing and nanotechnology: Promises and challenges. *Rapid Prototype Journal, vol. 19, pp.353–364.*
3. Oakridge national laboratory, 2010, Aerospace workshop planning session summary report, Department of energy efficiency and renewable energy.
4. Albert E. Patterson, Sherri L. Messimer, Phillip A. Farrington, 2017, **'**Overhanging Features and the SLM/DMLS Residual Stresses Problem', *Review and Future Research need, Additive manufacturing and technologies. 5*(2), pp.15 doi:*10.3390/technologies5020015*
5. Rafi H.K., Karthik N.V., Gong H., L. Starr, Stucker E., 2013, 'Microstructures and mechanical properties of Ti6Al4V parts fabricated by selective laser melting and electron beam melting," *Journal of Materials Engineering and Performance, vol. 22(12), pp. 3372-3383.*
6. Todd M. Mower, Michael J. Long, 2015, 'Mechanical behaviour of additive manufactured, powder-bed laser-fused materials', Materials Science & Engineering A. Vol.651, pp.198-213.
7. Molestsane, Krakhmalev, Du Plessis, Yadroitsava, Yadroitsev and Kazantseva**,** 2016. 'Tensile properties and microstructure of direct metal laser-sintered Ti-6Al-4V (ELI) alloy', *South African Journal of Industrial Engineering November 2016 Vol 27(3) Special Edition, pp.110-121.*
8. Azam Farooq, Ahmad Majdi, Khurram Altaf1, Haizum Zaharin, 2018**,** 'An in depth review on direct additive manufacturing of metals*', IOP conf. series: Material Science and Engineering 328-012005. doi:10.1088/1757-899X/328/1/012005*.
9. Ravi Yatnalkar, 2010, 'Experimental investigation of plastic deformation of Ti-6Al-4V under various loading conditions', *Master's thesis, The Ohio State University*.
10. Lee, Woei-Shyan, Ming-Tong Lin, 1995, 'The effects of strain rate and temperature on the compressive deformation behaviour of Ti6Al4V.
11. Lee WS, Lin CF, 1998, Plastic deformation and fracture behaviour of Ti–6Al–4V alloy loaded with high strain rate under various temperatures*. Mater Sci Eng A. vol.*241 (1–2), pp: 48-56.
12. Peng-Hui L., Wei-Guo G., Wei-Dong H., Yu S, Xin L.,Kang-BoYuan,2015,'Thermomechanical response of 3D laser-deposited Ti–6Al–4V alloy over a wide range of strain rates and temperatures' Materials Science & Engineering A Vol:647,pp.34–42.
13. Masood S., Mohammadhosseini A., Fraser D M. Jahedi M., 2015, 'Dynamic compressive behaviour of Ti-6Al-4V alloy processed by electron beam melting under high strain rate loading', Adv. Manuf. Vol.3 pp.232–243.*DOI 10.1007/s40436-015-0119-0.*
14. Biswas N, Ding J.L., Balla V.K. , Field D.P., Bandyopadhyay A, 2011. 'Deformation and fracture behaviour of laser processed dense and porous Ti6Al4V alloy under static and dynamic loading' Materials Science and Engineering A, Vol 549 pp. 213-221.
15. William F.Hosford, 2010, 'Mechanical Behaviours of Materials', *2nd Edition Cambridge University Press.pp.92.*

6. BIOGRAPHY

AM Muiruri holds a BEng (First class) Degree in Mechanical Engineering from the Department of Mechanical and Mechatronic Engineering of the Technical University of Kenya. He worked as an Assistance Mechanical Engineer at Collaboration Engineering Solutions and Products, CESP, Africa after completion of his first degree. He is currently registered for a MEng Degree in Mechanical Engineering at Central University of Technology, Free State.

Prof. Willie du Preez is Associate Professor in the Department of Mechanical and Mechatronics Engineering and Director: CRPM Research at Central University of Technology, Free State.

A registered Professional Materials Scientist, directing and supervising research and development on additive manufacturing of titanium alloys and polymers for application in the fields of medical implants and devices, aerospace parts, advanced tooling and direct end-use in industry.

Masu LM is a Professor of Mechanical Engineering of the Vaal University of Technology. His area of specialisation is on numerical modelling, as applied to high pressure vessels. He has read and written extensively in this area over the last 30 years. He is registered engineer of ECSA and an active member of SAIMEch.

Maina Maringa is a Professor of Mechanical Engineering of the Technical University of Kenya with specialisation in the field of composite materials. He has read and written extensively in this area over the last 26 year.

He joined the Central University of Technology, Free State (CUT) in the year 2016 where he has and is engaged in research on additive manufacturing of metals and polymers at CRPM.

Metal Additive Manufacturing of Ti6Al from Blended Elemental Powders

A. Zenani[1], W. B. du Preez[2], T. C. Dzogbewu[3] & I. Yadroitsev[4]
[1,2,3,4] Department of Mechanical and Mechatronics Engineering, Central University of Technology, Free State, Bloemfontein, South Africa

Abstract
TiAl is considered a key material for high performance gas turbines for aircraft engines and automotive applications, in particular for the production of engine valves and turbine wheels for turbochargers. This is due to the attractive high oxidation resistance and strength retention of the TiAl alloy at high temperatures. The limitations of the conventional methods of manufacturing TiAl alloy components with precise geometry and near net shape functional characteristics have limited the usefulness of the TiAl alloys. However, additive manufacturing (AM) methods such as direct metal laser sintering (DMLS) have proven to have the capacity to uncover the full potential of the TiAl alloy by producing components with complex shapes. Ti elemental powder was mixed with Al elemental powder to form a Ti6Al alloy. Geometrical characteristics of single tracks were investigated at a range of laser powers and scanning speeds. To study the distribution of the Al in the Ti matrix, a SEM-EDS investigation was performed. Effects of hatch distance and scanning strategy on the layer surface morphology were investigated. Optimum process parameters for manufacturing Ti6Al components with specific geometrical and functional near net shape characteristics with homogenous microstructure were determined.

Keywords
Metal additive manufacturing, Blended elemental powder, Ti6Al

1 INTRODUCTION

Titanium powder development has become a significant aspect of research, with possibilities of contributing to the growth of the titanium metal industry in South Africa. This interest is driven by the abundant availability of titanium, because South Africa is the second largest producer of titanium raw material in the world. Titanium is classified as a nonferrous and light metal and it stands out primarily because of its high specific strength, corrosion resistance and strength at high temperatures. However, the maximum application temperature is limited by the increased oxidation of the alloy at temperatures above 600ºC. Titanium aluminides (TiAl) partly overcome the oxidation disadvantage; and because of this reason, they have become the subject of intense alloy development efforts [1].

TiAl alloys represent an important class of high temperature structural materials providing a unique set of physical and mechanical properties that can lead to substantial payoffs in aircraft engines, industrial gas turbines and automotive parts due to their unique thermo- mechanical properties. TiAl based alloys have a strong potential to increase the thrust-to-weight ratio in an aircraft engine. This is the case especially with the engine's low-pressure turbine blades and the high-pressure compressor blades. These were traditionally made from Ni-based superalloys, which are nearly twice as dense as TiAl-based alloys [2,3]. Conventional methods of manufacturing have been used to produce TiAl components for high temperature operations with limited success due to their inability to produce near net shapes [2,4].

The mechanical properties of a material are the main determining factors for qualifying it for a specific application. However, for improved energy efficiency and power for highly engineered products, such as aircraft engines, industrial gas turbines and automotive parts, net-shapes that would enhance the geometrical, technical and functional properties of the components have become paramount. These functional requirements of near-net shapes of complex geometries (e.g. back tapers, intricate cooling channels, customized porous structures and special lattices or hollow structures) make additive manufacturing (AM) an attractive manufacturing technology to be exploited for manufacturing TiAl-based alloys with specific geometrical characteristics for high-temperature operations. In addition, the AM processing routes are more promising to produce more isotropic microstructures [4, 5].

Clark [6] and Bewlay et al. [7] concluded that successful manufacturing of near-net TiAl shapes would yield the production of aircraft engines with great propulsion efficiency of a 20% reduction in fuel consumption, a 50% reduction in noise, and 80%

reduction in NOx emissions compared to similar prior engines.

Fundamentally, the direct metal laser sintering (DMLS) technology is a powder bed fusion technique which is applied to build objects layer-by-layer from 3D CAD models, instead of removing material through subtractive methods such as machining.

The essential operation in DMLS is the laser beam scanning over the surface of a thin powder layer previously deposited on a substrate to fuse powder particles together. The forming process proceeds along the scanning direction of the laser beam. Each cross-section (layer) of the part is sequentially filled with elongated lines (tracks) of molten powder [8].

Characteristics of the produced DMLS object such as porosity, microstructure and mechanical properties are influenced by the parameters selected for the DMLS process. Parameters such as layer thickness, as well as powder particle size distribution and shape, are important for deposition of uniform powder layers. Laser power, temperature, spot size, scanning speed of the laser and the deposition time, determine the effectiveness and success of the process and the quality of the part produced [8].

Figure 1 shows the effect of different DMLS process parameters on the characteristics of Ti6Al single tracks.

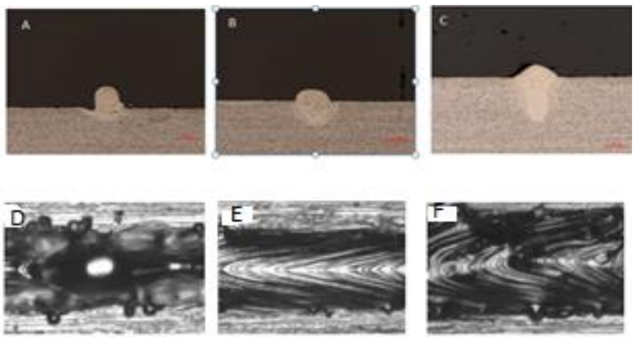


Figure 1 - Cross-sectional and top views of Ti6Al DMLS single tracks on a substrate

The process of single track formation is directly influenced by the process parameters and therefore at improved process parameters, DMLS single tracks are continuous and defect free (Fig 1.B and Fig 1.E)

At high laser power, excessive energy input can lead to a keyhole mode resulting in deep re-melting of the substrate and pores forming inside the molten pool, and this would contribute to porosity in the final product (Fig. 1C and Fig 1.F) [9].

If the scanning speed decreases, the time of irradiation is increased and it results in a higher energy input and irregular tracks are formed, because the heat affected zone becomes larger, more powder is involved and tracks tend to have satellites or dislodged sections that form during solidification.

High scanning speed leads to insufficient re-melting depths and long molten pools. Drop formation (balling effect) occurs when surface tension breaks the continuous molten pool into individual droplets (Fig. 1A and Fig 1D). At low laser power, the molten pool exists at lower temperatures, the surface tension coefficient, as well as melt viscosity, increases and leads to drop formation [8].

This paper reports on an investigation into the feasibility of in-situ alloying of Ti6Al from elemental powders as a cost and time effective means to produce a wide range of alloy compositions.

2 MATERIALS AND METHODS

The experiments were conducted with pure Ti (CP Ti, grade 2) and 99.8% pure Al spherical gas atomized powders procured from TLS Technik GmbH. The Ti powder particle size was <45 µm while the Al powder particle sizes ranged from 40 - 75 µm. For these experiments 94 wt.% of Ti was mechanically mixed with 6 wt.% Al.

The single track experiments were conducted in an EOSINT M280 machine supplied by EOS GmbH. The laser spot diameter was ~80 µm. Based on prior knowledge [5], single tracks and single layers were produced at laser power (P) of 150 W and 350 W over a wide range of scanning speeds (V = 0.4 – 3.4 m/s). A Ti6Al4V substrate was used with a uniform powder deposition of 60 µm. All tracks were of length 0.02 m. For each scanning speed, three single scan lines were produced. The samples were mounted and metallurgically prepared for optical analysis according to well-known procedures described in the literature [5].

The prepared samples were examined to measure the width of the tracks, the height of the tracks above the build plate and the depth of the tracks into the build plate.

The homogeneity of Al in Ti and the amount of Al lost were determined with scanning electron microscopy (SEM) and X-ray energy dispersive spectroscopy (EDS).

3 RESULTS AND DISCUSSIONS

Top surface analysis of the single tracks was conducted.

The results of track width as a function of scanning speed are graphically presented in figure 2.

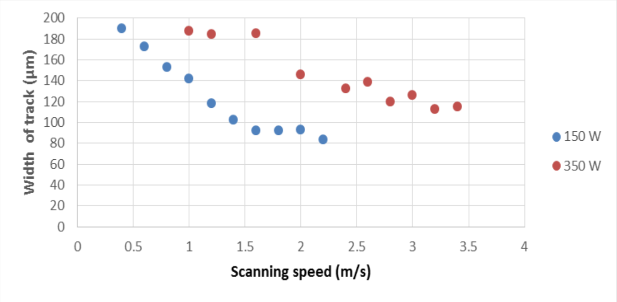


Figure 2 - The width of single tracks at different scanning speeds for two laser power settings

From figure 2, it is noted that the widths of the single tracks decrease with increasing scanning speed and decrease with increasing laser power.

With the increase of the laser scanning speed, heat loss by conduction decreases, and thus, the absorbed energy of laser radiation goes directly into fusing the material, and the width of the tracks is comparable to the laser beam diameter [8].
Progressive increase in the scanning speed has a significant effect on the capillarity instability of the molten laser cylindrical liquid tracks [10].

An understanding of the width to scanning speed relationship of the single track helps in determining hatch distance in order to ensure accurate overlapping between the tracks [11].

The results of linear energy density as a function of laser scanning speed are shown in figure 3.

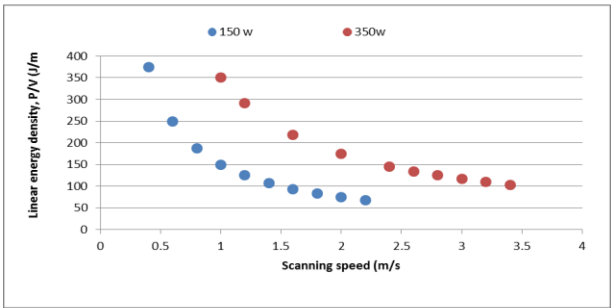


Figure 3 - Linear energy density of single tracks at different scanning speeds for two laser power settings

From figure 3, it is noted that an increase in the scanning speed reduces the linear energy density.

A reduction in energy density leads to a reduction of the temperature in the molten pool. The lower melting temperature would produce a limited molten fluid of high viscosity, which would obstruct the free movement of the unmelted powder particles in the molten pool. The high viscous liquid is expected to form semi-discontinuous tracks known as pre-balls and balling effect [12].

Balling is also caused by inadequate penetration of the input energy to melt and wet the underlying substrate due to surface tension, which tends to spheroidise the liquid [8].
The higher energy input increases the temperature of the molten pool and subsequent production of a large amount of liquid phase with low viscosity which enhances easy flowing of the molten pool. At low scanning speeds the heat conduction promotes heating of the substrate and powder material through the microscopic collision of the powder particles and their movement; thus creating conditions for involvement of new material into the fusion process [8].

During the in-situ DMLS process, if the selected process parameters lead to sufficient laser energy density to melt the powder and penetrate into the previous layer, bonding between the substrate (previous layer) and the next layer takes place. [8,13]

The relationship between the remelted depth and scanning speed is shown in Figure 4.

The track width and the remelted depth have a linear dependence: they decrease with the scanning speed and they decrease with laser power. When the scanning speed decreases, the length of the molten pool decreases as well and its width increases. Molten pools with greater circumference-to-length ratio show a more stable behaviour [8].

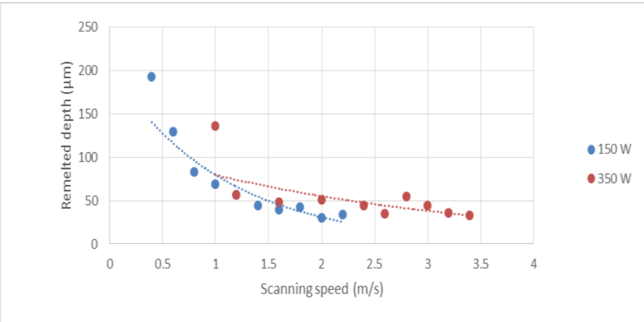


Figure 4 - Remelted depth for continuous tracks at different scanning speed for two laser power inputs.

In Figure 5 the remelted depth of the single tracks at 150 W laser power and corresponding scanning speed of 0.4 m/s – 0.8 m/s and 350 W at 1 m/s shows a V shape and signifies deep penetration of the laser into the substrate. This is an example of the effect of a combination of a high energy input and a low scanning speed. If the scanning speed decreases, the time of irradiation is increased and it results in a higher energy input and irregular tracks are formed, because the heat affected zone becomes larger, more powder is involved and tracks tend to have satellites or dislodged sections that form during solidification [9].

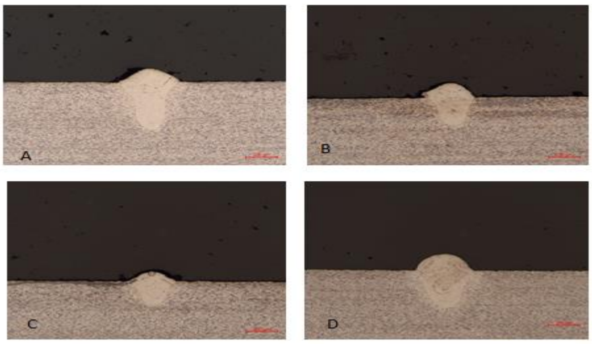


Figure 5 – Cross sectional views of the single track penetration into the substrate following a V shape *(A) 150 W at 0.4 m/s (B) 150 W at 0.6 m/s (C) 150 W at 0.8 m/s and (D) 350 W at 1 m/s.*

In Figure 6 the shape of the single tracks formed at laser power of 150 W and the corresponding scanning speeds of 1 - 1.4 m/s and 350 W at 1.2 m/s, respectively, resembles a semi-circular (*U*-shape) and signifies sufficient laser power at a good scanning speed. Such conducive mode would lead to good bonding between the tracks and the substrate, hence could be referred to as the window of optimum processing parameters [14].

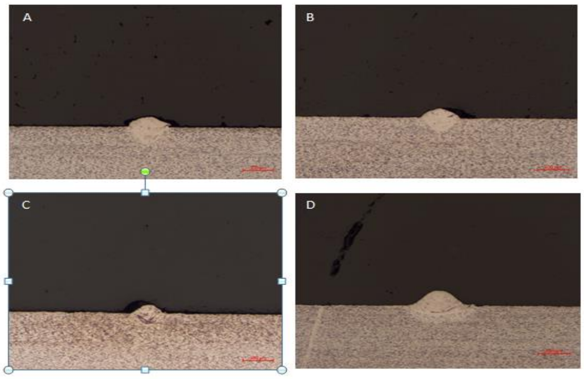


Figure 6 – Cross sectional views of the single track penetration into the substrate following a U shape *(A) 150 W at 1 m/s (B) 150 W 1.2 m/s (C) 150 W at 1.4 m/s and (D) 350 W at 1.2 m/s.*

In Figure 7, at scanning speeds in the range of 1.6-2.2 m/s for corresponding laser powers of 150 W and 1.6 – 2 m/s at 350 W, respectively, the laser energy melts the powder and drills very deep into the substrate. This is as result of the excessive laser energy which causes the laser beam to create a topological depression in the substrate. [14].

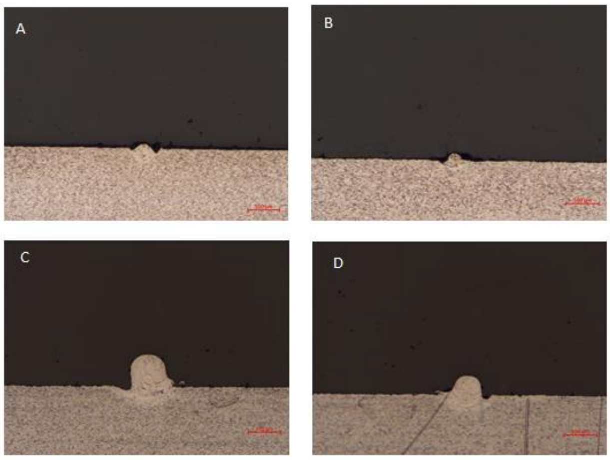


Figure 7 – Cross sectional views of the single track penetration into the substrate resembling a keyhole *(A) 150W at 1.6 m/s (B) 150W 2.2m/s (C) 350W at 1.6 m/s and (D) 350W at 2 m/s*

To investigate the influence of scanning strategy on single layer morphology, two types of surfaces were produced, namely, a single scan and rescanned surfaces at different hatch distances of 80 µm, 90 µm and 100 µm. These results are shown in Figure 8.

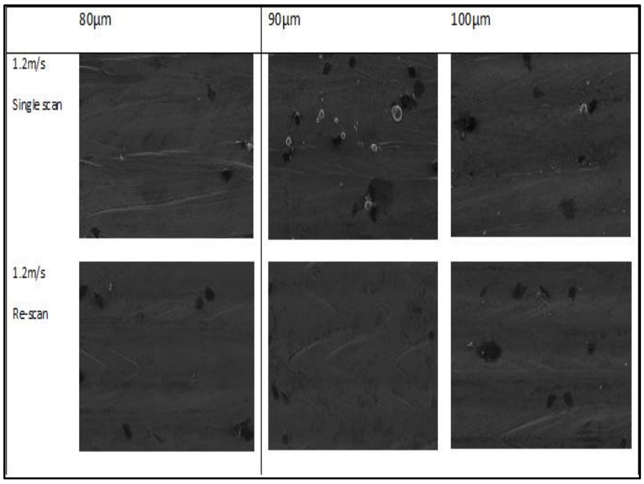


Figure 8 - Top view of the surfaces at hatch distances of 80 µm, 90 µm and 100 µm at *P*= 150 W, *V*= 1.2 m/s

From the results in figure 8, it can be concluded that good single track overlapping can be achieved at more than one hatch distance and that rescanning strategy enhances homogeneity of the in situ alloyed material, by ensuring that the laser could melt all powder particles and satellites, even though rescanning increases the production time [11].

A cross-sectional examination of the Ti6Al samples (Figure 9) reveals that Ti and Al particles melted completely. However, there are Al rich areas where there is a random distribution of Al in the Ti alloy matrix. The process of rescanning helps to homogenise the Al distribution in the Al rich areas.

During the rescanning process, the previously rough surface enhances laser radiation absorption. The increase in laser absorptivity due to the rougher surface leads to an increase in temperature and more prominent flow of the molten pool. The increased temperature and the more prominent liquid flow causes the Al concentrated zones to homogenise [15].

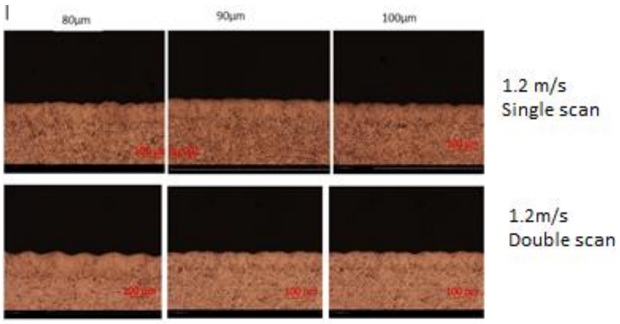


Figure 9 - Cross sectional view of DMLS Ti6Al *in-situ* alloy at 150 W laser power, and 1.2 m/s laser scanning speed of single and double scans at 80 µm, 90 µm and 100 µm hatch distances.

An EDS elemental mapping analysis was conducted to verify the distribution of the Al in the Ti near the surface of the alloy. The results given in Figure 10 suggest that Al was not evenly distributed in the Ti during the single scan, however, rescanning has improved the distribution of the Al in the Ti. The difference in hatch distance seems not to have any significant effect on the Al distribution in the Ti. All the samples demonstrated a similar level of Al distribution at each hatch distance (80 µm, 90 µm and 100 µm) in the Ti alloy matrix at a given constant laser speed.

SEM observations of the manufactured samples demonstrated that the alloy did not homogeneously mix and areas rich in Al were observed. On average, the concentration of Al in the alloy varied in the range of 0.6-4.98 wt.%.

The observed effect of energy input on *in situ* alloying a Ti and Al powder mixture can be explained with the differences in the material properties. Ti and Al have a specific heat capacity of 528 J/(kg.K) and 921 J/(kg.K), thermal conductivity of 17 W/(m.K) and 205 W/(m.K), density of 4500 kg/m^3 and 2710kg/m^3, latent heat of fusion of 435 kJ/kg and 398 kJ/kg, respectively. Al has a higher specific heat capacity and a higher thermal conductivity that result in Al absorbing more laser radiation than Ti. Add to this the much lower melting point of Al (660 ºC) than Ti (1668 ºC) and this explains why there is a significant loss of Al compared to Ti. The density of Al is only 60% of that of Ti, which will cause Al to move to the top and Ti to sink to the bottom in the molten pool. This brings Al even closer to the laser beam and promotes an accelerated melting and loss of Al [16].

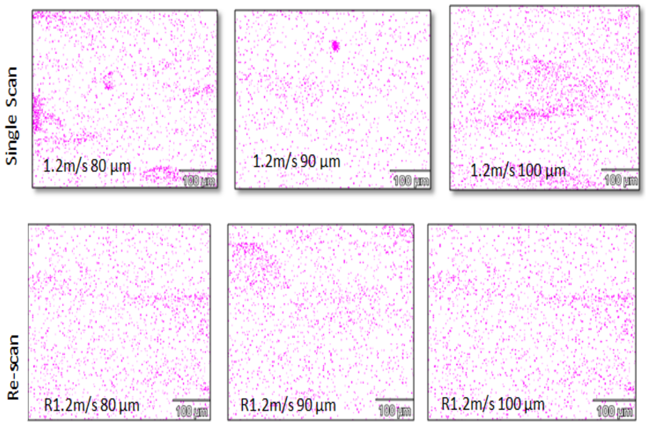


Figure 10 - Al EDS distribution map in backscattered electron (BSE) mode of DMLS Ti6Al *in-situ* alloy for a single scan and rescan at laser power 150 W, scanning speed 1.2 m/s and 80-100 µm hatch distances.

4 CONCLUSION

An understanding of the optimum process parameters of the single tracks and layers in a DMLS in-situ alloying process is essential before any further processing into 3D parts is attempted. As demonstrated in the current study and previous experiment results [10], the combination of the process variables shows a non-linear relationship between the different process parameters.

In the in situ alloying of blended Ti6Al powders, a melting of the Ti and the Al powders is achieved. Areas where there is uneven distribution of the Al powder are observed. The samples were rescanned and notable improvements in the homogeneity and distribution of the Al in the Ti matrix were observed. Therefore, the rescanning process has proven to yield noticeable improvements in the homogeneity and surface quality of the layers.

Improving the mixing process of the Ti6Al alloy before the experiments has the potential of further reducing the uneven distribution of the Al powder in the Ti matrix, since the current mixing of the powder was done manually.
The DMLS process was able to produce homogenous tracks of TiAl alloy with a complete melting of the powder after the rescanning process, showing an improved homogeneity on the single tracks, compared to what other previous authors have reported [4-5, 17-19].

From the analysis of the Ti6Al alloy samples, the most preferable samples seem to be at 150 W laser power, 80 µm hatch distance and 1.2 m/s scanning speed, using a rescanning process. These samples show a good melting and homogeneity of the Al particles on the surface and in the cross sections. They display a U shaped penetration depth into the substrate, which leads to good bonding between the tracks and the substrate [14].

5 ACKNOWLEDGEMENT

The active support and funding of the South African Department of Science and Technology through the Collaborative Program in Additive Manufacturing, Contract No.: CSIR-NLC-CPAM-15-MOA-CUT-01, is gratefully acknowledged.

The authors acknowledge the support provided by the team of the CRPM of the Central University of Technology, Free State.

6 REFERENCES

[1] Matthew J.Donachie, Jr.2000. Titanium: A Technical Guide. Second Edition, Materials Park, Ohio 44073-0002

[2] Lapin, J. 2009. TiAl-based alloys: Present status and future perspectives. In Conference Proceedings Metal, 19 (21.5).

[3] Bewlay, B.P., Nag, S., Suzuki, A. & Weimer, M.J. 2016. TiAl alloys in commercial aircraft engines. Materials at High Temperatures, 33(4-5), pp.549-559.

[4] Thomas, M., Malot, T., Aubry, P., Colin, C., Vilaro, T. & Bertrand, P. 2016. The prospects for additive manufacturing of bulk TiAl alloy. Materials at High Temperatures, 33(4-5), pp.571-577.

[5] Vilaro, T., Kottman-Rexerodt, V., Thomas, M., Colin, C., Bertrand, P., Thivillon, L., Abed, S., Ji, V., Aubry, P., Peyre, P. & Malot, T. 2010. Direct fabrication of a Ti-47Al-2Cr-2Nb alloy by selective laser melting and direct metal deposition processes. In Advanced Materials Research, 89, pp. 586-591. Trans Tech Publications.

[6] Clark, S. F. 2012. 787 Propulsion Systems, Aero Quarterly. Available at:www.boeing.com/boeingedge/aeromagazin e (Accessed: 05 September 2018).

[7] Bewlay, B. P., M. Weimer, T. Kelly, A. Suzuki, and P. R. Subramanian. 2013. "The Science, Technology, and Implementation of TiAl Alloys in Commercial Aircraft Engines." In Materials Research Society Symposium Proceedings, 1516:49–58.

[8] Yadroitsev, I. 2009. Selective laser melting - Direct manufacturing of 3D-objects by selective laser melting of metal powders, Lambert Academic Publishing, p 270.

[9] Rai.R, E. W. 2007. Heat transfer and fluid flow during keyhole mode laser welding of tantalum, Ti–6Al–4V, 304L stainless steel and vanadium. Journal of Physics D: Applied Physics, 40(18), 5753.

[10] Yadroitsev. I,. 2007. Parametric analysis of the selective laser melting process. Applied Surface Science, 8064–8069.

[11] Yadroitsev.I,. (2015.). Hierarchical design principles of selective laser melting for high quality metallic objects. Additive Manufacturing, 45-56.

[12] Yadroitsev, I.. (2010). Single track formation in selective laser melting of metal powders. . Journal of Materials Processing Technology, 210(12), pp. 1624-1631., 210(12), pp. 1624-1631.

[13] Kruth.J.P,. (2005). Binding mechanisms in selective laser sintering and selective laser melting,. Rapid Prototyping Journal 11, 26-36.

[14] Eagar, T. W. (1983). Temperature fields produced by traveling distributed heat sources. . Welding Journal, 62(12), pp. 346-355.

[15] Bergström, D., Powell, J. & Kaplan, A. F. H., 2007. A ray-tracing analysis of the absorption of light by smooth and rough metal surfaces. Journal of Applied Physics, 101(11), pp. 113504.

[16] AZO,. (2015, June 23). AZO materials. Retrieved from Grade 23 Ti-6Al-4V ELI (UNS R56401) data sheet: http://www.azom.com

[17] Gussone, J., Garces, G., Haubrich, J., Stark, A., Hagedorn, Y.C., Schell, N. and Requena, G. 2017. Microstructure stability of γ-TiAl produced by selective laser melting. Scripta Materialia, 130, pp.110-113.

[18] Gussone, J., Hagedorn, Y.C., Gherekhloo, H., Kasperovich, G., Merzouk, T. and Hausmann, J. 2015. Microstructure of γ-titanium aluminide processed by selective laser melting at elevated temperatures. Intermetallics, 66, pp.133-140.

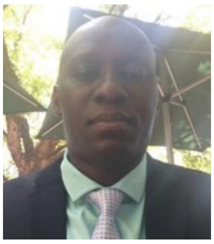
[19] Löber, L., Schimansky, F.P., Kühn, U., Pyczak, F. and Eckert, J. 2014. Selective laser melting of a beta-solidifying TNM-B1 titanium aluminide alloy. Journal of Materials Processing Technology, 214(9), pp.1852-1860.

7 BIOGRAPHY

Asavela Zenani obtained a Bachelor of technology degree in Mechanical Engineering from the University of South Africa in 2015 and a Bachelor of technology degree in Quality from the Nelson Mandela Metropolitan University of South Africa in 2013. She is currently pursuing Masters in Engineering with the Central University of Technology in the Free state, Bloemfontein.

Prof. Willie du Preez is Associate Professor in the Department of Mechanical and Mechatronics Engineering and Director: CRPM Research at Central University of Technology, Free State. He is a registered Professional Materials Scientist, directing and supervising research and development on additive manufacturing of titanium alloys and polymers for application in the fields of medical implants and devices, aerospace parts, advanced tooling and direct end-use in industry.

Prof. Igor Yadroitsev (Ihar Yadroitsau) is Research Professor in Central University of Technology, Free State. He is an international expert in applied optics, laser technologies and additive manufacturing. Since 2015 he is also the South African National Research Foundation Chair in Medical product development through Additive Manufacturing. Prof. Yadroitsev has over 30 years of academic experience. He published more than 140 papers and he is a holder of 10 patents in optics and additive manufacturing.



Dr. Thywill C. Dzogbewu is a postdoctoral research fellow at Central university of Technology, where he obtained his PhD. He received his MSc. and BSc. in Physics from Kwame Nkrumah University of Science and Technology. He is registered candidate materials scientist. His major research interest includes material physics, additive manufacturing of titanium alloys, engineering material properties, heat treatment of metals and microstructural analysis.

Laser Based Powder Bed Fusion of Pure Platinum

I. Yadroitsava[1], I. Yadroitsev[1], I. van Zyl[1], D. Kouprianoff[1], W.B. du Preez[1]
[1]Department of Mechanical and Mechatronics Engineering, Central University of Technology, Free State, South Africa

Abstract
Laser based powder bed fusion (LBPF) is an innovative technology which allows the production of fully dense, fine-structured parts with complex shapes. Platinum is a precious metal which forms part of the family of platinum group metals. Typically, platinum is used in the jewellery, glass and medical industries, in the automotive industry as catalyst in catalytic converters, and in the aerospace industry. Processing of platinum has conventionally been done through casting and normal metalworking methods for the conversion into useful components, however, in the jewellery industry casting of platinum is a well-known challenge to get exact non-porous components. Additive manufacturing of platinum opens up novel possibilities in terms of the complexity of the geometry that could be obtained. Successful laser based powder bed fusion of metal components require fine calibration and control of the process. Simulation allows for the estimation of thermal distribution during processing. Microscopy of single tracks allows insight into the effectiveness of process controls. In this paper the approach followed to produce platinum parts in an EOSINT M280 direct metal laser sintering machine is discussed.

Keywords
Laser based powder bed fusion, Simulation, Platinum powders

1 1 INTRODUCTION

Laser Powder Bed Fusion (LPBF) is a layer-by-layer manufacturing method in which powder layers are scanned by a laser beam to fuse powder particles together, as well as onto the previous layer. Sequentially, track–by-track, layer-by-layer, a 3D part is sintered. LPBF can produce fully dense parts of complex geometry with internal structures, including light-weight lattice structures. This technique opens up new avenues towards complex shape parts with inner structure, especially in the automotive, aerospace, medical and jewellery fields where unique designs, light weight and topology optimisation are required [1]. To produce solid non-porous 3D objects, a hierarchical approach was suggested, including mutual analysis of LPBF parameters necessary to control the final product quality on every level – the track, the layer and the final 3D object [2]. This approach was used for manufacturing of fully dense 3D samples from pure platinum powder supplied by Lonmin Plc.

2 2 MATERIALS AND METHODS

2.1 2.1Materials

The experimental work was done on a 50 mm x 50 mm x 6 mm pure platinum substrate with mass of 259.6 g and 99.968 % purity, (i.e. the overall total impurity content is not more than 0.032 wt.%). Surface roughness of the substrate was Ra=0.5±0.04 µm, Rz=7.4±0.9 µm.

Water-atomised powder with similar chemical composition was used. Powder particles were spherical in shape and had size <45 µm (Fig. 1). Before experiments, “powder for single tracks” was dried at 80°C for 2 hours. Only this powder was used for the following experiments with single tracks and layers.

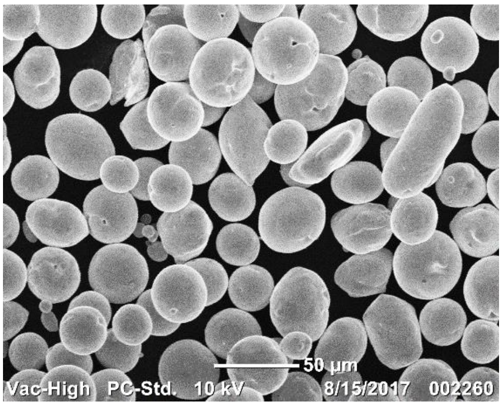


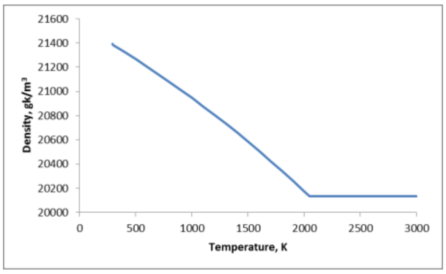
Figure 1 - SEM micrograph of pure Pt powder.

Surface roughness was measured with a Mitutoyo SJ-210 roughness tester in accordance with ISO1997 (λc=2.5 mm).

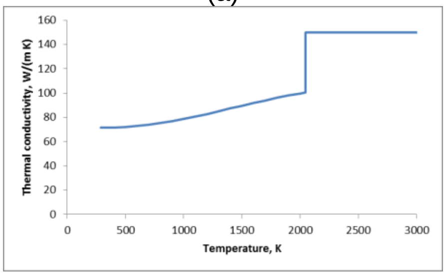
Single track scanning was done in the EOSINT M280 400 W system in Ar atmosphere at different scanning speeds and laser power. The laser beam spot size was about 80 µm.

The substrate with LPBF samples was divided by electro-discharge machining (EDM) wire cutting for cross-sectional analysis purposes. Templates were prepared in accordance with metallographic and materialographic specimen preparation practice for noble metals [3].

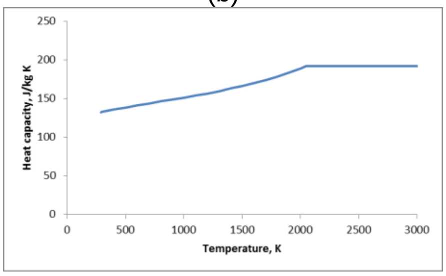
Temperature-dependent properties such as density, thermal conductivity and heat capacity of pure Pt were used for numerical simulations of single lines (Fig. 2).



(a)



(b)



(c)

Figure 2 - Temperature-dependent properties of pure Pt: density (a), thermal conductivity (b) and heat capacity (c) [8]

2.2 2.2Numerical Simulation

The energy needed for fusion of a unit volume of the material can be estimated with equation (1):

$$Q_{U.V.} = \rho[c \times (T_m - T_0) + L_f] \;, \qquad (1)$$

where ρ is the density, c is the specific heat capacity, T_m is the melting point, T_0 is… and L_f is the latent heat of fusion (133.6 kJ/kg). To melt 1 mm^3 pure Pt alloy 7.44 J/mm^3 has to be absorbed by the Pt substrate during laser beam scanning. This calculated value is 1.3 times higher than for Ti6Al4V alloy. Pt also has higher thermal diffusivity (25×10^{-6} m^2/s); this value for Ti6Al4V alloy is 2.9×10^{-6} m^2/s at room temperature, thus the ratio of ability to conduct thermal energy versus to store the energy is higher in pure Pt than in Ti6Al4V. A high thermal conductive media such as pure Pt requires high energy density and, at the same time, optimal scanning speed has to be used.

Temperature-dependent properties of pure Pt were used for numerical simulations of single lines produced by laser beam with Gaussian energy distribution on the solid substrate. The energy supplied by a laser beam with Gaussian distribution is defined by (2)

$$Q = Q_0(1-R)\frac{f \cdot A_z}{\pi r_0^2} e^{-\frac{f(x^2+y^2)}{r_0^2}} e^{-A_z z} \qquad (2)$$

where Q_0 is the total laser power; R is the surface reflectivity of the material for a given wavelength; A_z is the coefficient of absorption; r_0 is the characteristic radius of the laser beam; f is the power distribution factor; $x=x_0+V_t$, where V is the laser scanning speed. The heat balance is [4]:

$$\rho c(\frac{\partial T}{\partial t} + V\nabla T) = \nabla(k\nabla T) + Q \qquad (3)$$

where T is temperature, t is time and V is laser scanning speed, c is specific heat capacity, ρ is density of material, k is thermal conductivity. For numerical simulation the time-dependent “Heat Transfer in Solids” module of Comsol has been chosen in order to solve the heat balance equation. In the liquid molten pool Marangoni flow contributes significantly to the temperature distribution and shape of the molten pool. To simulate metal welding, enhanced thermal conductivity of liquid metal has been used by Zhang et al [5]. Following a similar approach, for the numerical simulation of Pt by laser melting, the effective thermal conductivity k_{eff} was chosen as

$$k_{eff} = \begin{cases} k \ \text{ for } T < T_m \\ m \cdot k \text{ for } T \geq T_m \end{cases}$$, where k is the thermal conductivity in solid state, T_m is the melting temperature, m is the coefficient >1 for the liquid metal.

For numerical simulations f=2 (Gaussian beam) and m=2 k_{eff} were chosen.

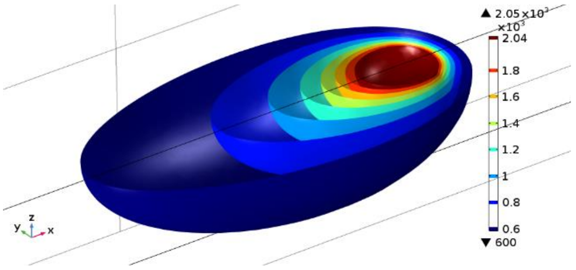
Except the top surface, all other boundaries are assumed to be thermally insulated. The heat flux on the top surface simulates convective cooling. Heat losses due to convection is expressed by

$$k\frac{\partial T(x,y,L_z,t)}{\partial z}\Big|_{top} = h_c(T - T_0) \;,$$

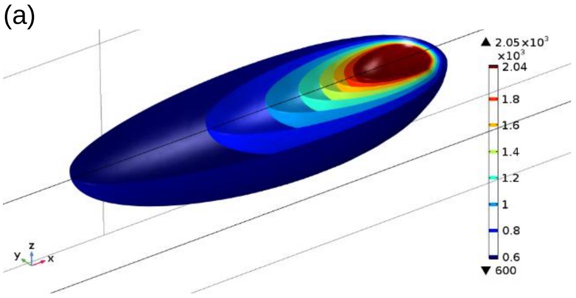
where T_0 = 293 K is the initial temperature, h_c = 10 W/(m^{2}K) is the convection coefficient. In order to obtain accurate results, the density of the mesh in the region around the irradiation, and in the top region of the sample (50–200 µm), was higher than in the sample as a whole. The minimum mesh size was 0.5 µm.

Shiraishi and Tilley [6] studied the spectral reflectivity of homogeneous binary gold alloys and also pure Pt. They found a great flattening of the reflectivity curve in the visible spectrum for pure Pt and for Pt and Pd alloys with gold. Thus, for numerical simulations a reflectance value of R=0.8 was used.

Numerical simulations showed that the penetration depth was more than 50 µm. This value is larger than the size of the powder particles, so all powder will be molten. With the scanning speeds used, the isothermal contours were shallower and longer (Fig. 4). The ratio between the length of molten pool to its width changed from 1.1 to 1.3, thus humping and balling effects can occur at higher scanning speeds.



(a)



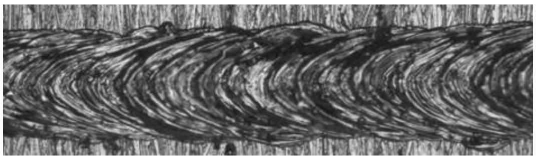
(b)

Figure 3 - Isothermal contours for numerical simulations of pure Pt at constant laser power with 80 µm spot size and parameter set 1 (a) parameter set 4 (b).

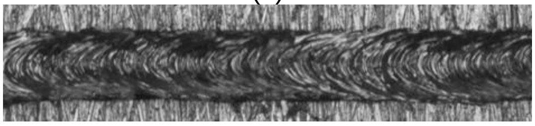
3 3 EXPERIMENTAL RESULTS AND DISCUSSION

3.1 3.1Single tracks

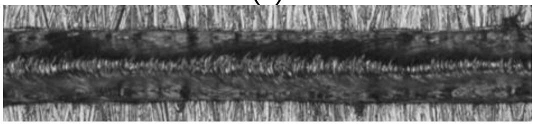
Based on the pilot study that was done earlier, a constant laser power and varying scanning speeds were used to produce single tracks on the substrate without powder (Fig. 5). Tracks were 20 mm in length. Solidification lines after laser scanning were clearly visible at certain scanning speeds on the top view of the substrate. Single tracks with pure Pt were conducted at the same process parameters on powder layer thickness of 50 µm (Fig. 6). Also 3 tracks were produced and analysed for each set of the process parameters.



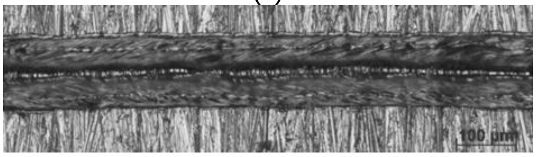
(a)



(b)

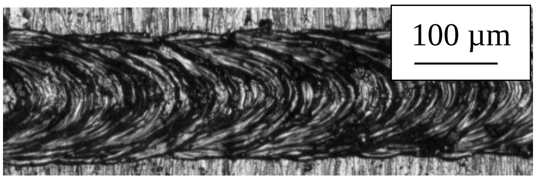


(c)

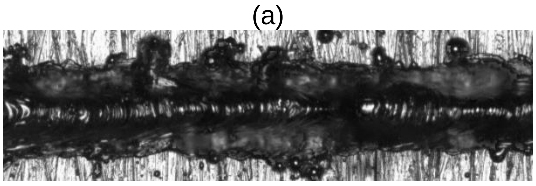


(d)

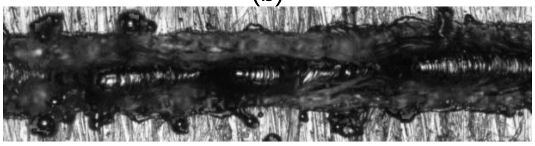
Figure 4 - Top view of single tracks at constant laser power, scanning speeds listed from parameter set 1 (a) parameter set 2 (b) parameter set 3 (c) parameter set 4 (d).



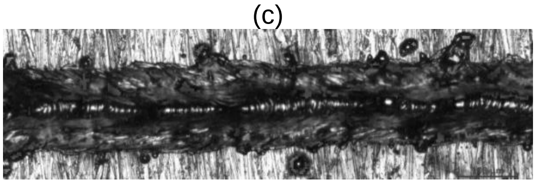
(a)



(b)



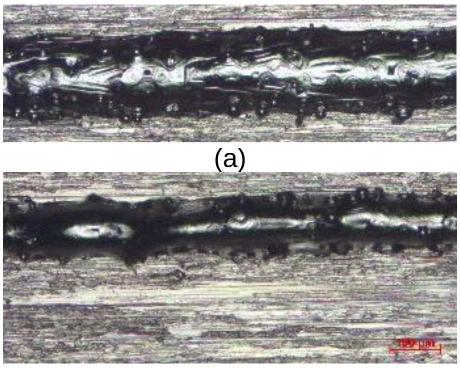
(c)



(d)

Figure 5 - Top view of single tracks from Pt powder on Pt substrate: constant laser power, scanning speeds listed from parameter set 1 (a) parameter set 2 (b) parameter set 3 (c) parameter set 4 (d).

The balling effect started at 0.8 m/s (Fig. 6c). For Ti6Al4V single tracks, the balling effect started at higher scanning speed: at 300 W laser power and scanning speed up to 1.8 m/s, tracks were stable (see Fig. 7). It can be assumed that the lower absorption coefficient than Pt (R=0.63 for Ti, Steen, [9]) and higher surface tension at melting point resulted in this phenomenon.



(a)

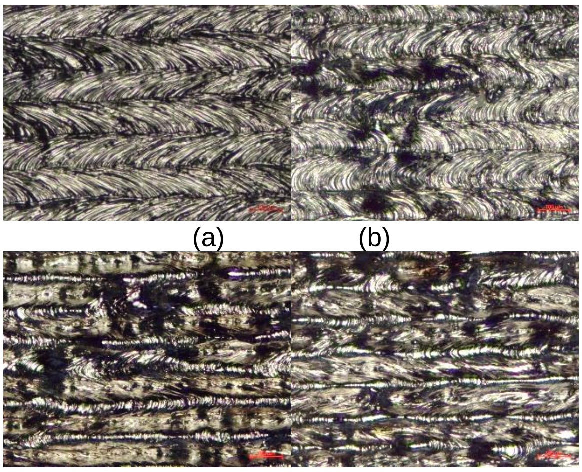
(b)

Figure 6 - Top view and cross-section of LPBF Ti6Al4V single tracks: laser power 300 W, scanning speeds 1.2 (a) and 1.8 m/s (b).

Similar to single tracks without powder, at scanning speeds of parameter set 1 and parameter set 2, the width of the tracks decreased significantly and then was approximately constant. Analysis of the cross-sections confirmed significant humping in the molten pool at faster scanning speeds. Since single tracks were more than 120 µm wide (Fig. 6), a hatch distance of 100 µm was chosen for manufacturing single layers.

3.2 3.1Single layers

LPBF single layers were produced at a constant laser power, 100 µm hatch distance and scanning speeds of parameter set 1 to 4. Geometrical characteristics of the single tracks influenced the morphology of the layers. Homogenous and very smooth layers were achieved at parameter set 1 scanning speed; at parameter set 4 the layers were more irregular (Fig. 9).



(a) (b)

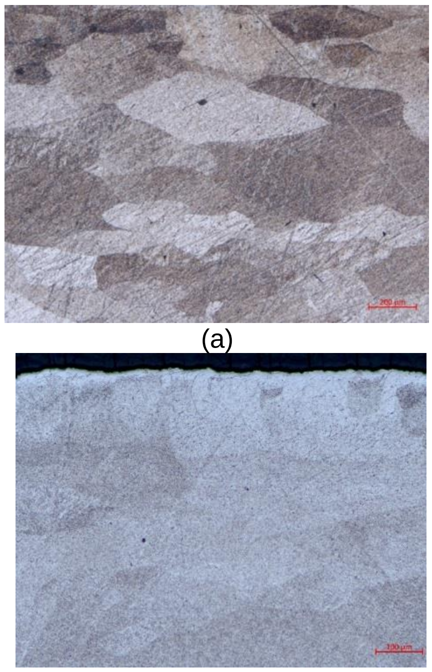
(c) (d)

Figure 7 - Top view of single layers manufactured at 100 µm hatch distance, laser power and scanning speeds of parameter set 1 (a), parameter set 2 (b), parameter set 3 (c) and parameter set 4 (d).

In LPBF, surface roughness depends on powder layer thickness, laser power density, scanning speed and thermo-physical properties of material, etc. When the hatch distance is shorter than the denudation zone, consistent diminishing of track width and height occur [7]. When the surface is very smooth, powder delivery will be difficult because spherical powder particles easily slip off the substrate when the powder is levelled with the blade or brush. For LPBF the profile of the layer at parameter set 2 looks preferable; layers produced at parameter set 3 and 4 had very rough surfaces; it can provoke porosity in final 3D LPBF samples.

3.3 3.33D LPBF objects

For high mechanical properties, 3D LPBF objects have to be without pores and must have good metallurgical bonding between the layers. So penetration depth is very important. After etching of cross-sections, it was found that the basic Pt substrate had a coarse structure with grain size of more than 200 µm. Grains were elongated in shape along the rolling direction (Fig. 10a). Recrystallization after laser melting led to very fine microstructure and electro-etching even at 10 V did not reveal the fusion boundaries (Fig. 10b).



(a)

(b)

Figure 8 - Cross-sections of the substrate (a) and LPBF Pt single layer at parameter set 1, at 100 µm hatch distance (b).

Analysis of cross-sections showed that a solid powder layer was produced. Based on these experiments, non-porous 3D objects were manufactured (Figure 11). High density of the parts was confirmed by micro CT scans at Stellenbosch University.

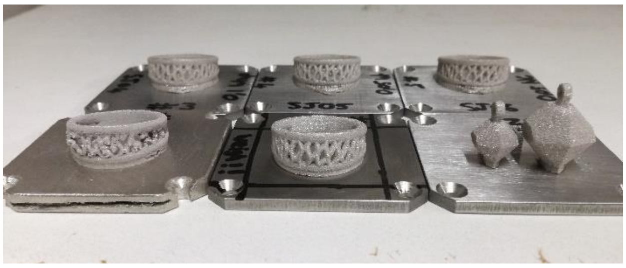


Figure 9 - 3D pure platinum objects manufactured by LPBF at optimal process parameters.

4 CONCLUSIONS

Based on numerical simulations, constant laser power with 80 µm spot size was chosen for fusion of Pt substrate and powder. More stable tracks were obtained at parameter set 1 and 2. Single layers produced at these scanning speeds and hatch distance of 100 µm also showed lower surface roughness. Appropriate morphology of LPBF single layers fabricated at parameter set 2 permitted the identification of this laser power and scanning speed as optimal for manufacturing 3D LPBF porosity-free samples.

5 ACKNOWLEDGEMENTS

The authors are grateful for the support of the project by Platforum and the platinum producer Lonmin, Plc also the Centre of Rapid Prototyping and Manufacturing for their continued technical support.This work is based on research supported by the Collaborative Program in Additive Manufacturing and the South African Research Chairs Initiative of the Department of Science and Technology and the National Research Foundation of South Africa.

6 REFERENCES

[1] Wohlers Associates, Inc., 2013. Wohlers Report 2013: Additive Manufacturing and 3D printing state of the industry, Fort Collins, Colorado, USA: Wohlers Associates.

[2] Yadroitsev, I., Krakhmalev, P. Yadroitsava, I., 2015. Hierarchical design principles of selective laser melting for high quality metallic objects. Additive Manufacturing J., 7, pp. 45–56.

[3] Bjerregaard L., K. Geels, B. Ottesen, M. Rückert. Metalog Guide, Struers A/S, Copenhagen, Denmark, 2000.

[4] Mazumder, J. and Steen, W. Heat transfer model for CW laser material processing. J. Appl. Phys. 1980, 51, 941-947. 10.1061/1.327672.

[5] Zhang, Q., Zhang, D., Jia, S., Shong, W., 1995. Microstructure and properties of some dispersion strengthened platinum alloys, the influence of yttrium and zirconium additions. Platinum Metals Review, 39 (4), pp. 167-171.

[6] Shiraishi T., Tilley, R. J. D., 2014. An estimation of the reflectivity of some monophasic binary gold alloys. Gold Bull, 47, pp. 75–82.

[7] Yadroitsev, I., 2009. Selective laser melting: Direct manufacturing of 3D-objects by selective laser melting of metal powders. LAP Lambert Academic Publishing, 280 p.

[8] Paradis P.F., 2014. Thermophysical properties of Platinum group metals in their liquid undercooled and superheated phases. Johnson Matthey Technol. Rev., 58, (3), pp. 124–136.

[9] [Steen, W. M., 2003. Laser Material Processing. Springer Science & Business Media, 408 p.

7 BIOGRAPHY

Dr. Ina Yadroitsava is Senior Researcher in the Department of Mechanical and Mechatronics Engineering, Central University of Technology, Free State. Since 2010 her specific focus and area of research interest is in Additive Manufacturing. She is co-author of 40 papers in this field.

Prof. Igor Yadroitsev (Ihar Yadroitsau) is Research Professor in Central University of Technology, Free State. He is an international expert in applied optics, laser technologies and additive manufacturing. Since 2015 he is also the South African National Research Foundation Chair in Medical Product Development through Additive Manufacturing. Prof. Yadroitsev has over 30 years of academic experience. He has published more than 140 papers and he is a holder of 10 patents in optics and additive manufacturing.

Ian van Zyl obtained an M.Eng in Mechanical Engineering in 2016 from Central University of Technology (CUT), Free State. He is presently employed by the Central University of Technology Innovation Services. Since 2018 he has been working on his doctoral dissertation.

Dean-Paul Kouprianoff holds an M. Eng in Mechanical Engineering since 2018 (Cum laude) from Central University of Technology (CUT), Free State. He is presently employed by CRPM Research at CUT. Since 2018 he has been working on his doctoral dissertation and has published 5 papers.

Prof. Willie du Preez is Associate Professor in the Department of Mechanical and Mechatronics Engineering and Director: CRPM Research at Central University of Technology, Free State. He is a registered Professional Materials Scientist, directing and supervising research and development on additive manufacturing of titanium alloys and polymers for application in the fields of medical implants and devices, aerospace parts, advanced tooling and direct end-use in industry.

Investigation of an Economical Stereolithography 3d Printer for Rapid Prototyping and Mass Production

C.I. Basson[1], G. Bright[1]
[1]Department of Mechanical Engineering, University of KwaZulu-Natal, South Africa

Abstract
Rapid prototyping justifies the requirement for advancements in customizable and mass manufacturing environments. Masked Stereolithography, (MSLA), 3D printing provides the potential platform to manufacture a number of components across a large surface area and produces high-resolution components. Industrial stereolithographic 3D printers with high production and customizable capabilities are an expensive capital investment. The study provides a proposed solution to an affordable 3D printer employing a large printable area. The proposed system was designed to utilize mechatronic motion control and photopolymerization curing rates of resins through the variable intensity of ultraviolet light. The design of the prospective MSLA 3D printer ensures production complexity and minimal filament consumption. A modular reservoir was designed with an automated refill system applying a fluid level control. The system was tested according to UV LED wavelength and hardening rates. The refill system was tested according to resin replenishment rates. Conclusions and recommendations were made according to the feasibility of a low-cost stereo-lithography 3D printer.

Keywords
Flexibility, MSLA 3D printing, rapid prototyping, photo-polymerisation, customizable manufacturing

1 INTRODUCTION

Manufacturing environments for mass production require products to be produced numerously, accurately, with high surface resolution and respond to the market rapidly [1]. Stereolithography, (SLA), 3D printing possesses the ability to satisfy consumer and production needs of customizable products and mass production through short manufacturing phases and requires minimal part surface finishing [2]. SLA is associated with additive manufacturing and products are created through a chemical reaction utilizing ultraviolet, (UV), to cure a photopolymer. SLA bonding enables the rapid design and decreases iterations in the developmental stage of product conception. SLA bonding also reduces costs involved in maintenance, physical space requirements, infrastructure and equipment installation [3]. SLA 3D printing uses a process called photo-polymerisation to produce accurate and high-resolution components [4]. Prior to the curing of the layer new polymer coating is glazed across the hardened level and the hardening process is repeated until a final component is fashioned.

Light Emitting Diodes, (LEDs), retain high light intensity and are suitable for photopolymerization [5]. LED technology possesses several thousand hours of operation lifetime and requires no light filters during the curing process [6]. LED light sources have a curing depth of up to 4 mm and can potentially replace conventional halogen light sources. New generations of Ultra-Blue LEDs could possibly possess light intensity up to 1400 mW/cm^2 and the wavelength range of 400 – 470 nm. Photopolymer resins require a minimum wavelength range of 225 – 405 nm and light intensity of 280 mW/cm^2 [7]. Control of curing times to avoid under-curing and over-curing should be carefully monitored. Under-curing is the result of the absence of polymer chain bonds and over-curing is the result of light penetrating into successive layers which are greater than the specified layer thickness [8]. The design considerations with regards to mechanical and chemical properties of products manufactured from photopolymers are the tensile strength, flexibility, chemical resistance, harness and non-yellowing of the material [9].

The components manufactured through SLA processes possess increased mechanical properties when compared to machining of the same product produced from the same material. The anisotropic behaviour of orientated isotropic layering affects the mechanical properties of the build orientation [10]. The mechanical properties of the SLA material are improved through post-curing of up to four (4) hours. The Elastic modulus, elongation and elongation at break was improved through post-curing [11]. The main contribution of this research are summarized as follows:

- Investigation and development of a high volume MSLA 3D printer.
- Development of a resin replenishment system.
- Examination of a UV LED array platform for photopolymerization.
- Experimental and feasibility testing of the proposed 3D printer.

2 MSLA 3D PRINTERS IN LITERATURE

Masked Stereolithography, (MSLA), consists of an array of light masked by a Liquid Crystal Display, (LCD), film. The projected light is dispersed across a large surface area and cures multiple resin planes simultaneously parallel to the mask. Employing large surface area functionality of MSLA initiates the possibility of the production of multiple or large products in sprinting cycles [12].

An LED projection micro-MSLA 3D printer was suggested by [13]. The system utilized 395 nm LED light source to cure 0.3-micron resin layers with a resolution of 0.1-micron. The products produced were no greater than 1.5 mm in diameter. An alternative micro-MSLA 3D printer implementing a digital micromirror device to project patterns up to 10 microns was developed by [14].

A low-cost MSLA 3D printer that consisted of a hybrid laser/projector light source was proposed by [15]. The projector cures a large surface area of the resin. The boundary of the large surface is staggered due to the limited pixelation of the projector. The laser continues to cure the remaining staggered edge to produce a fine boundary. The proposed system utilizes an intelligent contour construction algorithm for the purpose of fine boundary curing of the laser.

3 CONCEPTUAL DEVELOPMENT

3.1 System integration

An MSLA 3D printing system was designed and developed to enhance the printing of large surface area parts and improve the surface finish of printed parts. The design consisted of four (4) components namely: The mechanical system, the system control, the UV projection system and the resin reservoir, (VAT), system (see Figure 1).

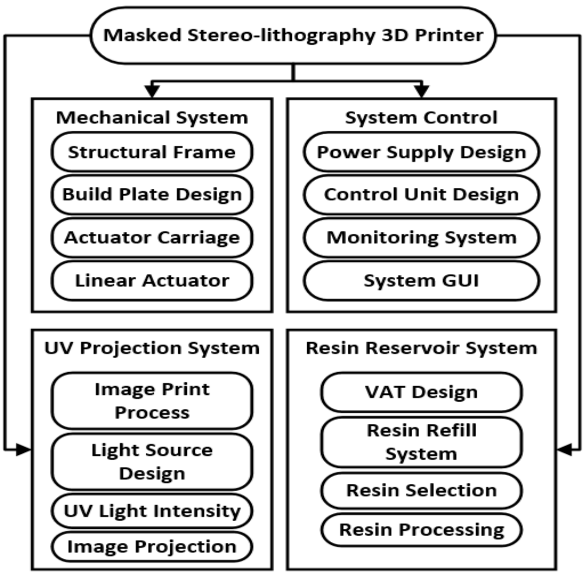


Figure 1 - System layout of the MSLA 3D printer.

The conceptual model was modelled to enhance modularity, ease of manufacturing and modern aesthetic design. The mechanical system was intended for rapid printing and part retrieval. The structural frame accommodates a UV LED projection mask (see Figure 2). An airtight and UV protection screen protects users against any air and light pollution that can potentially cause injury.

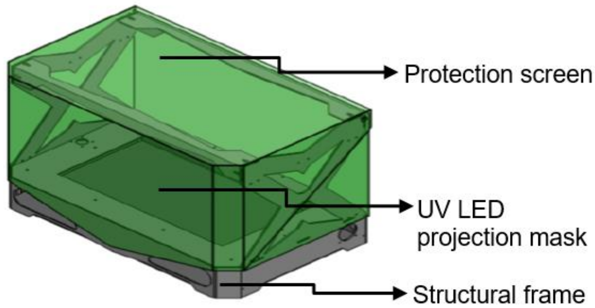


Figure 2 – Conceptual assembly of MSLA 3D printer.

The mechanical structure was designed to support a build plate, an actuator carriage, a linear actuator and structural frame (see Figure 3). The support of the structural frame houses a dampening material to reduce vibration transfer. The system was designed and constructed to reduce the amount of dynamic noise that affects the print quality.

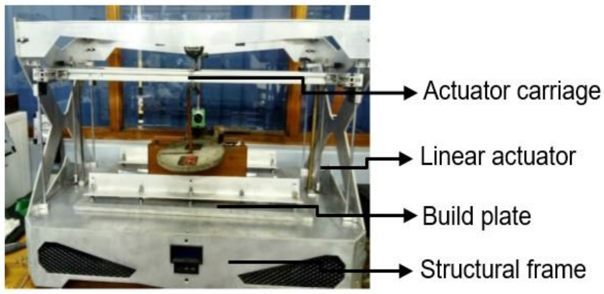


Figure 3 – Manufactured assembly MSLA 3D printer.

3.2 MSLA 3D printer development

The mechatronic architecture of the MSLA 3D printer consists of a mechanical system (as discussed previously), a software architecture and an electronic system (see Figure 4). Printing geometrics containing a CAD model is entered into the printing software. The model is sliced into layers and a path plan in a G-code format is generated for the Z-axis to travel vertically. The command code is transmitted to the Raspberry Pi microcontroller. The microcontroller sends all relevant data signals to the motor driver and the projection screen. User informational data to a display hub where the user monitors the print progress. The motor driver actuates the build plate vertically according to layering specifications. The projection screen regulates UV light intensity and resin replenishment. The process is completed through a finished 3D printed product.

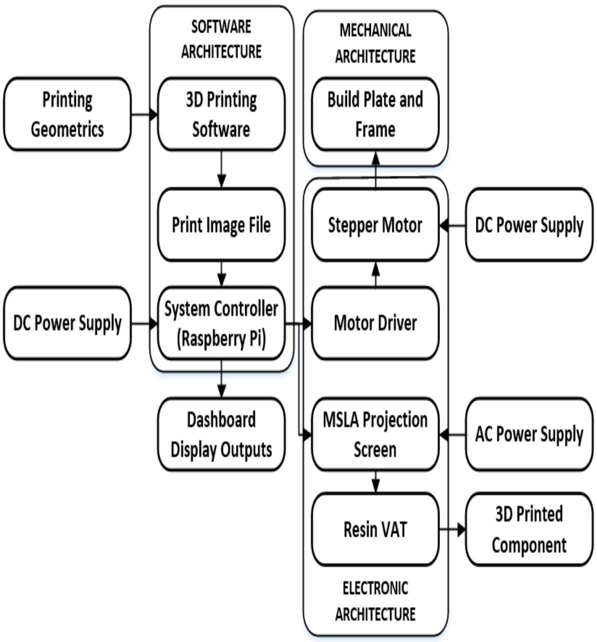


Figure 4 – The mechatronic architecture of the MSLA 3D printer.

The pseudocode described the closed loop operation program of the proposed MSLA 3D printer (see Figure 5). The process was initiated through the uploading of a 3D CAD model. The model was sliced and a path plan was generated. Each layer was printed according to cross-sectional dimensions. The command print followed the actuation of the Z-axis for vertical printing otherwise the print was ended if the parameter conditions were not met. If the layer to be printed was the last layer the print would be terminated otherwise the next layer was printed until the last layered was queued.

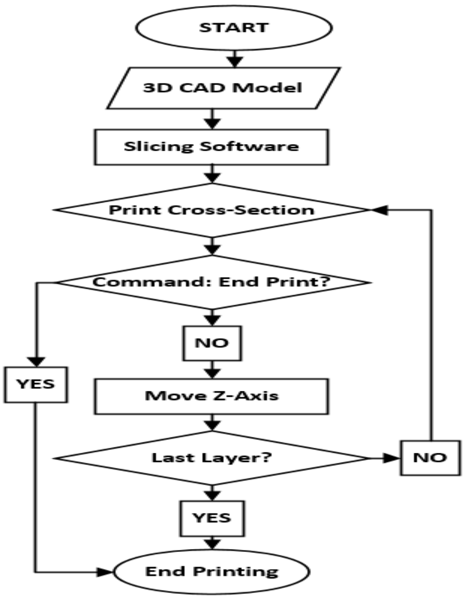


Figure 5 – Pseudocode of printing procedure.

3.3 Automated refilling system

The automated refilling system was designed to store uncontaminated resin and replenish the resin level as the liquid was consumed during the print (see Figure 6). The resin was pumped by means of displacing air in an airtight resin storage diaphragm through an air pump. The level was measured through an ultrasonic sensor and controlled the air pump when the resin level was low.

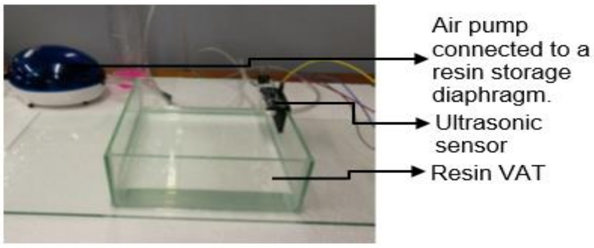


Figure 6 – Resin refilling system.

3.4 LCD screen with LED array

The LED projection mask was manufactured through soldering individual UV LEDs to a breadboard array panel (see Figure 7). The LEDs were placed 2 cm apart from one another. A 720p flat screen television set was dismantled the LCD screen was exploited. Light intensity and pigment of each pixel are controlled through the imaging software. The disadvantage of an LCD display was the limited resolution of the manufactured pixels. The assembled LED projection array board was inserted between the LCD screen and the backend panel of the television (see Figure 8). The height between the UV dispersion from the array panel and LCD mask was adjusted by a spacer. Height alterations were prepared for testing.

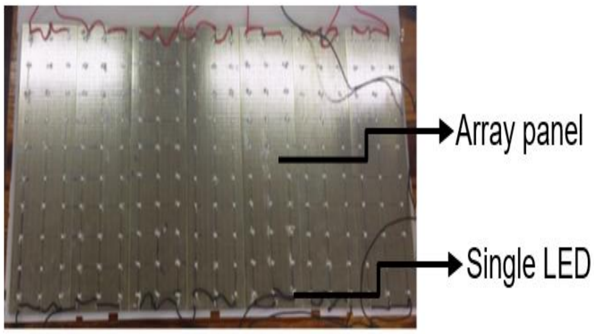


Figure 7 – Assembled LED array.

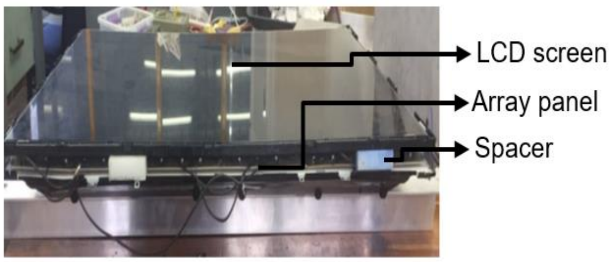


Figure 8 – Television LCD display mask.

4 TESTING OF MSLA 3D PRINTER

4.1 VAT refilling system testing

The experiment was configured to measure the sensitivity and accuracy of the ultrasonic proximity sensor and refill system. Liquid soap was determined to possess the same density as the photopolymer resin and was utilized during the experiment. The data captured was related to the change in volume over time and the change in sensor height over time. The testing was done multiple times to produce accurate results. The change in volume over time experiment illustrated a linear relationship between the increase in volume in relation to the time period passed (see Figure 9). The gradient decreases as the volume reached a certain amount, this is due to a back pressure build-up causes a reduction in the replenishment period. The experiment was repeated employing other experimental conditions. The fluid was altered to water for a second experiment and the level surface was changed to aluminium for a third experiment. The graph results show a linear behaviour between the change in height during refill and the time taken with the exceptional outlier (see Figure 10).

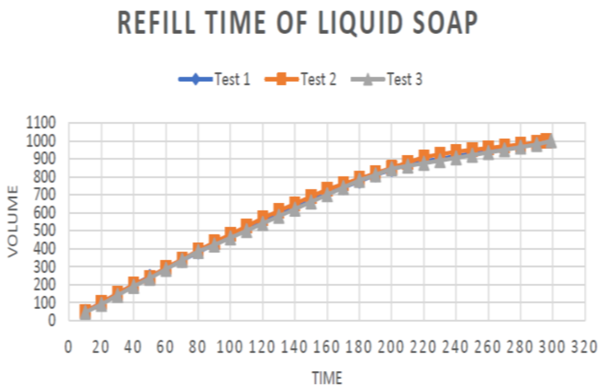


Figure 9 – Liquid soap volume vs filling time graph.

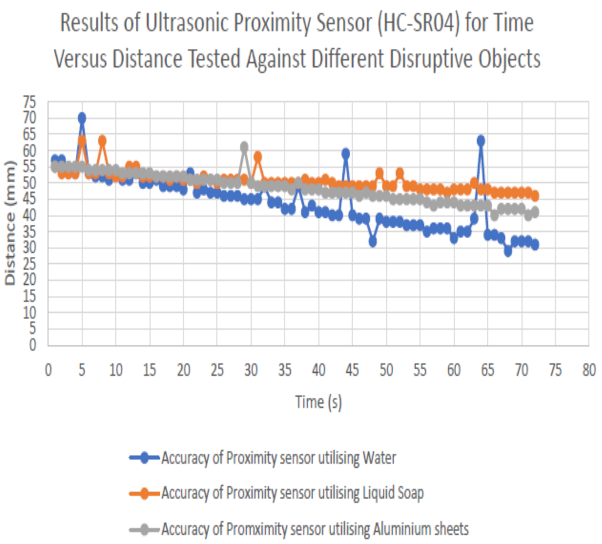


Figure 10 – Graph representing ultrasonic sensor during filling times.

4.2 Resin curing testing

The testing of resin curing was tested by applying a coating to a specified area on the bottom glass screen of the reservoir (see Figure 11). The UV LEDs was energized and the resin coating was exposed to UV light. The light intensity was altered through changed the height between the mask and the project panel through manipulating the spacers. The light intensity was measured per height alteration. The resin exposure was timed accordingly and plotted.

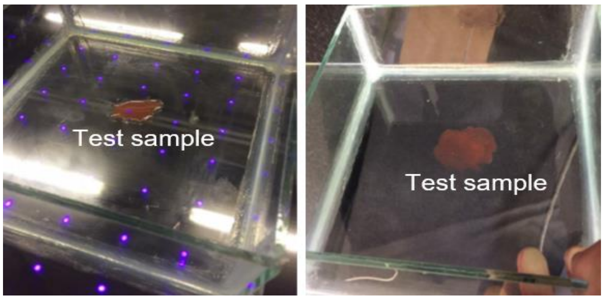


Figure 11 – Liquid resin test sample (Left) and cured test sample (Right).

The experiment was performed to determine the hardening time of the resin with respect to the wavelength emitted from the light source. One (1) Lux light intensity is equivalent to 0.0001464 mW/cm^2. The results showed a linear relationship between the light intensity and spacer distance with a negative gradient (see Figure 12). At 20 mm from the spacer, the light intensity altered minimally. Results were plotted for the hardening time versus the wavelength and generated a linear relationship with a negative gradient (see Figure 13). The critical results were tabulated and quantified the actual average hardening times according to the tested wavelength. The fastest curing rate was 15.3 seconds at a wavelength of 410 nm (see Table 1).

UV Wavelength (nm)	Avg hardening time (s)
390	54.7
400	28
410	15.3

Table 1 - Example of table centred across two columns

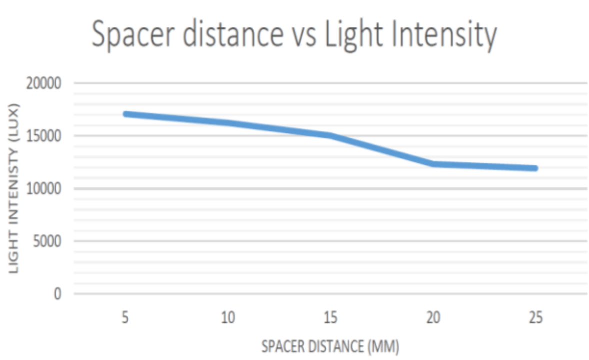


Figure 12 – Spacer distance vs Light intensity graph.

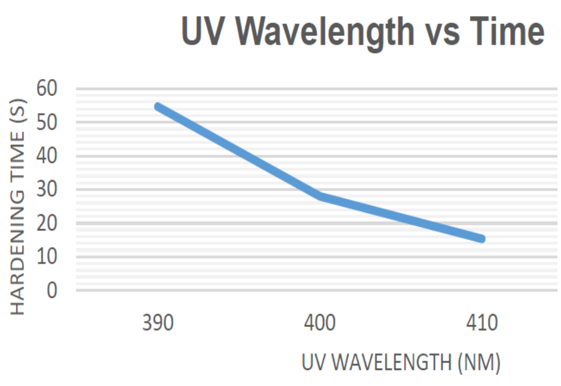


Figure 13 – UV wavelength vs Hardening time graph.

5 CONCLUSION AND RECOMMENDATIONS

The proposed system proved to be feasible through testing as an economical approach for MSLA printing. The printer surface area was adjustable and modular according to the required surface area size required. The refill system proved to sufficiently provide uncontaminated resin on-demand during the photopolymer consumption throughout the printing procedure. The testing provided evidence of sufficient resin curing for the feasible operation of the printer for production of components.

The light intensity of the of the UV projection was not sufficient for thorough printing. The light source intensity should be increased to a sufficient light concentration. The resolution of the LCD photomask was 720p and caused staggered edges when cured. The LCD resolution should be increased to a 4K screen to provide sufficient resolution for printing. Misaligned leadscrew caused observable vibration and chatter during printing and can potentially cause defects in surface finish. The lead screw should be redesigned and manufactured for stable linear actuation. The experimental testing should be evaluated with a complete projected image and full surface resin flood to print layers accordingly.

6 ACKNOWLEDGEMENTS

Acknowledgement is given to students Samitan Givindsamy, Sandile Goqo, Dhashan Naidoo and Ugendran Naidoo for their research contribution. The author's special thanks are extended to the staff of UKZN for the opportunity to present my work.

7 REFERENCES

[1] Padayachee, J., Bright, G., 2013, The Design of Reconfigurable Assembly Stations for High Variety and Mass Customisation Manufacturing, South African Journal of Industrial Engineering, 24/3, 43-57.

[2] Paulo, J, B., Ian, G., 2011, History of Stereolithographic Processes, Stereolithography: Materials, Processes and Applications, Berlin, Springer, 37-56.

[3] Amin, R., Knowlton, S., Hart, A., Yenilmez, B., Ghaderinezhad, F., Katebifar, S., Messina, M., Khademhosseini A., Tasoglu, S., 2016, 3D-printed microfluidic devices, Biofabrication, 8/ 2, 1-16.

[4] Tyler, F., 2015, High Definition 3D Printing – Comparing SLA and FDM Printing Technologies, The Journal of Undergraduate Research, 13/ 3, 9-25.

[5] Mark, V., Hermann, S., 2014, A New Micro-Stereolithography-System based on Diode Laser Curing (DLC), International Journal of Precision Engineering and Manufacturing, 15/10, 2161-2166.

[6] Kramer, N., Lohbauer, U., Garcia-Godoy F., Frankenberger, R., 2008, Light curing of resin-based composites in the LED era, American journal of dentistry, 21/ 3, 135-142.

[7] Calixto, L, D, R., Lima, M., Queiroz, R, S., Rastelli, A, N., Bagnato V, S., Andrade, M, F., 2008, Curing Depth of Composite Resin Light Cured by LED and Halogen Light-Curing Units, Laser Physics, 18/ 11, 1365–1369.

[8] O'Neill, P, F., Kent N, Brabazon, D., 2017, Mitigation and Control of the Overcuring Effect in Mask Projection Micro-Stereolithography, Proceedings of the 20th International ESAFORM Conference on Material Forming, Dublin.

[9] Shukla, V., Bajpai, M., 2004 Overview on basic chemistry of UV-curing technology, Pigment and Resin Technology, 33/5, 272-279.

[10] Hague, R., Mansour, S., Saleh N., Harris, R., 2004, Materials analysis of stereolithography resins for use in rapid manufacturing, Journal of Materials Science, 39/7, 2457-2464.

[11] Chantarapanich, N., Puttawibul, P., Sitthiseripratip, K., Sucharitpwatskul S., Chantaweroad, S., 2013, Study of the mechanical properties of photo-cured epoxy resin fabricated by stereolithography process, Journal of science and technology, 35/ 1, 91-98.

[12] Mao, Y., Miyazaki, T., Sakai, K., Gong, J., Zhu M., Ito, H., 2018, A 3D Printable Thermal Energy Storage Crystalline Gel Using Mask-Projection Stereolithography, Polymers, 10/10, 1-14.

[13] Zheng, X., Deotte, J., Alonso, M, P., Farquar, G, R., Weisgraber, T, R., Gemberling, S., Lee, H., Fang N., Spadaccini, C, M., 2013, Design and optimization of a light-emitting diode projection microstereolithography three-dimensional manufacturing system, Review Scientific Instruments, 83/12, 1-6.

[14] Lee, M, P., Cooper, G, J., Hinkley, T., Gibson, G, M., Padgett M, J., Cronin, L., 2015, Development of a 3D printer using scanning projection stereolithography, Scientific Reports, 23/5, 1-5.

[15] Zhou, C., Ye H., Zhang, F., 2015, A Novel Low-Cost Stereolithography Process Based on Vector Scanning and Mask Projection for High-Accuracy, High-Speed, High-Throughput, and Large-Area Fabrication, Journal of Computing and Information Science in Engineering, 15/ 1, 1-8.

8 BIOGRAPHY

Christian Basson holds an MSc Eng degree in Mechanical Engineering from the University of KwaZulu-Natal.He is currently doing his PhD in the Department of Mechanical Engineering. His specific focus and area of research are robotics and reconfigurable manufacturing technologies.

Professor Bright graduated with a BSc (Mechanical Engineering), MSc (Engineering) and PhD (Engineering) degree at the Ex-University of Natal in Mechatronics, Robotics and Advanced Manufacturing Systems. Professor of Mechatronics, Robotics and Advanced Manufacturing Systems since 2002.

Optimizing the Machining Strategy to Control the Residual Stress State of a Typical Ti-6Al-4V ELI Component

P.A. le Roux[1], R.F. Laubscher[1]

[1]Department of Mechanical Engineering, University of Johannesburg, South Africa.

Abstract

The surface integrity induced by a manufacturing process and its effect on the functional performance of parts has become more significant in recent years. This paper investigates the effect of cutting strategy in terms of the induced residual stress and resultant surface roughness of machined Ti-6Al-4V ELI (extra low interstitial) stepped shafts. Cutting speed is varied between the typical recommended cutting speed of 40m/min to an intermediate speed of 110m/min. Literature evidence suggests that the surface residual stress state may change appreciably towards the intermediate cutting speeds. Titanium alloys may typically show an increased compressive residual stress state at these intermediate speeds before becoming more tensile at elevated cutting speeds. This behaviour is linked to the ability of the heat generated due to the cutting process to dissipate via chip formation and conduction into the workpiece and tool. Increasing the cutting speed increases the heat generation rate and therefore the cutting temperature and the heat dissipation ratio between the tool and workpiece. Stepped shafts are therefore machined at various cutting speeds before being subjected investigation by XRD (X-Ray Diffraction) analysis and 2D surface profilometry as a function of cutting speed. Conclusions are presented, and recommendations made in terms of selection of a smart machining strategy of Ti-6Al-4V ELI to enhance functional performance.

Keywords

Ti-6Al-4V ELI; Residual stress measurements; surface roughness

1 INTRODUCTION

Ti-6Al-4V ELI has been widely adopted for medical, aerospace and military applications for its superior mechanical performance over a large temperature and applied stress range [1]. Titanium alloys typically display a compressive residual stress state after machining which is different to carbon steel that may typically display tensile stresses. A variety of published work indicated an increase in residual stress imposed on titanium alloys when machining at intermediate cutting speeds [2, 3, 4]. Titanium alloys may typically show an increased compressive residual stress state at these intermediate speeds before becoming more tensile at elevated cutting speeds [5, 6, 7, 3].

This behaviour is linked to the ability of the heat generated due to the cutting process to dissipate via chip formation and conduction into the workpiece and tool [8]. Increasing the cutting speed increases the heat generation rate and therefore the cutting temperature and the heat dissipation ratio between the tool and workpiece.

The machining of titanium alloys poses a significant opportunity to investigate the smart selection of a machining strategy to enhance part performance to a level commensurate with other post machining processing operations such as shot-peening. Turning was selected for this investigation due to its relative simplicity as it utilizes a single cutting edge in constant contact with the same directionality with the workpiece, thereby enhancing the thermal conductivity between the cutting tool and workpiece.

As turning produces a relatively constant lay in the circumferential direction, respective to tool geometry and wear, vibrations, etc. [9], it is proposed that the induced residual surface is directional in nature and that the ultimate functional performance of the part may also be directional.

This paper investigates the effect of cutting strategy in terms of the induced residual stress and resultant surface roughness of machined Ti-6Al-4V ELI (extra low interstitial) stepped shafts. Cutting speed is varied between the typical recommended cutting speed of 40m/min to an intermediate speed of 110m/min.

2 EXPERIMENTAL PROGRAMME

2.1 Workpiece material

Specimens were machined from a 32mm diameter Ti-6Al-4V ELI (Grade 23 titanium) round bar produced in accordance to the ISO 5832-3 / ATSM F136 standard. The material had the following chemical composition: 5,91% Al, 3,86% V, 0,08% Fe, 0,01% C, 0,01% N, 0,09% O, 0,001% H and 90,039% Ti. The material has a tensile strength of 955 MPa and yield strength of 862 MPa.

2.2 Specimen preparation

Specimens were cut and machined to length prior to CNC machining on an Efamatic RT-20 CNC lathe. Machining was performed using conventional flood cooling using physical vapour deposition (PVD) coated cemented carbide tip tools with chip breaking technology for roughing and finishing operations. The roughing operation reduced a 75mm length of the bar from a diameter of 32mm to 20mm, with a 0,25mm cutting depth per pass at a cutting speed of 55m/min, utilising a (80° rhombus Sandvik CNMG 12 04 08-XF GC15) C-shaped insert with a nose radius of 0.8mm. A tool change was performed to machine the near net shaped specimen (0,6mm oversize) using a (55° rhombus Sandvik DNMG 15 06 12-SMC 1115) D-shaped insert with a nose radius of 1,2mm at a cutting depth of 0,25mm. A final tool change was conducted for finishing parameters, utilising a new (55° rhombus) D-shaped insert to conduct the final three passes at a cutting depth of 0,2mm at a preselected cutting speed of 40m/min, 75m/min and 110m/min. A feed rate of 0.2 mm/rpm were used throughout. The final geometry of the specimen can be seen in Figure 1.

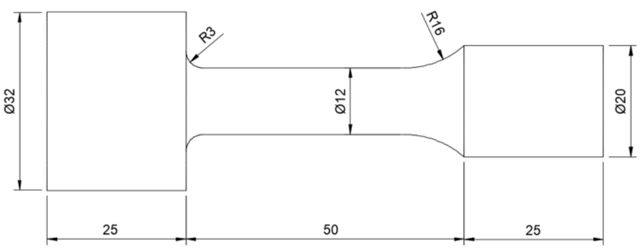


Figure 1 - Test specimen

2.3 Surface roughness measurement

Surface roughness measurements were conducted in accordance to the ISO 4288:1996 standard on the finished surface for the three cutting speeds. Three measurements were performed per specimen using a Jenoptik Hommel Etamic T8000 2D surface profilometer.

2.4 Residual stress measurement

Residual stress measurements were conducted by X-Ray Diffraction (XRD) analysis at three locations as indicated in Figure 2 per specimen. A Proto iXRD machine was implemented to analyse the residual stress with a Copper (Cu) X-ray source through a 2mm round X-ray aperture. The XRD measurement parameters were as follow:

- Voltage: 20kV
- Current: 4mA
- Filter: Nickel (Ni)
- Bragg angle: 139,69° on 231 hkl
- Exposure time: 3s
- Number of exposures per angle: 12
- Psi osc: 3°
- Calibration standard: ASTM: E915-10
- Calibration date: 8 March 2018

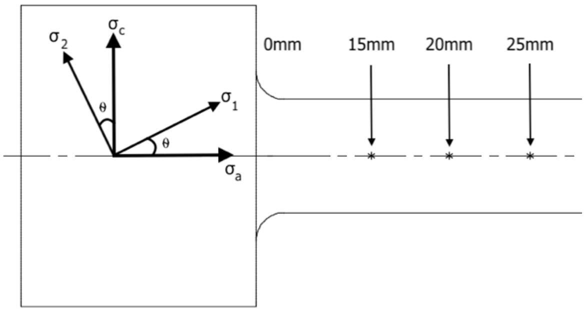


Figure 2 - XRD measurement locations and principle stress orientation

Measurements were taken at each of the three locations at an angle of 0, 45 and 90 degrees relative to the longitudinal axis of the specimen. The XRD results were then translated into both the principal stresses and the translation angle through plane stress analysis [10].

3 RESULTS

3.1 Surface roughness

The surface roughness measurements are shown in Table 1. The roughness values compared well to the ideal theoretical arithmetic average surface roughness (based on the nose radius and feed) of 4.2 μm. A decrease in all the surface roughness parameters were evident as a function of increased cutting speed from 40m/min to 110m/min. Indicating that changing a single cutting parameter can affect the surface roughness, which for this experimental exercise is advantageous to the surface roughness which may typically be linked to an enhanced fatigue performance.

Cutting speed	Ra	Rz	Rmax	Rmax/Ra	Rz/Ra
m/min	μm				
40	2,25	10,67	11,21	4,98	4,74
75	1,95	10,03	10,30	5,28	5,14
110	1,57	7,86	9,24	5,88	5,00

Table 1 - Surface roughness results

When comparing Rmax/Ra and Rz/Ra as a function of cutting speed, an increase in both ratios are observed. This is contradicting the norm that a reduction in surface roughness is associated with increased compressive residual stress.

Table 2 depicts a cross sectional view of specimens machined at 40, 75 and 110m/min which clearly indicates that grain pull-out and grain deformation occurred during machining at all 3 cutting speeds. Grain pull-out are typically associated with machining of Ti-6Al-4V alloys at cutting speeds ranging from 80-120m/min with a feed rates of between 0,25-0,4 mm/rev at a cutting depth of 0,25mm [11, 12]. As grain pull-out is a randomly occurring surface defect, it could be associated with the increase of Rmax/Ra and Rz/Ra ratios as a function of cutting speed.

SEM analysis of randomly selected specimens that were machined at each of the three cutting speeds revealed that pockmarks are visible at the faced surfaces that could be as a result of grain pull-out as depicted in Figure 3.

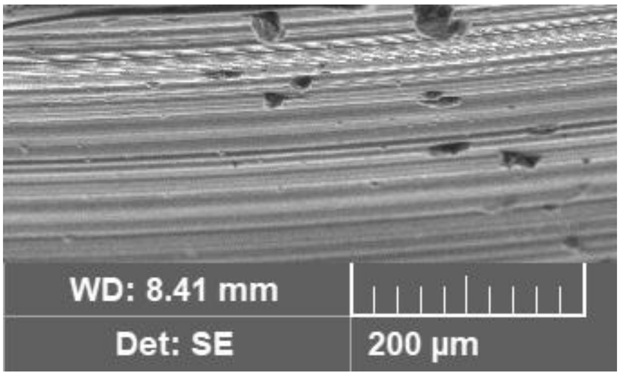


Figure 3 - Pockmarks on faced surface machined at 40m/min

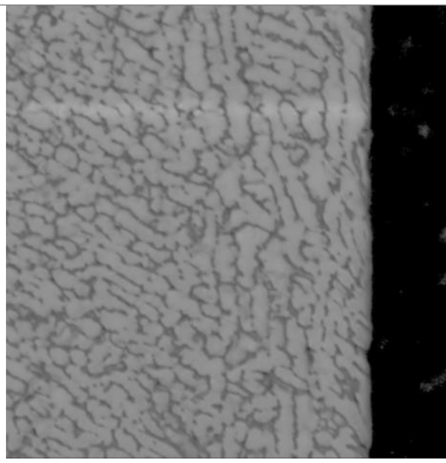
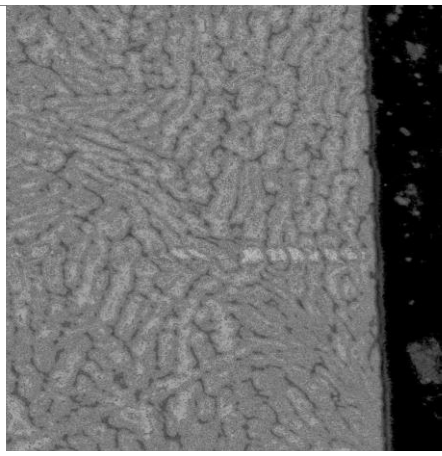
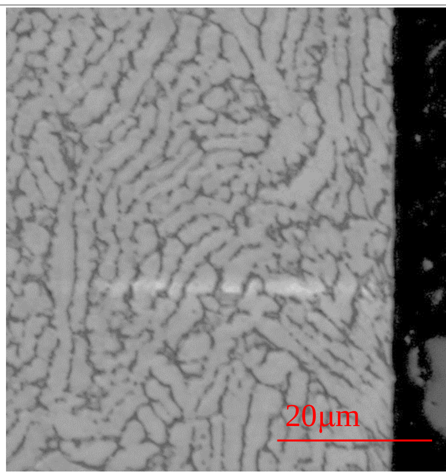
Cutting speed (m/min)	
40	
75	
110	

Table 2 - Near surface microstructure at different cutting speeds (100x magnification)

3.2 Residual stress analysis

The results of the residual stress measurements and subsequent analysis are presented in Table 3.

Cutting speed	σ_c	σ_a	σ_1	σ_2	Theta
m/min	MPa	MPa	MPa	MPa	Degree
40	-266,4	-254,2	-178,2	-342,4	47,1
75	-450,0	-260,5	-328,1	-382,4	64,9
110	-442,4	-389,8	-348,0	-484,3	54,1

Table 3 - Residual stress results

The data shows that the surface residual stress state is indeed compressive and that this increases significantly with rising cutting speed. The maximum compressive principal residual stress rises from approximately 340 MPa to approximately 480 MPa, a 41% increase (see Figure 4). The directional vector of this principal stress displays a mildly decreasing trend but hovers between 47,1 to 64,9 deg. This implies that it is aligned in a direction somewhere between the main cutting direction and the transverse direction but more closely aligned with the main cutting direction.

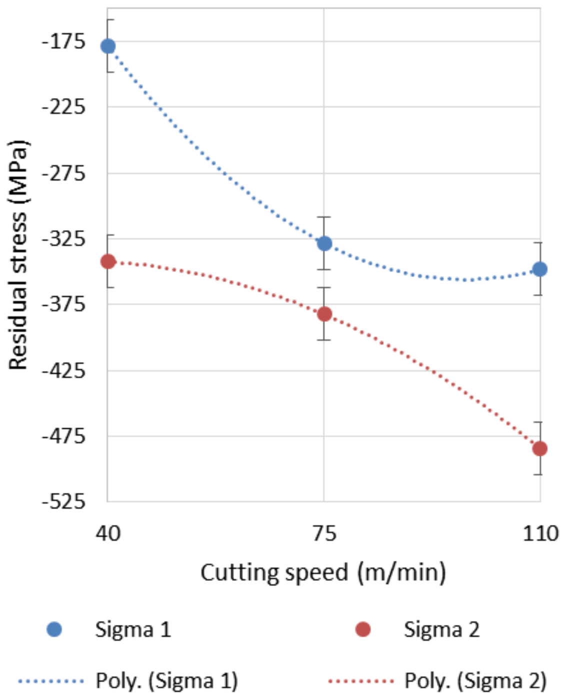


Figure 4 - Principal stresses vs cutting speed

The fact that the main principal stress is not more closely aligned with the cutting direction is indicative of the single point cutting process. Essentially cutting occurs along a leading edge of the tool nose as it is fed into the material for simple outside turning. This introduces a transverse plastic strain component that sums as a vector along with the pure circumferential strain to achieve the final resultant vector or direction.

The fact that the principle stress direction seems to indicate a decreasing trend may be indicative of the heat generated component of strain starting to play a more significant role at the higher cutting speeds.

It is also worth noting that the direction of the principle stress state denotes the direction of cut and may therefore be used to identify the cutting direction.

Typically, residual stresses are formed during machining due to two main contributing factors: A residual stress is induced during the shearing action associated with the cutting action. This component is typically directional with the principal stress aligned with the cutting direction. Upon this is superimposed a residual stress due to uneven cooling associated with frictional heat created due to the tool workpiece interaction. This component is less directional and may be the dominant factor for materials that display a poor heat conductivity such as titanium alloys.

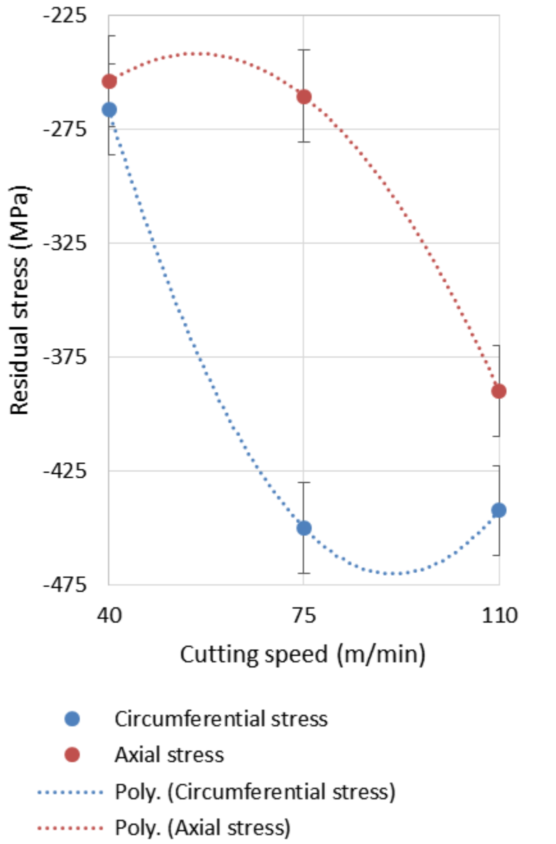


Figure 5 - Circumferential and axial residual stress vs cutting speed

The sum of these two effects results in the final stress state. An increase in cutting speed leads to an increased rate of heat production that causes localised temperature increases both at the tool tip and the workpiece. Due to the lower conductivity of titanium alloys this heat takes longer to dissipate before it is then eventually quenched by the coolant from a higher temperature leading to an increased residual stress state.

Figure 5 **Error! Reference source not found.**indicates the variation of the circumferential stress σ_c (cutting direction) along with the axial direction σ_a. In both cases significant increases in the compressive residual stress is realized with an increase in cutting speed from 40 m/min to 110 m/min. The residual stress aligned with the cutting direction increases from -266 MPa to -442 MPa or 66%. For the axial direction an increase of 53% is realized.

Typically during dynamic loading the compressive residual stress needs to be overcome first for defects to form and propagate effectively. This implies that depending on how the stepped shaft is loaded a minimum and maximum residual stress increase between 40% to 66% is realized that will add to fatigue life. The results indicate that all the load cases that a typical stepped shaft may be subjected to i.e torsional, bending, axial and or a combination of the above will all benefit from an increased compressive residual stress state due to an intelligent selection of the cutting strategy that includes an increase of cutting speed to intermediate levels.

4 CONCLUSION

The functional performance of machined mechanical parts are closely related to their resultant surface integrity due to the processing. Improved surface roughness and a conducive residual stress field suggest an improved fatigue performance. The final surface integrity achieved is closely related to the cutting strategy employed.

The results indicate that the surface finish could be moderately improved due to an increase in cutting speed. However, the effect of pockmarks and grain pull-out has to be investigated through fatigue testing.

An XRD analysis of outside turned stepped shafts indicated that residual stress changes of up to 66% could be realized by increasing the cutting speed from 40 m/min to 110 m/min. Surface compressive residual stresses of op to 484 MPa was realized along the principal directions. These were aligned at approximately 55 degrees relative to the main cutting direction. A minimum of 348 MPa (compressive) was measured at right angles to these.

The results indicate that by intelligently selecting an appropriate cutting speed ranging from the typically recommended value of 40 m/min to the intermediate speed of 110 m/min and appreciable modification of the residual stress field could be engineered.

Applications that require low residual stress fields such as in corrosive environments would be better suited with a final finishing strategy utilizing lower cutting speeds. Dynamic loading applications on the other hand would benefit from an elevated compressive residual stress field.

This paper indicates that a change in cutting strategy imposed on the finishing cutting operation can have a significant effect on the surface roughness and residual stress state induced into the near surface of machined components from Ti-6Al-4V ELI. Other materials with different conductivity and material properties will be different.

5 REFERENCES

[1] M. Peters, J. Kumpfert, C. Ward and C. Leyens, “Titanium Alloys for Aerospace Applications,” in *Titanium and Titanium Alloys: Fundamentals and Applications*, WILEY-VCH GmbH & Co. KGaA, 2005, pp. 419-427.

[2] D. Madyira, R. Laubscher, N. Janse van Rensburg and P. Henning, “High speed machining induced residual stresses in Grade 5 titanium alloy,” *Journal of materials: Design and applications,* vol. 227, no. 3, pp. 208-215, 2013.

[3] J. Puerta Velásquez, A. Tidu, B. Bolle, P. Chevrier and J. Fundenberger, “Sub-surface and surface analysis of high speed machined Ti–6Al–4V alloy,” *Materials Science and Engineering,* no. 527, pp. 2572-2578, 2010.

[4] D. Umbrello and G. Rotella, “Fatigue life of machined Ti6Al4V alloy under different cooling conditions,” *CIRP Annals - Manufacturing Technology,* 2018.

[5] Z. Xiangyu, L. Zhenghui, S. He and Z. Deyuan, “Surface Quality and Residual Stress Study of High-speed Ultrasonic Vibration Turning Ti-6Al-4V Alloys,” *4th CIRP Conference on Surface Integrity (CSI 2018),* pp. 79-82, 2018.

[6] G. Ortiz-de-Zarate, A. Madariaga, A. Garay, L. Azpitarte, I. Sacristan, M. Cuesta and P. J. Arrazola, “Experimental and FEM analysis of surface integrity when broaching Ti64,” *4th CIRP Conference on Surface Integrity (CSI 2018),* pp. 466-471, 2018.

[7] J. Holmberg, J. M. Rodríques Prieto, J. Berglund, A. Sveboda and P. Jonsén, “Experimental and PFEM-simulations of residual stresses from turning tests of a cylindrical Ti-6Al-4V shaft,” *4th CIRP Conference on Surface Integrity (CSI 2018),* pp. 144-149, 2018.

[8] S. Patil, S. Joshi, A. Tewari and S. S. Joshi, “Modelling and simulation of effect of ultrasonic vibrations on machining of Ti6Al4V,” *Ultrasomics,* pp. 694-705, 2014.

[9] B. Griffiths, Manufacturing Engineering Modular Series: Manufacturing Surface Technology: Surface Integrity & Functional Performance, London: Penton Press, 2001.

[10] G. E. Dieter, Mechanical Metallurgy [SI Metric Edition], London: McGraw-Hill, 1998.

[11] X. Liang, Z. Liu and B. Wang, “State-of-the-art of surface integrity induced by tool wear effects in machining process of titanium and nickel alloys: A review,” *Measurement,* vol. 132, pp. 150-181, 2018.

[12] J. Hughes, A. Sharman and K. Ridgeway, “The Effect of Cutting Tool Material and Edge Geometry on Tool Life and Workpiece Surface Integrity,” *The Journal of Engineering Manufacture,* 2006.

6 BIOGRAPHY

Rudolph Laubscher holds a D.Ing from RAU. He is currently an associate professor at the University of Johannesburg. His research interests are associated with physical metallurgy, manufacturing and FEA.

Pieter le Roux holds a M. Eng degree in Mechanical Engineering from the University of Johannesburg. He is currently a doctoral student at the University of Johannesburg. His specific focus and area of research interest is in the surface integrity and fatigue performance of titanium alloys.

Development of an ANSYS Simulated Discrete Machining Approach to Titanium Milling.

K-A. Airey[1], L. Lobmeyer [2], R. Kuppuswamy[1], A. Hoffmann[2], N. Vilakazi [1]
[1]University of Cape Town, South Africa
[2] Frankfurt University of Applied Sciences, Germany

Abstract

The next generation of fuel-efficient aircraft is to employ significant tonnage of titanium alloys (20% by weight [1]) in critical components to take advantage of titanium's superior strength-to-weight ratio of 260 – 300 kN.m/kg, high temperature performance of up to 400°C and corrosion resistance. Hindering this implementation are the challenges of machining titanium - including cutting forces rising to 5 times higher than conventional material machining [2]. Shear banding and sever plastic deformation indicate twice the plastic strain is generated, with consequentially high tool wear generation, shape errors of 2µm and surface roughness values twice that of acceptable standards [3]. General trends in modulation-assisted machining (MAM) show marked improvements to cutting of titanium, with corresponding reductions to tool wear, cutting temperatures, machining forces and cycle times. Controlling the functional attributes of machined surfaces, regarding the microstructural and residual stress (surface integrity) changes, allows the simultaneous increasing removal rates and reducing tool wear. Understanding the generation of these cutting behaviours is significantly advanced through the application of advanced finite element analysis and simulation modelling, which provides a holistic understanding of the machining process without exposing valuable equipment to potentially damaging machining conditions. This characterisation of the complex deformation and microstructure of Titanium as it is machined, facilitates the predictions of severe plastic deformation, chip morphology and allows the better design of optimal machining performance. Experimentation, with in situ measurement of strain, strain rate and temperature, complements this characterisation, enabling model development and validation, while providing phenomenological insights into microstructure evolution of SPD. The application of this understanding to discrete machining of titanium alloys is expected to improve product quality and productivity across low-frequency modulation in machining (MAM).

Keywords: Ansys, Discrete Machining, Milling, Titanium alloy, Ti6Al4V, Severe Plastic Deformation (SPD), Finite Element Analysis (FEA.

1 BACKGROUND

Discrete machining is a powerful machining technique, that records performance improvements through the interruption of the cutting action (Modulation-Assisted Machining -MAM) of 50% cutting force reductions in turning applications, with consequent temperature reductions[4] and overall tool wear improvements. Applying this technique to difficult-to-machine materials – including titanium, would reduce the formation of the usually aggressive machining environment typical of super-hard materials. This preserves the chemical stability of cutting tools by moderating the conditions under which chemical affinity develops between the workpiece and the cutting tool. Optimising and testing the success of these machining parameters under experimental conditions can be destructive and costly. Virtual simulation is far more efficient and powerful in determining parameter optimisation, once calibrated.

2 INTRODUCTION

This study focuses on the development of a simulation model for discrete milling of titanium.

For the purposes of this study, simulations will be conducted through the ANSYS software platform.

2.1 MAM as it applies to simulation building

The key aspects of simulating discrete milling, as opposed to the traditional simulations of discrete turning (for example), focus on the progression from orthogonal cutting actions to oblique actions. This is achieved by the addition of rotatory motion to the cutting tool, added via the explicit dynamics application within ANSYS. This model is then extended through the addition of discrete machining principles. For 1D MAM this entails adding a cyclical vibration modulation to the cutting tool velocity. More frequently discrete machining is applied as an elliptical cutting action, with vibrations on both the velocity and feed directions.

The ANSYS MAM simulations focused on the generation of vibration in the direction of the milling feed, with the model assuming a face milling operation

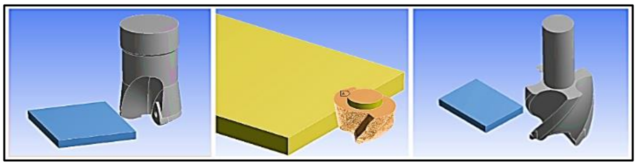


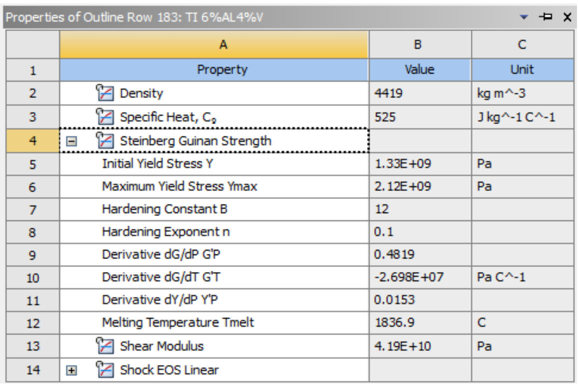
Figure 1 - 3D Milling Simulation Models

The workpiece material is TI-6Al-4V and the material properties can be updated, or customised from the onboard ANSYS default [5]:

Symbol	Property	Unit	Value
σ_y	Tensile Strength, Yield	MPa	880
σ_{UTS}	Tensile Strength, Ultimate	MPa	950
τ	Shear Strength, Ultimate	MPa	550
ρ	Density	$Kg.m^{-3}$	4419
E	Modulus of Elasticity	GPa	113.8
ν	Poisson's Ratio		0.342
K_{IC}	Fracture Toughness	$MPa.m^{1/2}$	75
C	Specific Heat	J/g°C	0.5263
λ	Thermal Conductivity	W/m.K	6.7
$\beta_{transus}$	Beta Transus	°C	980

Table 1 - Ti-6Al-4V Material Properties [6]

With the onboard value for material properties given as [5]:



Properties of Outline Row 183: TI 6%AL4%V

	A	B	C
1	Property	Value	Unit
2	Density	4419	kg m^-3
3	Specific Heat, C_p	525	J kg^-1 C^-1
4	Steinberg Guinan Strength		
5	Initial Yield Stress Y	1.33E+09	Pa
6	Maximum Yield Stress Ymax	2.12E+09	Pa
7	Hardening Constant B	12	
8	Hardening Exponent n	0.1	
9	Derivative dG/dP G'P	0.4819	
10	Derivative dG/dT G'T	-2.698E+07	Pa C^-1
11	Derivative dY/dP Y'P	0.0153	
12	Melting Temperature Tmelt	1836.9	C
13	Shear Modulus	4.19E+10	Pa
14	Shock EOS Linear		

Figure 2 - ANSYS Ti-6Al-4V Material Properties

The cutting tools have been modelled off an indexable SECOMax tool R217.21-1820.0-R100.2A and an end mill.

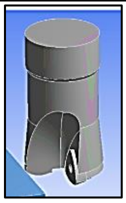
		Indexable Tool	End Mill
			
Material		CBCN	CBCN
Diameter (mm)	D	10	8
Cutting Edges	N	2	4
Speed Range (rpm)	V	0-10 000	0-30 000

Table 2 - Cutting Tool Properties

Some complex tool geometries were simplified from assembly models, to single solid models in order to limit the processing time of the software simulation (e.g. indexable cutting tool). A reduced material model was also developed to isolate a single cutting edge with the intention of reducing the simulation time required. On completion of simulation model optimisation and calibration, the simplified model can be replaced by the fully assembly solid model.

2.2 Experimental Configuration

The experimental setup was developed at the University of Cape Town's Advanced Manufacturing Laboratory and is configured as below:

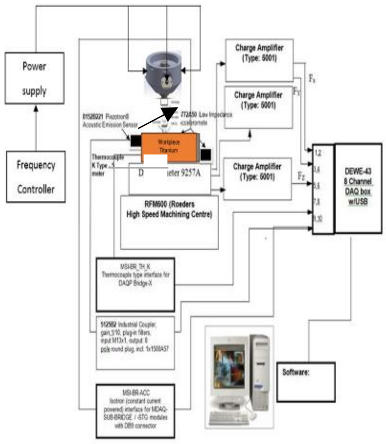


Figure 3 - Proposed HSM experimental system plan

The setup informs the development of the simulation model in terms of; relative displacements of cutting actions, application of vibration modulations and relative interaction of these components with the workpiece. The frequency of modulation simulation is limited by the available laboratory equipment. E-501

Series modular piezo controller, single channel, 90-120/220-264V AC, 50-60Hz [7]

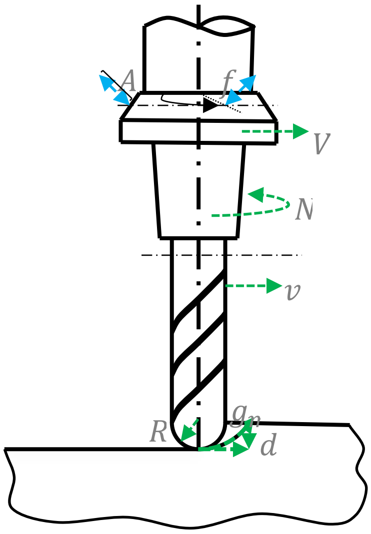


Figure 4 - Schematic Analysis of Milling Parameters

Operating Parameter	Sym	Unit	Value
Travel range at 0-100V	A	µm	15
Push Force Capacity	F_{PIPus}	N	1000
Pull Force Capacity	F_{PIPull}	N	50
Torque on tip	To	Nm	0.35
Resonant Frequency (No Load)	f_{reson}	Hz	18
Length	l	Mm	32

Table 3 - Piezo actuator P-840.10 Series

2.3 Theoretical Approach

The theoretical basis of the simulation model is based on the 2D application of discrete machining applied to a milling configuration. This results in both velocity, as well as feed modulation. The dimension characterisation of the modulation action as applied to milling, is given as the vertical tool position relative to the workpiece $\mathrm{Z(t)}$, and its subsequent velocity $\mathrm{Z'^{(t)}}$:

X Position $$\mathrm{X(t)} = A_x\cos(2\pi \mathrm{f t}) + \mathrm{Vt} \quad (1)$$

X Speed $$\mathrm{X'(t)} = -2\pi\mathrm{f}A_x\sin(2\pi\mathrm{ft}) + \mathrm{V} \quad (2)$$

Y Position $$\mathrm{Y(t)} = A_y\cos(2\pi\mathrm{ft}) + \mathrm{Vt} \quad (3)$$

Y Speed $$\mathrm{Y'(t)} = -2\pi\mathrm{f}A_y\sin(2\pi\mathrm{ft}) + \mathrm{V} \quad (4)$$

Z Position $$\mathrm{Z(t)} = A_z\sin(2\pi\mathrm{ft}) \quad (5)$$

Z Speed $$\mathrm{Z'^{(t)}} = 2\pi\mathrm{f}A_z\cos(2\pi\mathrm{ft}) \quad (6)$$

Where Z represents the vertical direction, and A_z represents vertical amplitude, f is the frequency, t the time and V the machine speed. A 2D duty cycle (DC) for VAM calculates the length of the elliptical toolpath cut during one vibration cycle as is given as:

$$DC = \frac{arc(\theta_2 - \theta_1)}{2\pi\sqrt{(A^2 + B^2)/2}} \quad (7)$$

Vector analysis of the force as acting on the tool is resolved as follows:

Z-Axis Force $$F_z = (F_1 + F_2 + F_3)sin\alpha \quad (8)$$

X-Axis Force $$F_x = F_1 sin\beta - (0.5\,F_2 sin\beta + 0.5F_3 sin\beta) = sin\beta(F_1 - 0.5(F_2 + F_3)) \quad (9)$$

Y-Axis Force $$F_y = (-F_2 sin\beta sin60 + F_3 sin\beta sin60) = 0.866 sin\beta(F_3 - F_2) \quad (10)$$

For the purposes of ensuring the simulation will in fact be modelling modulated cutting action, the separation threshold (a.k.a. critical upfeed velocity) must be preserved throughout the cutting action. This characterisation is defined as the limit which ensures contact between the tool and workpiece interface is unequivocally broken during each vibration cycle.

$$2\pi f A < V \quad (11)$$

If the critical velocity strays above this threshold, the tool does not adequately leave the rake face, and discrete cutting action is not achieved. This value must also be balanced with the appropriate depth of cut. Unbalanced application is evident in previous simulations conducted in the laboratory [8], where simulation of subsonic orthogonal discrete machining using a single edged cutting tool acting on titanium material, with finite element analysis was conducted using ANSYS software.

Parameter	Symbol	Value	Unit
Amplitude	A	0.02	
Frequency	f	0, 200, 500, 1000	Hz
Cutting Speed	V	50, 100,175, 200	m/min
Dynamic friction	μ	0.3 (assumption)	

Table 4 - Prior ANSYS Modelling: Discrete Machining Parameters

One of the fundamental assumptions this study makes is that the model is isothermal in nature. As one of the fundamental advantages of discrete machining is that it improves machinability by reducing heat generation, this model is useful only in showcasing the power of software modelling as a tool, rather than being an accurate representation of FEA discrete machining of Ti6Al4V

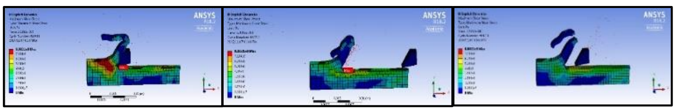


Figure 3 - Discrete Machining model at 175 m/min at 200Hz and 0.02 amplitude.

This study exhibits a good example of how modelling can detect incomplete withdrawal of the cutting tool from the cutting interface, as indicated by A in the image below:

To correct for this separation threshold, the simulation parameters that have been selected to ensure discrete machining [9], [10], are given below, with the parameter range for simulation chosen to suit Taguchi Methods.

Chip thickness	Amplitude	Frequency	Speed
0.1mm	0.05mm	50Hz	4000rpm
0.15mm	0.15mm	1000Hz	30000rpm

Table 5 - Simulation Parameters

The upfeed index (f_{up}) characterises the distance moved by the tool with respect to the workpiece per one vibration cycle:

$$f_{up} = \frac{V}{f} \tag{12}$$

While the Horizontal Speed Ratio (HSR) is given as:

$$HSR = \frac{V}{2\pi f A} \tag{13}$$

Theoretically, if both f_{up} and HSR are maintained at low values, the surface roughness is correspondingly lower. Analysis of the cutting forces considers a non-zero force to indicate the tool is contacting the workpiece and engaged with active cutting processes. Cutting forces can thus be characterised as peak or average values. Average values are presented as a

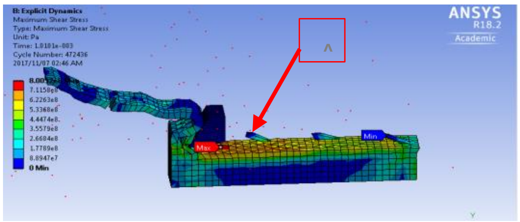


Figure 6 - Incomplete Withdrawal ANSYS Simulation

proportion of the peak correlating to the duty cycle. This value is consistently higher than other methods which seek the integral of the signal for the period of interest as this method ignores the dynamic response of both the mearing and cutting systems.

Simulation analysis of this machining progression must be seen to account for the theoretical understanding proposed through previous literature studies, where effects such as severe plastic deformation (SPD), governed by the Hall-Petch equation, are present at the cutting interface of the machined surface and facilitate our understanding of chip removal mechanisms.

$$\sigma_{yield} = \sigma_{friction} + Ad^{-1/2} \tag{14}$$

Where, σ_{yield} is the yield strength, $\sigma_{friction}$ is the friction stress, A a constant and $d^{-1/2}$ is the diameter of the workpiece material grain [9], [10].

2.5 ANSYS Simulation Model

Given the characteristics of the milling process as of short-duration and friction-based body interactions, the cutting actions were simulated using the ANSYS® explicit dynamics suite [11]. Since the explicit dynamics suite was developed to get insights into short term events, the model is solved with the Euler method and a fixed mesh which is well suited for severe deformations. In addition, the meshfree method of Smooth particle hydrodynamics (which is developed for hypervelocity impacts or crack propagation in brittle materials) was included in the analysis selection.

The Ansys explicit dynamics Settings were chosen as follows:

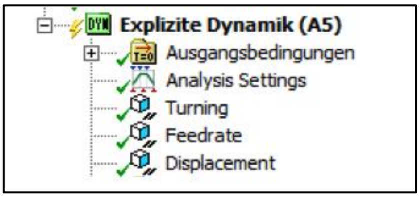


Figure 7 - Explicit Dynamics Settings

The analysis duration was set as t_{max}= 0,05 seconds which defines the number of turns for the process for both tools as [12]:

Tool 1: (30.000 rpm): 25 turns (9000°)

Tool 2: (10.000 rpm): 8.3 turns (3000°)

Assuming the equation for the feed rate is given as [12]:

$$v_f = N * z * f_z \tag{15}$$

where N – turning speed [m/min]; z – number of cuttings, and f_z – tooth feed [mm]. The effects of an interrupted cutting with a linear feed and a nonlinear feed can be analysed. The initial focus of this simulations was a basic understanding of a feed direction modulated milling process. Superimposing a sine curve movement, one can identify two cases for the feed rate setting:

Linear feed rate: for every time step

$$x = t * v_f \tag{16}$$

Linear feed rate + sine movement: for every time step:

$$x = t * v_f + 0{,}02mm * sin(t * 2\pi * 1000Hz) \tag{17}$$

The overall movement in second case can be seen in the feedrate modelling below, as the grey line combining the sine movement (blue) and the actual feed rate (orange):

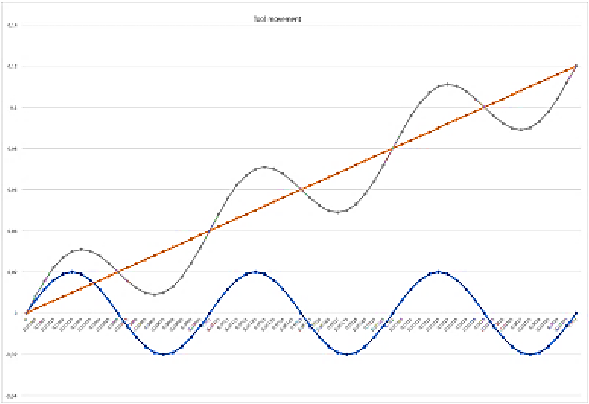


Figure 8 - Linear Feedrate & Sine Movement

To be able to compare the results for the simulations the model was asked to solve the following:

- Total and directional deformations
- Plastic strain
- Maximum principal and equivalent stress
- Maximum temperature
- Energy

The results should allow conclusions about the influence of a modulated assisted machining (MAM) application of a titanium milling process on tool temperature, surface characteristics and chip thickness.

3 RESULTS

The simulation shows agreement with prior temperature improvements [13] [14]. The simulation model shown below exhibits the successful formation

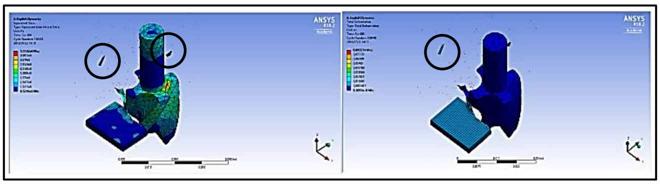


Figure 9 - Left: Equivalent Stress; Right: Total Deformation

and ejection of chips from the cutting interface (circled in the image above). This allows for more detailed chip analysis to be conducted following experimental calibration and simulation validation.

As anticipated from the titanium material properties, there is a strong link between the temperature and chip thickness of the process as evidenced in Figure 10.

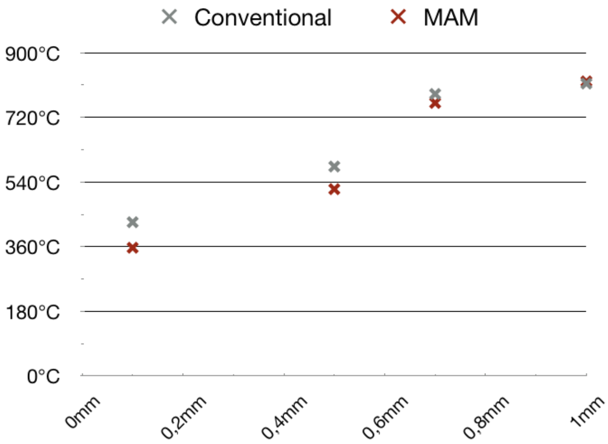


Figure 10 - Temperature results for conventional and discrete machining with increasing chip thickness.

At 30000rpm with a chip thickness of 0.1mm and discrete machining parameters of 50 Hz and an amplitude of 0.05mm, the simulation clearly exhibits interrupted cutting actions, with clear chip ejection. The focused tool edge model created, has also reduced simulation time from 1.5 days, for full model simulation, to 3 hours for the reduced model. This reduced model shows two clear cutting engagements for the arc of the cutting path (Indicated with arrows), as well as a clear ejection of a chip from the cutting interface (circled).

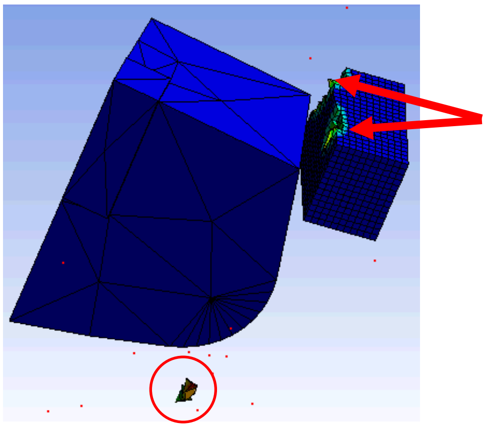


Figure 11 - Reduced Mass Simulation model for discrete machining

a)

Results	
Minimum	20 °C
Maximum	820,19 °C
Minimum Occurs On	Plate
Maximum Occurs On	Plate

b)

Results	
Minimum	20 °C
Maximum	814,72 °C
Minimum Occurs On	Plate
Maximum Occurs On	Plate

c)

Results	
Minimum	20 °C
Maximum	428,67 °C
Minimum Occurs On	Plate
Maximum Occurs On	Plate

d)

Results	
Minimum	20 °C
Maximum	357,81 °C
Minimum Occurs On	Plate
Maximum Occurs On	Plate

Figure 4 - Calculated Temperature at a) chip thickness 1mm conventional machining b) chip thickness 1mm discrete machining c) chip thickness 0.1mm conventional machining d) chip thickness 0.1mm discrete machining

Comparing a) and c) or b) and d) this link verifies the calculated results, because as one can see the temperature decreases drastically by reducing the chip thickness. As the results in Figure 12 and the graphical evaluation of all the simulation in Figure 13 show, the effect of discrete machining on the temperature are bigger if the chip thickness becomes smaller.

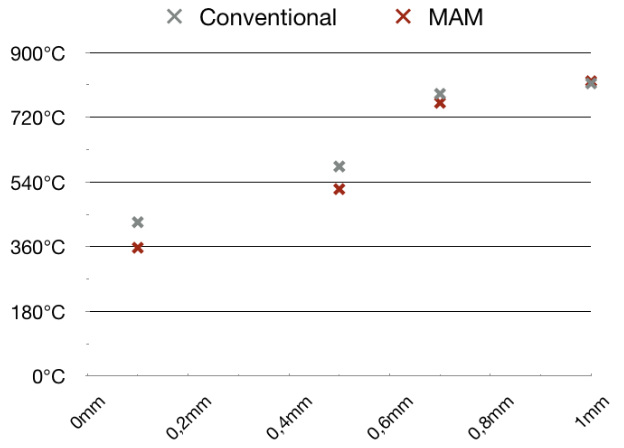


Figure 5 - Temperature results for conventional and discrete machining according to chip thickness

Looking at the average temperature decrease in percentage in Figure 14, one can see the strong link between temperature and chip thickness again, while it clearly indicates that discrete machining has positive effects on the temperature if the chips are small. Furthermore, the results indicate that a positive impact of discrete machining can only be expected if the material removal rate is small.

To analyse the impact of discrete machining on the chip formation, the next steps for the project include further simulations with a smaller simulated duration but a higher resolution. These simulations identify the differences between the chip formation during conventional and discrete machining. Moreover, the simulations will analyse the system regarding stress on the tool and workpiece and cutting forces.

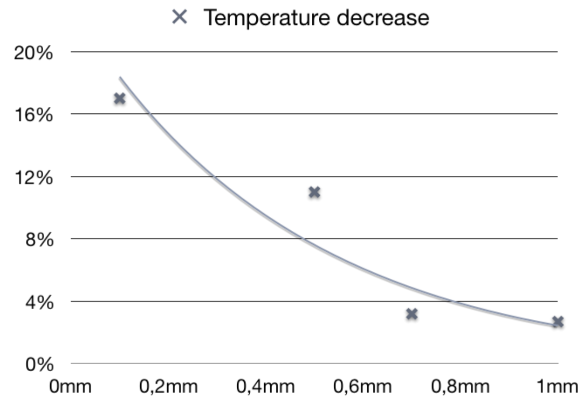


Figure 6 - Average Temperature decrease comparing conventional and discrete machining according to chip thickness

4 CONCLUSION

This study shows successful simulation of discrete machining utilising the ANSYS software. There is clear depiction of the formation of multiple, discrete chips adhere to the separation threshold requirements of discrete machining. The simulation also successfully ejects this chip from the cutting interface, which will allow for future characterisation the chip size and formation that can be supported by experimental validation.

The mesh size selected shows clear colour graduations that track the deformation of the material removed from the chip site.

The reduced mass model has been hugely successful in reducing processing time to 8% of the original time frame and has enabled to expansion of the number of testing parameters to be analysed in the future.

The depiction of the thermal generation by the simulation shows expected conventional machining values reaching 720°C. The deviation in thermal generation when modulation vibration is not present, conforms to prior research in this field that anticipates temperature reduction under discrete machining. In this case, the temperature reduction ranged from 4% to 20%.

5 RECOMMENDATIONS

The motivation of discrete machining is that, as long as the tool material possesses sufficient toughness, a discontinuous mode of material removal, in place of a continuous one, could decrease the flank/crater wear rate on the tool and extend the tool life beneficially.

However, the simulations to date are unable to predict the end point of tool life, as the simulation takes an unreasonable length of time to process. This would be of significant future interest to the researchers

6 ACKNOWLEDGEMENTS

The financial assistance of the Advanced Manufacturing Laboratory (UCT) and the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed, and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF

.

7 REFERENCES

[1] Boeing, *Current Market Outlook: 2017-2036*. 2017.

[2] Brehl, D.E. and T.A. Dow, *Review of Vibration-Assisted Machining methods for precision fabrication*. 2013, North Carolina State University, Raleigh, North Carolina, USA: Precision Engineering Center.

[3] Brehl, D.E. and T.A. Dow, *Review of vibration-assisted machining.* Precision Engineering, 2008. **32**(3): p. 153-172.

[4] Ahmed, N., et al., *Analysis of forces in ultrasonically assisted turning.* Journal of Sound and Vibration, 2007. **308**: p. 845-854.

[5] ANSYS Inc. *ANSYS Homepage*. 2018 [cited 2018; Available from: ansys.com.

[6] Aerospace Specification Metal Inc. *Titanium Ti-6Al-4V (Grade 5), Annealed*. 2018 [cited 2018 13 August]; Available from: http://asm.matweb.com/search/SpecificMaterial.asp?bassnum=mtp641.

[7] Physik Instrumente (PI) GmbH & Co. KG, *Modular Piezo Controller: E501 Series User Manual*. 2018.

[8] Vilakazi, N.S., *An Analysis of a High Speed Discrete Machining System*. 2017, University of Cape Town: Cape Town.

[9] Azushima, A., et al., *Severe plastic deformation (SPD) processes for metals.* CIRP Annals - Manufacturing Technology, 2008. **57**: p. 716–735.

[10] Hu, H.X., Y.G. Zheng, and C.P. Qin, *Comparison of Inconel 625 and Inconel 600 in resistance to cavitation erosion and jet impingement erosion.* Nuclear Engineering and Design, 2010. **240**: p. 2721-2730.

[11] ANSYS Inc *Explicit Dynamics*. 2018.

[12] Shaw, M., *Metal cutting principles*. 1984, Clarendon: Osford University.

[13] Kountanya, R., *Cutting tool temperatures in interrupted cutting—The effect of feed-direction modulation.* Journal of Manufacturing Processes, 2008.

[14] Potdar YK, Z.A., *Measurements and simulations of temperature and deformation fields in transient metal cutting.* Journal of Manufacturing Science and Engineering, 2002.

8 BIOGRAPHY

Kerry-Anne Airey obtained her MSc degree in Mechanical Engineering from the University of Cape Town and is pursuing her PhD in the same.

Lucas Lobmeyer obtained his BSC from and is pursuing his MSc in Mechanical Engineering from the Frankfurt University of Applied Sciences.

Ramesh Kuppuswamy obtained his PhD in Mechanical Engineering from the Nanynag Technological University, in Singapore and was appointed Associate Professor in Advanced Manufacturing at the University of Cape Town.

Axel Hoffmann obtained his PhD in from and was appointed Lecturer at the Frankfurt University of Applied Science.

Nomfundo Vilakazi recently obtained her BSc in Mechanical Engineering from the University of Cape Town

Control of Grinding Induced Surface Cracks through Innovation of a Force Control Device for Vane Blades

R. Kuppuswamy[1], A. Alao[1], K. Sookdawe[1]
Department of Mechanical Engineering, University of Cape Town, South Africa

Abstract
The working conditions of the critical aerospace parts such as: turbine blades, nozzle guide vanes and combustion hardware necessitate refurbishment after every 5000 flight hours as the component experiences severe thermal damage. However, the refurbishment process especially the abrasive assisted process often induces surface damage in the form of cracks and hence multiple reworks has become a mandatory practice This project unveils the findings of a grinding technology by which the conditions of grinding wheel-work interface were configured to avoid the surface damage conditions for the aerospace components. A force control device was developed and tested for grinding titanium based alloys. The device was featured with hydraulic pad base, hydraulic pad, disc spring, plunger and check valves. Pre-defined force was set in the system both through hydraulics and disc springs. When the actual grinding force exceeds the pre-set value, the hydraulic oil would tend to escape through the check valve to facilitate a slip at the wheel-work interface. As a result, the conditions that are favourable to crack initiation and formation were avoided. Prior to design and development of the device, the wheel-work interface was equated to indentation process and the growth behaviour of indentation was studied. The key findings suggest that the force control device has become effective at cylinder pressure of magnitude 9 bars and above, at any pressures below 9 bars has exhibited periodic variation in the grinding forces that lead to chatter and surface damage. The device has enabled to reduce the severe plastic deformation for Ti6Al4V alloy and significantly control the crack growth for hard material weldment fillers such as: Triboloy 800.

Keywords
Grinding, surface damage, force control device.

1 INTRODUCTION

Recent material blending innovations on refurbishment of critical aerospace parts such as: nozzle guide vanes, stator vanes, turbine blades, seal segments, turbine discs, rotor blades and combustion hardware have mitigated the several challenging issues such as: thermal erosion, corrosion, overheating and foreign object damage considerably. In general grinding process is extensively used for refurbishment of aerospace components, but the process often induces surface damage in the form of cracks, residual tensile stress and hence multiple reworks has become a mandatory practice [1,2]. A study on manufacturing process of jet engine turbine blades suggests that fatigue failure, creep, and fracture are initiated by surface cracks which lead to part failure [3]. Table 1 enumerates the major grinding induced surface defects of turbine components. Grinding is still one of the primary operations in the finish machining of critical gas turbine engine components, due to the strict requirement for achieving tight dimensional tolerances of <10 µm and superior surface finish with roughness (R_a) in the order of < 0.5 µm together with acceptable workpiece quality/integrity. Prior research towards avoiding the surface damage has portrayed the use of grinding wheel peripheral speed parameter as it offers high grinding efficiency and surface integrity, especially for brittle materials and thin-fragile components [4]. However, the great potential of this process has not been exploited mostly due to the huge capital outlay associated with ultra-high-speed grinding spindle, drive, lubrication system and thermal management system.

Surface Defect	Description
Cracks	Cracks can be internal or external.
Craters	Craters are shallow depressions in the material.
Laps, folds and seams	Material overlapping during processing.
Metallurgical transformations	Result of thermal cycling of the material.
Pits	Surface depressions that results from physical or chemical attack.
Residual stresses	Compression or tension and are a result of a non-uniform deformation and temperature distribution.
Splatter	Small metal particles get re-solidified onto the surface.
Plastic deformation	Result of friction, die and tool geometry as well as processing methods.

Table 1 - Major grinding induced surface defects of turbine components

Past research efforts also suggest the use of high-speed creep feed grinding (CFG), continuous dress creep feed (CDCF) grinding to preserve the wheel profiles and sharpness of the Al_2O_3 grits [5]. Turbine blades are exposed to working conditions such as "Very High Cycle Fatigue" regime and undergo more than 10 million cycles. An ultrasonic fatigue tests on Ti6Al4V titanium alloy suggests the critical role of surface defects against the fatigue life [6]. A past research has also revealed that brittle fracture is a major drawback in grinding turbine blades and recommends the use of force-controlled grinding technology [7]. Collectively all the past research concludes that a decrease in the actual depth of cut of the individual grain when all other parameters are constant is beneficial to control the grinding induced cracks. Grinding forces, wheel wear and most importantly ground surface roughness are also reduced with a control on the undeformed chip thickness. This project unveils the finding of force control grinding technology by which the conditions of grinding wheel-work interface are configured to avoid the surface damage conditions for the aerospace components.

2 INDENTATION EXPERIMENTS

Prior to design and development of the force control device, the grit-work interface was equated to indentation process and the growth behaviour of indentation was studied. Shown in Fig.1 is the indentation behaviour for various loads on the Ti6Al4V alloy. The indentation test for the Ti6Al4V alloy was performed after polishing with a 800 grit abrasive paper followed by a 1200 grit abrasive paper in the perpendicular direction. Finally, the sample was polished with a 3 µm polishing pad in the same direction as similar to the 800-grit abrasive paper. The indentation results suggest that when the indentation load was 100 N and above the smear with an extensive plastic deformation was apparently seen. The results of indentation test on Ti6Al4V have enabled to configure the force-controlled grinding device for reducing the severe plastic deformation.

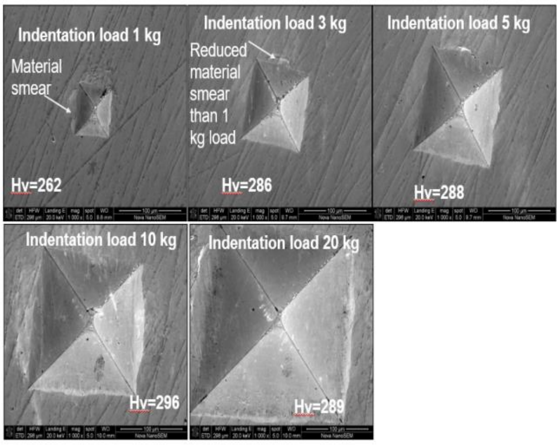


Figure 1 – Indentation test behaviour for Ti6Al4V alloy for various indentation loads

3 DESIGN AND DEVELOPMENT OF A FORCE CONTROL DEVICE

A force control device was developed and tested for grinding the titanium based alloys and hard weldment fillers Triboloy-T800. The device was featured with hydraulic pad base, hydraulic housing, disc spring, plunger and check valves. Pre-defined force was set in the system both through hydraulics and disc springs. Shown in Fig.2 is the developed force control device. When the actual grinding force exceeds the pre-set value, the hydraulic oil would tend to escape through the check valve to facilitate a slip at the wheel-work interface. As a result, the conditions that are favourable to crack initiation and formation were avoided for the titanium based materials.

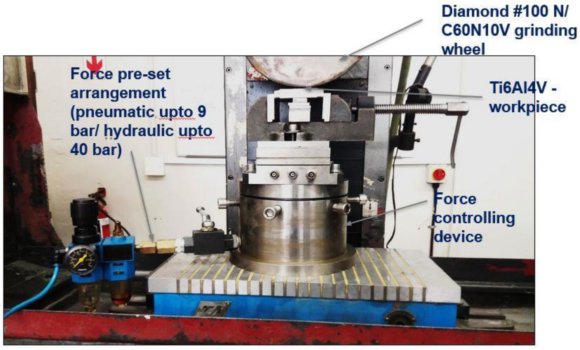


Figure 2 - Development of a force-controlled device for grinding the Ti6Al4V and Triboloy T800

Shown in Fig. 3 is the pre-set value in disc spring and hydraulic cylinder for facilitating the slip at conditions of excessive grinding force. The plunger in the force control device could be adjusted. The hydraulic load was set proportional to the disc from 0-0.3 mm to set a pre-load of value 0-880 N. The hydraulic load was set proportional to the disc spring pre-load value through appropriate setting on the built-in pressure relief valve of pressure range 0-40 bar.

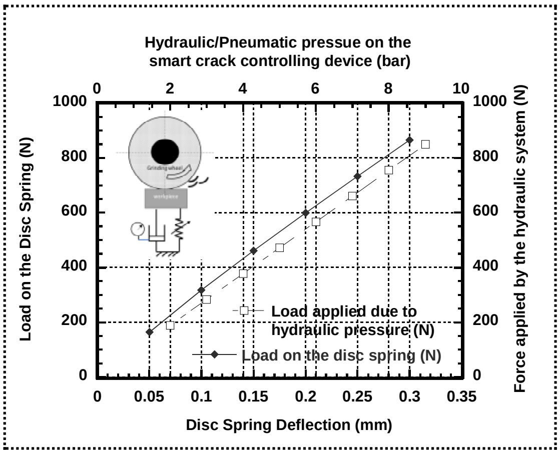


Figure 3- Load characteristics of the force controlled grinding device

4 EXPERIMENTAL SET UP

Grinding experiments were conducted on Ti6Al4V alloy and hard weldment filler:Triboloy-T800. Experiments were conducted on a precision surface grinder PSG-450DX to establish the grinding performance while grinding the Ti6Al4V and Triboloy T800 material. Grinding force signatures F_T and F_N were captured to understand the wheel-work interface behaviour. CBN grinding wheel of mesh #60, #80 and #200 were used for the grinding experiments. The grinding and wheel preparation conditions are given in Table 2 and Table 3 respectively. Surface finish on the work piece was measured using a Taylor Hobson Stylus Profilometer. The captured cutting force signatures were further analysed using the DEWEPORT 7 data acquisition system (see Fig. 4). Data acquisition typically involves the acquisition of signals / waveforms and processing the same to obtain the desired information.

DETAIL	VALUE
Grinding Machine	Precision surface grinder 450DX
Grinding wheel: CBN #60 N 100 EP (250 µm); CBN #80N 100 EP (177 µm); CBN #200 N 100 EP (74 µm)	
Wheel speed	33 m/s
Coolant	Water based (SC 526)
Coolant pressure and Flow rate	8 ~ 12 bar and 20 ~ 25 LPM
Table feed	200 mm/min for Triboloy T800; 15000 ~20000 mm/min for Ti6Al4V
Depth of cut	1.5 mm for Triboloy T800; 6 ~20 µm for Ti6Al4V

Table 2 - Grinding conditions for Ti6Al4V&Triboloy T800

DETAIL	VALUE
Roughing dressing	90 mm/min
Fine dressing	75 mm/min
Very fine dressing	50 mm/min
Dresser rotation speed	110 RPM
Wheel speed	250 RPM
Dresser roller	Diamond disc

Table 3 - Grinding wheel preparation conditions

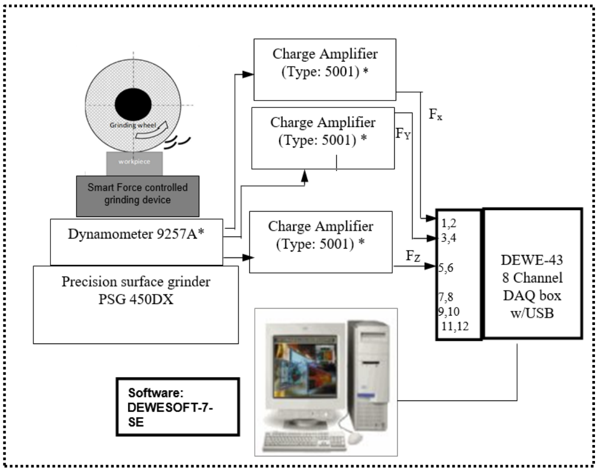


Figure 4 - DAQ arrangement for force-controlled grinding of Ti6Al4V alloy and Triboloy T800

5 RESULTS AND DISCUSSION

The key findings suggest that the force control device has become effective at cylinder pressure of magnitude 9 bars and above. At any pressures below 9 bars has exhibited periodic variation in the grinding forces that lead to chatter and surface damage. Shown in Fig. 5 is the grinding force behaviour for Ti6Al4V alloy and the corresponding pre-set values on the force control device.

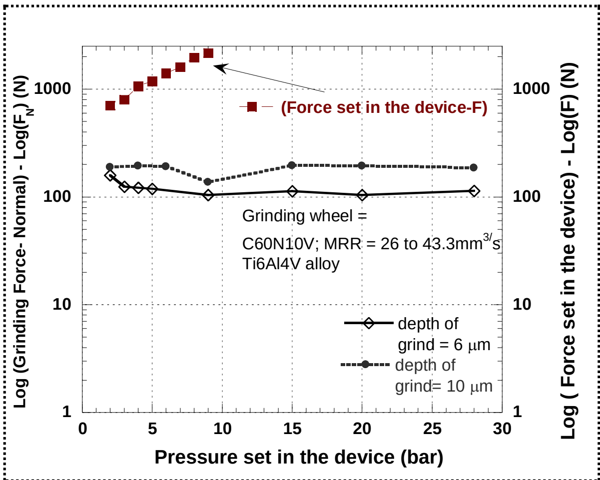


Figure 5 – Grinding force behaviour for Ti6Al4V alloy and pre-set forces in the force control device.

It was also found that at high cylinder pressures of value >9 bar and material removal rate 4000 mm^3/min has produced surface finish (R_a) of value 0.4 ~ 0.6 µm. A clear correlation was observed between the surface finish of the ground parts and grinding forces through surface textures study and material removal plots (see Fig.6). In general, a severe plastic deformation (SPD) in grinding is characterized by the formation of ultra- large plastic strains especially during plowing conditions of the wheel-work interface behaviour. Successful grits and wheel-work interface temperature resize the average grain structures to smaller sizes and the yield strength is given as [8];

$$\sigma_{yield} = \sigma_{friction} + \frac{A}{\sqrt{d}} \qquad (1)$$

where, σ_{yield} is the yield strength, $\sigma_{friction}$ is the friction stress, A is a constant and d is the diameter of the grain.

In the second phase of experiments force control grinding experiments were conducted on material: Triboloy-T800 as it is extensively used as a filler to achieve the final size and shape for the aerospace refurbishing components. Triboloy-T800 offer high strength and hardness of value 1.778 GPa and 498 Hv respectively [9]. The grinding performance results on welding coupons filled with Triboloy-T800 material suggests that the crack control is feasible with using of the developed crack-controlled grinding device. This is confirmed through examination of dye-penetration test and visual observation at magnification X50. The consolidated results are shown in Table 4 and Fig.7. Attempt was made to vary the crack pressure setting in the disc spring and the results are shown in Fig. 7. The results suggest that grinding induced cracks exists at pre-set force in disc spring of magnitude 1300 N than the pre-set force in the disc spring of values 650 and 520 N as shown in Fig.7 A, B & C respectively. Where, σ_{yield} is the yield strength, $\sigma_{friction}$ is the friction stress, *A* is a constant and *d* is the diameter of the grain.

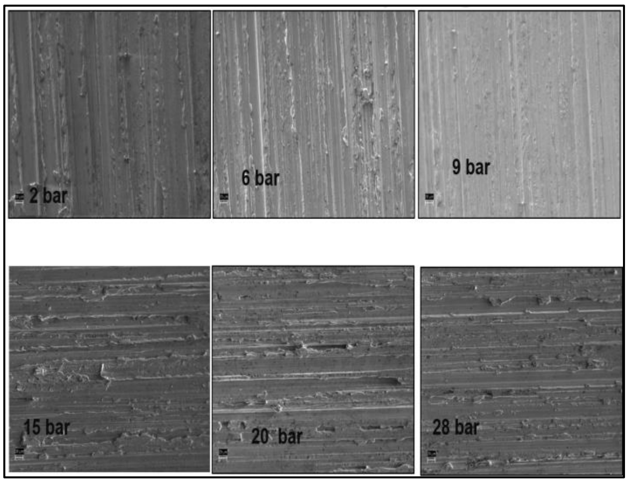


Figure 6 – Surface texture behaviour for Ti6Al4V against the different settings of hydraulic pressures

Material removal rate (mm^3/s)	Grinding force normal (F_N)	Grinding force Tangential (F_T)	Presence of crack (using dye penetration test) (X50)	
			Without crack controlling device	With crack controlling device
21.8	153	385	No	No
26.1	180	420	Yes	No
17.71	135	340	No	No
23.9	162	405	No	No
13.5	112	280	No	No
33.5	225	490	Yes	No

Table 4 - Grinding experiments results on welding coupons filled with Triboloy T800 material at grinding conditions: grinding wheel CBN #200, Creep feed grinder, wheel speed 33 m/s, feed rate 200 mm/min and depth of grind 1.5 mm.

The results were further confirmed through an extensive theoretical analysis using the Griffiths fracture mechanics applied to brittle materials. The details are given below;

In general, the Griffith's theory enables to derive the crack initiation conditions for brittle materials through equating the internal elastic energy to the surface energy of the intended workpiece. The internal elastic energy per unit thickness of the workpiece with a crack (W_1) is given as (See Fig.8) [10];

$$W_1 = (1 - \upsilon^2)\frac{\pi c^2 \sigma^2}{E} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

where, σ is a uniform tensile stress that produces a crack of length *2c*, *E* is the longitudinal elastic modulus (778 GPa for Triboloy) and υ is the Poisson's ratio (0.33).

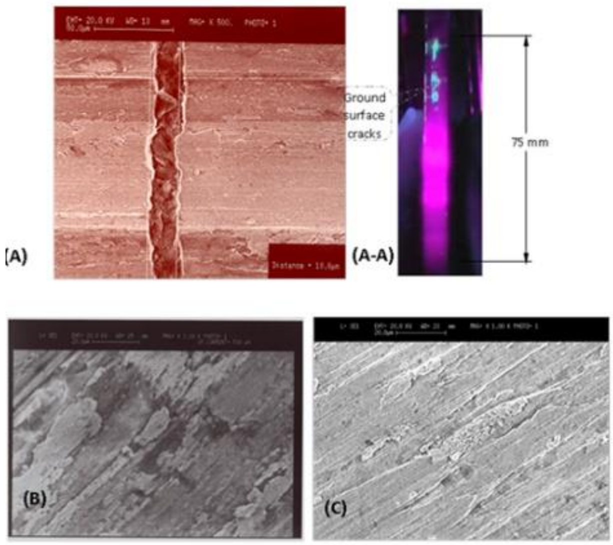


Figure 7 - Ground surface behaviour for Triboloy T800 at settings: A) Pre-set force in the disc spring is 1300N; B) Pre-set force in the disc spring is 650 N; C) Pre-set force in the disc spring is 520 N

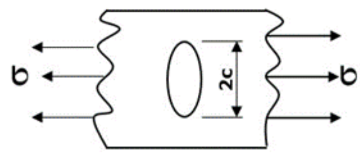


Figure 8 – Griffith's crack growth model of size 2c against the external working stress (σ) due to external loading

The surface energy per unit thickness of the workpiece resists the crack formation and given as;

$$W_2 = 4.\gamma c \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3)$$

where, γ is the surface energy density (1770 mJ/m^2) of the work piece (Triboloy T800).

At conditions of equating the surface energy and internal elastic energy, the critical crack size c_c and critical crack stress σ_c, are computed as;

$$C_c = \frac{4\gamma E}{2\pi\sigma^2} \text{ .. (3); } \sigma_c = \sqrt{\frac{2\gamma E}{\pi C(1-\vartheta^2)}} \text{...... (4)}$$

While applying the Griffith's theory to the grinding process, the crack length could be equated to the scratch length or grit contact length (l_c) that facilitates chip formation (See Fig.9).

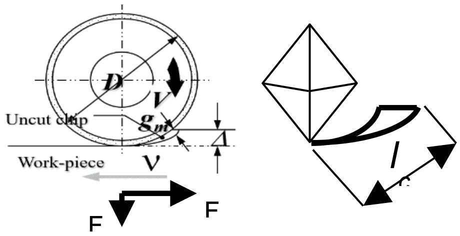


Figure 9 – Grinding wheel-work interface behaviour and grit contact length, l_c

The undeformed chip thickness of a grinding process is useful to characterize the process and given as [11];

$$g_m = \sqrt{\frac{4v}{VCr}\sqrt{\frac{\Delta}{D}}} \text{................. (5)}$$

where, Δ is depth of grind (Δ = 0.25 ~1.5 mm); V is the cutting speed (V=33 m/s); v is the feed rate (v =200 mm/min); C is the number of abrasives /mm^2 (C= 20, 40 and 232 for the grinding wheel #60, #80 and #200 respectively and r = ratio of width to grit protrusion depth

The undeformed chip thickness was used to compute the scratch length as;

$$2c = l_c = \sqrt{D.g_m} \text{.................. (6)}$$

Attempts were made to compute the critical stress, (α_C) that produces critical crack size (C_c) and the same are compared with the applied stress. The applied stress was computed using the grinding force/grit value and the contact area at the grit-work interface. To compute the applied stress the following grinding force equations were used. They are [12];

$$F_n = C_p \frac{\pi.v.\Delta.b}{2V} Tan\alpha \text{ ... (7) } \quad F_t = C_p \frac{v.\Delta.b}{V} + \mu F_n \text{ ... (8)}$$

where , C_p = 1.778 GPa for Triboloy T800;Δ= depth of grind (0.25 ~1.5 mm); b = grinding width (welding coupon) (6 mm); V =cutting speed (33 m/s); v= feed rate (200 mm/min); α=half included angle of the abrasive (°) Tan(80°)= 5.8 ; C = no of abrasives /mm^2;r = ratio of width to grit protrusion depth.

Using the equations 7 and 8 and applying the wheel work interface conditions the applied stress was computed and shown in Table 5. The computed critical stress results were compared with the applied stress and the results are shown in Fig. 10

Depth of cut , d	contact length (mm)	contact area (mm^2)	Grinding force (F_N) N	Ft (N)
			$F_n = C_p \frac{\pi v \Delta b}{2V} Tan\alpha$	$F_t = C_p \frac{v \Delta b}{V} + \mu F_n$
0.25	6.8	40.8	144.5	59.3
0.5	9.5	57	288.9	118.4
0.75	11.7	70.2	433.3	177.6
1	13.5	81	577.7	236.8
1.25	15	90	722.1	296
1.5	16.5	99	866.5	355.1

No of grits in contact	Force/ grit (N), FN/grit	No of grits in contact	Force/ grit (N)	No of grits in contact	Force/ grit (N)	Applies stress (MPa)		
#60	#60	#80	#80	#200	#200	#60	#80	#200
816	0.177083	1696	0.0852	8484	0.017032	140.9899	129.0554	150.6729
1140	0.253421	2400	0.120375	12000	0.024075	201.7684	182.3353	212.9777
1404	0.308618	2940	0.147381	14700	0.029476	245.7152	223.2419	260.7589
1620	0.356605	3392	0.170313	16960	0.034063	283.9211	257.9769	301.3314
1800	0.401167	3792	0.190427	18960	0.038085	319.4002	288.4452	336.9201
1980	0.437626	4156	0.208494	20784	0.041691	348.4286	315.8111	368.8139

Table 5 - Computed results of applied stress at the wheel-work interface for Triboloy T800

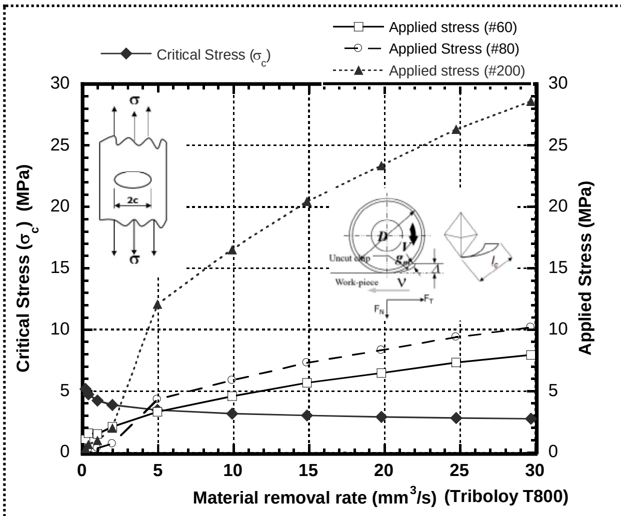


Figure 10 - Behaviour of applied stress at the wheel work interface for TriboloyT800 against the critical stress computed using the Griffith's theory

The computed results in Fig.10 suggests that when the applied stress is less than the critical stress the process of crack initiation could be avoided.

6 CONCLUSION

The main aim of the project is to design and develop a force control device for controlling crack formation as well as reducing the severe plastic deformation while grinding the blended refurbished aerospace components and OEM components respectively. The device was found to be effective at a cylinder pressure 9 bar and above. At any pressure below surface texture variation was widely observed. The required set force in the device was found to be much higher than the actual grinding force in order to avoid the grinding induced surface defects for the turbine components. Also, the achievement of grinding finish and surface texture suggests the suitability of device for further profile grinding that employs higher material removal rate and generates higher grinding force. This is evident from the grinding tests conducted on welding coupons. Furthermore, this project has enabled to reduce the cost through reduction of grinding related scrap with improved grinding productivity.

7 ACKNOWLEDGEMENTS

The authors wish to thank Emeritus Prof. Dimitrov, Stellenbosch University, South Africa for the foresight, market steering and project management towards the formation of the Titanium Centre of Competence at South Africa and the related project support. Further thanks go to: Mrs Penny Louw for her assistance with the optical microscope and her help with surface roughness testing; Mrs Miranda Waldron for her assistance with the scanning electron microscope; and Mr Horst Emrich for his assistance on the precision surface grinder. This project was supported through a NRF grant for the incubation of Titanium Centre of Competence, South Africa vides reference: TICOC/UCT 401582 on project: Resource Efficient Process Chains for Titanium Components in Aerospace, Automotive, and Medical Applications.

8 REFERENCES

[1] L. Nelson, H. H. K. Xu, S. Danyluk and S. Jahanmir, 1997, Subsurface damage in grinding titanium aluminide, Machining Science and Technology, 1/2: 289-297.

[2] D. Ulutan., T. Ozel., 2010, Machining induced surface integrity in titanium and nickel alloys: A review, International Journal of Machine Tools & Manufacture, 51:250-280.

[3] T A. Thong., 2013-2014,The manufacturing process of jet turbine blades," CA report, MA2700, School of mechanical and aerospace engineering, Nanyang Technology University.

[4] Ramesh, K., Huang, H., 2006, Use of wheel speed as a parameter to inhibit surface crack generation in the grinding of wear-resistant fillers, The International Journal of Advanced Manufacturing Technology, 28/7: 701-706.

[5] J. Beddoes.,W. Wallace., M. d. Malherbe.,1992, The technology of titanium aluminides for aerospace applications, materials and manufacturing Processes, 7: 527-559.

[6] S. Heinz., D. Eifler.,2016, Crack initiation of Ti6Al4V in the very high cycle fatigue regime, International Journal of Fatigue, 93: 301-308.

[7] T. R. Kurfess., H. A. Razavi., S. Danyluk.,2003, Force control grinding of Gamma Titanium Aluminide, International Journal of Machine Tools & Manufacture, 2/43 : 185-191.

[8] Azushima, A., Kopp, R., Korhonen, A., Yang, D. Y., Micari, F., Lahoti, G. D., Groche, P., Yanagimoto, J., Tsuji, N., Rosochowskij, A. & Yanagida, A., 2008, Severe plastic deformation (SPD) processes for metals, CIRP Annals - Manufacturing Technology, 57:716–735.

[9] Data sheet supplied by Rolled alloys against bulletin number :280USe07/12 with web reference: https:// www.rolledalloys.com/shared-content/technical-resources/datasheets/ TRIBALOY-T-800_DS_US_EN.pdf

[10] Norio Taniguchi, 1996, Book: Nano-technology – Integrated processing systems for ultra – precision and ultra fine products, Oxford University Press, USA

[11] Kalpakjian. S., 2016, Book: Manufacturing processes for engineering materials- 6th edition, Pearson publishers.

[12] Inada Y, Studies in ultra-high-speed grinding, PhD thesis, Tohoku University, 1996, Japan

9 BIOGRAPHY

Ramesh kuppuswamy obtained his PhD degree in Mechanical Enginnering at the Nanyang Technological University, Singapore. Since 2009 he was appointed as a Associate Professor in the Department fo Mechanical Engiineering, University of CapeTown, South Africa.

Alice Alao has obtained her BSC, MSc degree programmes at Federal University of Technology akure-Nigeria. Presently she is doing her PhD studies in the Department of Mechanical Engineering, University of CapeTown, South Africa.

Kitesh Sookdewe obtained his BSc in Mechanical engineering from the University of Cape Town in 2017. Presently he is a Engineer in training at Sappi Kraft-Tugela Mill, Durban , South Africa

Stream B

Novel Clamping System for Machine Tools

E. Relea[1], L. Weiss[1], M. Schlüssel[2], K. Wegener[3]
[1]Inspire AG, Switzerland
[2]Gressel AG, Switzerland
[3]Institute of Machine Tools and Manufacturing (IWF), ETH Zurich, Switzerland

Abstract
The need for higher productivity and efficiency in manufacturing of complex parts makes 5-axis machining centers preferred to a machine tool in industrial production. Along with the machine capability, 5-side accessibility of the workpiece, as well as safe and repeatable clamping is required. However, conventional clamping-technology by means of a vise gives full access on 3 sides of the workpiece only. Furthermore, it has disadvantages such as workpiece elastic and occasionally plastic deformation due to the high clamping forces, and, custom or additional chucks are required for clamping specific shapes. Alternative clamping techniques, for instance magnetic or vacuum, have a restricted use due to the requirement of ferromagnetic materials or shape constraints, respectively. In this work, a novel way of clamping workpieces was subject to research. Instead of clamping the workpiece on its surface, like conventionally carried out, pins were welded onto the workpiece, which were then used for clamping. Advantages of this innovative clamping method are amongst others the unlimited access on 5 sides of the workpiece, a lower workpiece deformation compared to traditional clamping, and clamping of workpieces with irregular surfaces. Furthermore, savings in raw material and consequently reduced power consumption, as well as wider possibilities to automate the process are additional advantages of such a clamping system. The research work focuses on the feasibility and quality issues of the stud welding process and its impact on the workpiece, the positioning and the number of studs, and qualitative and quantitative comparisons of conventional clamping with the novel pin clamping method.

Keywords
Clamping system, milling, stud welding, clamping automation

1 INTRODUCTION AND STATE OF THE ART

Strong global competition, faster paced developments, rapidly expanding options in materials as well as processes, increasing requirements for precision and quality push the machine tool industry to seek improvements in every single part of machining centers [1]. This also includes the clamping technology and processes for the workpiece, which, according to Stefan et al. [2], can be classified as visible in Figure 1:

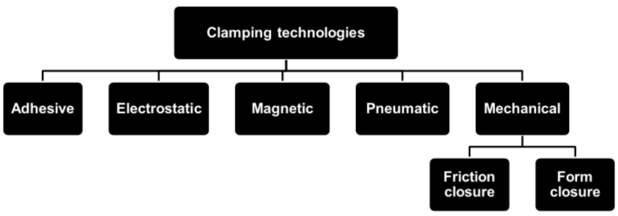


Figure 1 - Common clamping technologies according to [2].

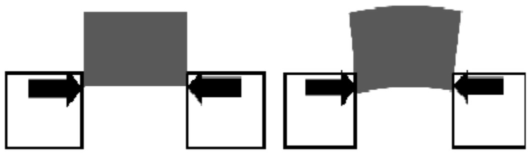
An example of adhesive clamping is the freezing of the workpiece on a cooled fixing surface. This method is suitable and commonly used for clamping materials such as ceramic, rubber or graphite due to the lack of clamping forces acting on the workpiece. However, during machining, the development of heat significantly reduces the holding forces, which makes this technique unsuitable for machining with high removal rates [3].

Electrostatic clamping requires the workpiece to be electrostatically charged and sets low limits regarding the workpiece weight [4]. Also, it must have a large contact surface in comparison to its height. This technique is mainly used for clamping semiconductor wafers, and is regarded unsuitable for machining purposes [5].

Magnetic clamping is another tension-free clamping technique. Besides the main limitation of the workpiece having to be magnetic, it also requires a clean contact surface, since even small air gaps affect the holding forces considerably. An additional disadvantage is the chip adhesion, which continues to take place due to residual magnetism even after the unclamping of the workpiece [4].

Pneumatic clamping technology uses a suction plate under the workpiece with the purpose of creating a vacuum; thereby the desired holding force can be generated. This provides optimal accessibility to the workpiece because the clamping device does not protrude. In addition, very thin and flat workpieces can also be securely clamped. However, porous, uneven workpieces cannot be clamped pneumatically, since the holding force is reduced too much by excessive air leakage [6].

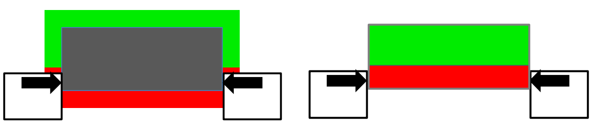
The mechanical clamping by means of a vise is the most used clamping technology in machining due to its universal applicability and adaptability to almost any workpiece geometry [7]. However, friction clamping has the disadvantage that its effectiveness is reduced by the friction coefficient. With a coefficient of friction of 0.1 between typically steel chucks and steel workpieces, it means that the force applied by the vise onto the workpiece surface must be 10 times greater than the holding one.



(a) (b)

Figure 2 - Bent workpiece once clamped (b) compared to its initial unclamped state (a)

This has two major disadvantages: firstly, the high clamping forces elastically and/or plastically deform the workpiece, as portrayed in Figure 2, with negative consequences regarding the manufacturing precision of the part. Secondly, the clamping devices, e.g. vise, chuck, collet, must be designed for high forces, with disadvantages in term of production costs and limited handling capabilities for automated processes caused by the excessive weight. Due to increasing personnel rates and lower automation costs in recent years, a surge in degree of production automation took place [8]. 5-axis machining requires clamping devices that improve the accessibility to the workpiece, which is highly obstructed by machine vises, as depicted in Figure 3a. Although clamping devices are designed to limit the accessibility to the workpiece as little as possible, all frictional methods impede at least 3 surfaces to be machined. This problem can be avoided by allowing extra material, visible in red in Figure 3b, which is removed in a subsequent processing step. However, this results in a systematic loss of material.



(a) (b)

Figure 3 - Side view: accessibility issues in a vise (a), and consequent material loss (b).

Also, another disadvantage of vises is the travel distance of the clamping device, which must be large enough to be able to clamp all workpiece sizes. On top of that, special chucks are needed for non-rectangular workpiece shapes.

A promising approach to achieve the above mentioned characteristics is the addition of form closure to the force closure [10]. By the means of an embossing station, a patented pattern is embossed into the workpiece, which is then used to form clamp the workpiece into a vise with custom chucks, as depicted in Figure 4. This allows for increased holding forces and reduced clamping forces. Hence, deformations due to clamping stresses are lowered significantly. Furthermore, the interface area can be reduced, which increases the accessibility to the workpiece in 5-axis machining.

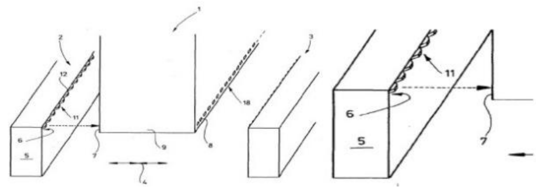


Figure 4 - Lang GmbH's „Präge-Fix" [11].

The matrix form clamping technology suggested by Matrix GmbH [9] is yet another option. As shown in Figure 5, the workpiece is clamped with pins instead of chucks. The former adapt individually to the workpiece geometry, which gives benefits in terms of flexibility and more uniform clamping pressures distributed on the workpiece.

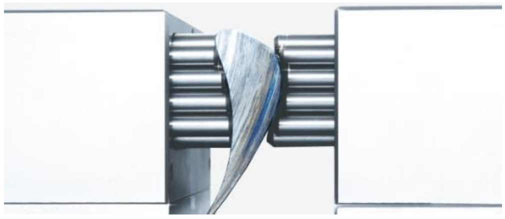


Figure 5 - Matrix GmbH clamping system [9].

A main disadvantage of this system is the susceptibility to contamination caused by chips and coolants. On top of that, the accessibility to the workpiece is made more difficult in the matrix form clamping system caused by the clamping pins in comparison to the workpiece size.

2 NOVEL CLAMPING SYSTEM

The aim of this work is to propose a method and suitable clamping technique which simultaneously meets the following requirements:

- Good ratio of available strength to clamping force.
- Minimal deformation of the workpiece.
- Maximum accessibility of the workpiece.
- Minimal material loss.
- Independence from the workpiece contour.
- Suitability for automated loading and unloading without a grid plate.
- Reproducible re-clamping within tight limits in case re-clamping has to take place in another machine between roughing and finishing.
- Low cost of the clamping device.

- Low costs for clamping a workpiece and the operation of the clamping device.

These requirements are met by a novel clamping method by welding studs on one of the workpiece's surfaces in such a way that the workpiece can be easily and reliably clamped on the studs themselves, as shown in Figure 6.

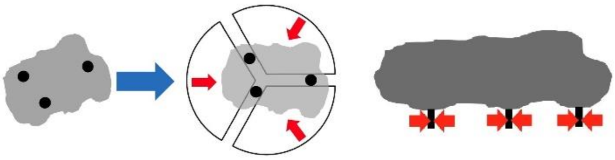


Figure 6 - Novel pin-clamping. Workpiece visible from above (left) and the side (right).

3 FEASIBILITY AND QUALITY ISSUES OF THE STUD WELDING

Stud welding is a well-established technology, developed for a long time in different fields. Reliable and sturdy bonding between stud and parent material has been a primary focus in all applications. Precision and accuracy of the fastening, for both manual and automatic stud welding, is typically in the tenth of millimeter range [12]. Usual application of stud welding include shipbuilding and building construction, automobile bodies, as well as electrical panels. The different stud welding processes are listed below.

The capacitor discharge (CD) process utilizes a flanged fastener with a timing tip at its center. There are two different methods of fastening the studs. The first one involves a direct contact between the ignition tip and the workpiece. The former is pressed by a spring in the welding gun onto the latter. Once the welding process is triggered, the current evaporates the tip and ignites the arc, after which the two are joined. The welding time is typically shorter than 4 ms. The second CD method holds the stud at a predefined distance from the workpiece. Once the welding process is triggered, the fastener is accelerated by a spring towards the workpiece. The welding time doesn't normally exceed 1 ms.

Drawn arc (DA) welding performs a lifting action of the stud to initiate an arc which produces a molten pool in the material. The stud is consequently fastened into the pool by the return spring pressure of the welding gun. Mild and stainless steels, as well as some aluminum alloys may be welded using this process, which is more tolerant of surface irregularities such as rust, oil and dirt. The weld duration can vary between 100 and 1000 ms.

A mix between the two methods described above is the Short cycle (SC) welding. The welding sequence is as with the DA welding, except for the use of higher currents and shorter welding times. The latter is comprised between 10 and 100 ms. In order to avoid weld spatter, shielding gas is usually employed.

In established application fields for stud welding, the mixture of materials for studs and parts tends to remain focused and limited. For clamping technology purposes, stud welding needs to cover a much larger spectrum of material combination, mainly of steel alloys, but also of aluminum. In order to simplify and speed up the process, it was necessary to find a welding process and stud type that could cover a wide range of materials. The stud welding tests included a total of 9 commonly machined steel alloys (C45E, 14NiCr14, 16MnCr5, S235JR, 11SMnPb30, 9SMn28, X5CrNi18-10, 16MnCrS5 and 42CrMoS4), as well as 3 aluminum alloys (AlMgSi0.5, AlMg3, AlMgSi1). The stud materials varied between S235JR and 1.4301 steel, and AlMg3 aluminum. Two diameter variations, Ø6 and Ø8 mm, were also considered upon economic considerations, as well as minimum stiffness requirements based on FEM simulations. To mention is also the fact that the surfaces of the workpieces were not cleaned, as to represent real life conditions, so rust, grease and cutting fluid were present on the surfaces, as visible in Figure 7.

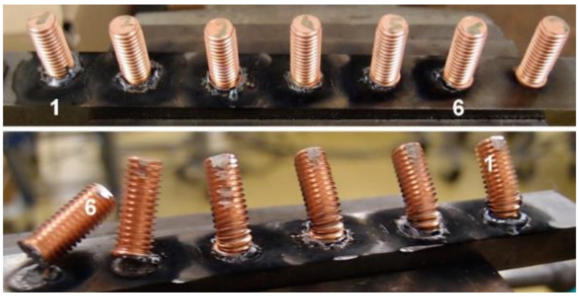


Figure 7 - Variation of stud welding parameters to achieve weldability standards.

The fastening of the studs was carried out according to DIN EN ISO 14555, which specifies both a visual examination, visible in Figure 8, and a 60°bending test of the studs, depicted in Figure 9.

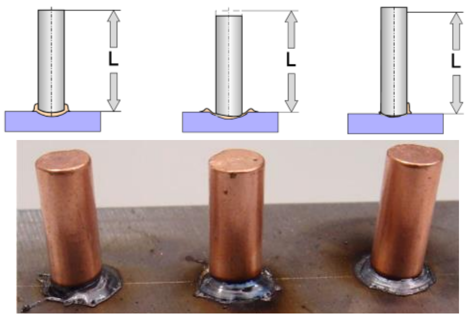


Figure 8 - Visual examination according to DIN EN ISO 14555 [13].

With regards to the steel alloys tests, the Ø6 mm fasteners welding results showed a compatibility of all 8 steel alloys to be joined with both S235 JR and 1.4301 studs utilizing the drawn arc method complemented by shrouding gas, visible in Figure 10. However, the Ø8 mm studs were exclusively weldable with the 8 steel alloys if made of S235 JR.

Figure 9 - 60° bending test according to DIN EN ISO 14555 [14].

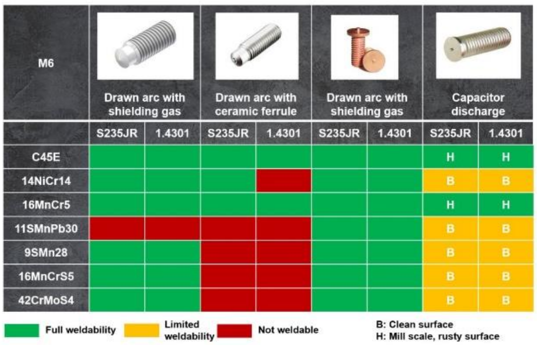


Figure 10 - Weldability tests for steel with M6 studs.

On the other hand, AlMg3 studs proved to be more challenging to fasten, and could exclusively be welded in the 6 mm diameter with the 3 aluminum alloys, utilizing the capacitor discharge with gap technology.

4 IMPACT OF THE WELDING PROCESS ON THE WORKPIECE

The depth of welding into the parent material was visually investigated by milling away layers of 0.25 mm at the time, and stopping where no observable effects could be detected anymore, as shown in Figure 11.

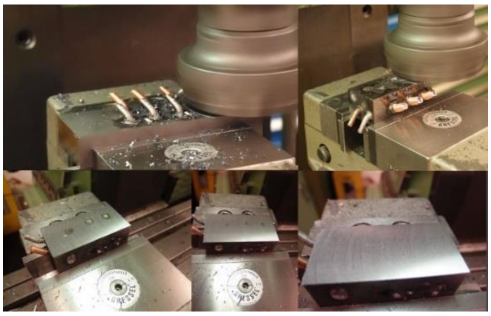


Figure 11 - Milled surfaces to investigate the welding penetration in the parent material.

In the case of drawn arc welding for steel alloys the welding depth could extend to one quarter of the stud diameter. Capacitive discharge with gap technology for aluminum alloys and 6 mm diameter fasteners reached a welding depth of up to 1 mm in the 3 aluminum alloys. The results can be seen in Table 1.

Workpiece Material Range	Stud Materials	Stud Ø [mm]	Welding depth [mm]
9 steel grades	S 235 JR 1.4301	6	1.5
9 steel grades	S 235 JR 1.4301	8	2
3 aluminum grades	AlMg3	6	1

Table 1 – Stud welding depths in the workpieces.

5 STUD POSITIONING OPTIMIZATION

The optimization of the positioning of the studs with regard to stiffness was carried out on an exemplary basis. Two steel workpieces were chosen: a plate with dimensions of 80x80x20 mm and a cube with dimensions of 80x80x80 mm. The studs, Ø6 and Ø8 mm, were fastened in a square arrangement, and clamped at two heights from the workpiece's bottom: 5 and 10 mm. The gap between studs was also varied between 48 mm and 68 mm. The studs were moved on the diagonal of the chucks with a fixed displacement of 1 µm, and the vertical and side displacements of the workpieces were calculated, as illustrated in Figure 12.

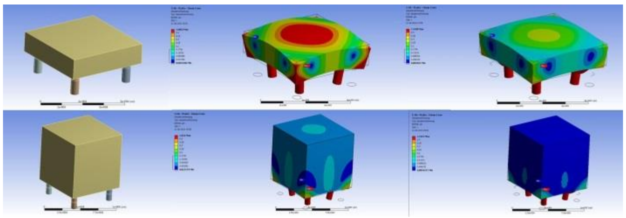


Figure 12 - FEM displacements.

As represented in the Tables 2 and 3, the stud diameter, the clamping height, as well as the stud gap are important parameters which can be varied to achieve minimum displacements in the workpiece and stiffest possible clamping.

Plate workpiece LxWxD= 80x80x20 mm				
Stud diam. [mm]	Clamp height [mm]	Stud gap [mm]	Vertical displacement [μm]	Side displacement [μm]
6	5	68	0.610	0.320
6	5	48	0.472	0.155
6	10	68	0.260	0.132
6	10	48	0.223	0.075
8	5	68	0.920	0.480
8	5	48	0.850	0.276
8	10	68	0.510	0.257
8	10	48	0.440	0.140

Table 2 – Displacements of the plate due to clamping forces.

Cube workpiece LxWxD= 80x80x80 mm				
Stud diam. [mm]	Clamp height [mm]	Stud gap [mm]	Vertical displacement [μm]	Side displacement [μm]
6	5	68	0.106	0.154
6	5	48	0.011	0.064
6	10	68	0.041	0.055
6	10	48	0.004	0.024
8	5	68	0.170	0.253
8	5	48	0.021	0.104
8	10	68	0.086	0.113
8	10	48	0.007	0.044

Table 3 – Displacements of the cube due to clamping forces.

6 COMPARISONS BETWEEN CONVENTIONAL AND THE PIN CLAMPING METHODS

FEM comparisons, as visible in Figure 13, were carried out to investigate the workpiece deformation caused by the clamping forces for the two different methods. Steel plates (LxWxD= 80x80x20 mm), as well as cubes (LxWxD= 80x80x80 mm) were taken as examples of workpieces. The stud arrangement was square with a gap of 60 mm between fasteners, while the clamping height was 5 mm. Considering typical clamping forces of 8 kN, the vertical deformations were as follows:

F=8 kN	Vertical deformation [μm]	
	Conventional clamping	Pin clamping
Plate	4.7	2.96
Cube	2.3	0.9

Table 4 – Comparison of vertical deformation

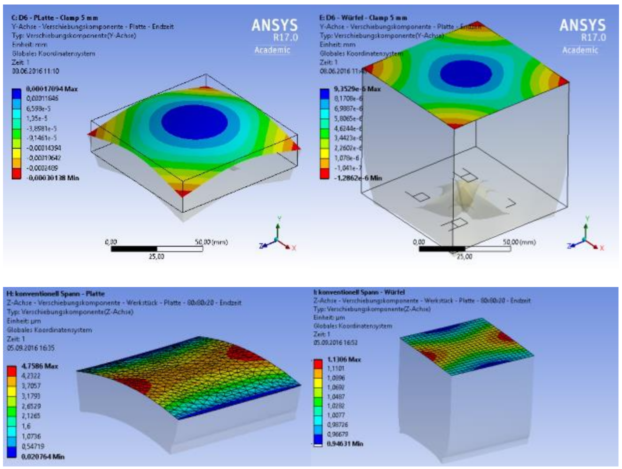
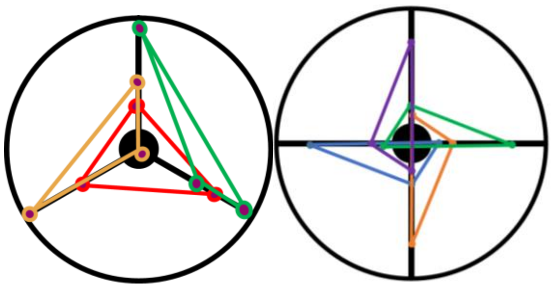


Figure 13 - Pin (top) vs. conventional (bottom) clamping FEM.

7 ASSESSMENT OF THE STUD NUMBER

The minimum number of studs, in order to get a defined system, is three. A comparison with four studs was carried out. The studs have a considerable cost since they are extra material that has to be purchased, welded, and finally removed, contributing to the material loss and adding further process steps.



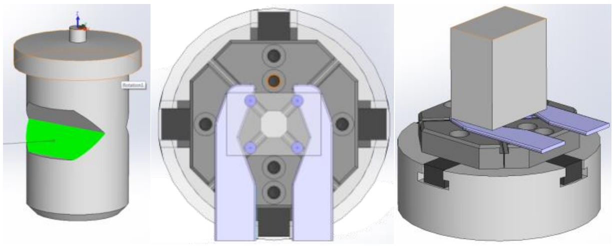
(a) (b)

Figure 14 - 3 vs. 4 jaw chucks.

The main reasons a four stud configuration is advantageous over a three one are as follows:

- Flexibility and ability to cover a wider range of workpiece shapes. As visible in Figure 14a, a three jaw chuck can exclusively clamp three stud configurations with a maximum spacing between studs lower than the whole diameter of the chuck. Besides, as shown by the orange triangle in Figure 14a, arrangements at 120° are also not possible. Also, when an angle is greater than 120°, two bolts of the side must be clamped in one jaw, as depicted by the green triangle in Figure 14a. On the other hand, a four jaw chuck has neither of the above mentioned limitations, and, on top of that, in case of need, triangular configurations can also be clamped in a four jaw chuck, as displayed in Figure 14b.

- Ease of automation –the studs can also be used for handling the workpiece. Compared to the workpiece, which can have different shapes, fasteners maintain standard geometries, as illustrated in Figure 15a. Besides, the studs' grooves created by the clamping can be advantageously used as a form closure, shown in Figure 15b and 15c.
- Safety – in case one of the fasteners fails, there are still three to securely withstand the load.



(a) (b) (c)

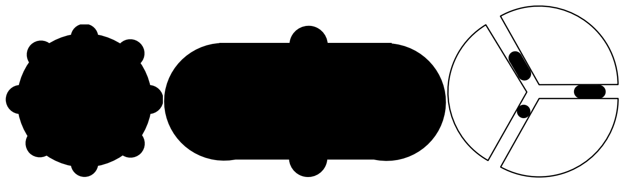
Figure 15 - Automatic handling of the workpiece [15].

8 CONCLUSION

In this paper a novel clamping pin system was characterized. The main advantage of this method is a lower workpiece deformation compared to traditional clamping, as the clamping forces deform the pins, not the workpiece. Moreover, this creates an additional form closure, rather than a purely friction one. 5-side accessibility for machining is also made possible by clamping on the pins, which are welded on one side of the workpiece only. Additionally, the clamping system is much more compact, which further increases accessibility to the workpiece in 5-axes machining. Another benefit of pin clamping is that a much broader workpiece geometry choice can be clamped. Furthermore, savings in raw material and consequently reduced power consumption, as well as wide possibilities to automate the process are additional advantages.

9 OUTLOOK

As an alternative to round studs, other contoured shapes are possible which increase stiffness and provide a feature that facilitates the form closure by plastic deformation, as illustrated in Figure 16a. Also possible are other basic fastener forms, which take into account the alignment of the jaws, shown in Figure 16b and 16c.



(a) (b) (c)

Figure 16 - Alternative stud contours and shapes

Cast parts and additive manufacturing could also be of interest, as the pins would be directly integrated into the workpiece, which would allow for a much wider freedom of design.

10 ACKNOLEDGEMENTS

This work was supported by the Swiss Commission for Technology and Innovation (CTI Project 18438.1 PFIW-IW) and GRESSEL Spanntechnik AG.

11 REFERENCES

[1] Nägeli, D., 2016, Industrie 4.0 nimmt Form an, in Unternehmer Zeitung, ed. Zürich: SWISS BUSINESSPRESS.

[2] Stefan, H., Heinrich, K., Dieter, E., 2012, Vorrichtungen, Begriffe und Gliederung, in Betriebsmittel Vorrichtung, ed: Carl Hanser Verlag GmbH & Co. KG, pp. 1-16.

[3] Blumenthal P., Raatz A., 2012, Adhesive Workpiece Fixturing for Micromachining, IFIP Advances in Information and Communication Technology, vol 371. Springer, Berlin, Heidelberg.

[4] VDI-Gesellschaft Produktionstechnik, 1992, Vorrichtungen: rationelle Planung und Konstruktion: Springer.

[5] Horwitz, C.-M., 1992, Elektrostatische spannvorrichtung, DE4216218A1.

[6] Wulfsberg, J.-P., Lehmann, J., 2003, Spanntechnik für die Mikrofertigung. In: wt Werkstattstechnik online 93(3) p. 146.

[7] Henriksen, E.-K., 1973, Jig and fixture design manual. Industrial Press, New York.

[8] Vogel business Media AG., 2011, Spannen ohne Deformationen, maschine+werkzeug, 62-64.

[9] Matrix GmbH website, 2018, http://www.matrixinnovations.com/uploads/productcategories/2/slides/3.jpg.

[10] LangTechnik, 2013, Präge-Fix Formschluss-Technologie / Makro-Grip Zentrierspanner.

[11] Lang Technik GmbH website, 2018, Prägemuster: http://www.langtechnik.com

[12] Trillmich R., Welz W., 2014, Bolzenschweissen Grundlagen und Anwendung, DVS Media GmbH, Düsseldorf.

[13] DIN EN ISO 14555-08, 2014, Schweissen, Lichtbogenschweissen von metallischen Werkstoffen, DIN, Berlin.

[14] DIN 54111-1, 1988, Zerstörungsfreie Prüfung, DIN Berlin.

[15] Gressel AG, 2018, http://www.gressel.ch/

12 BIOGRAPHY

Eduard Relea obtained his M.Sc.Eng. degree in Mechanical Engineering from the Polytechnic University of Turin. His specific focus and area of research interest at inspire AG is in the composite materials and machine tool structures.

Lukas Weiss holds M.Sc.Eng. degree in Electrical Engineering from the ETH Zurich. He is currently Head of the Machine Group at inspire AG, and his research interest is in the Technology and Innovation Management.

Marcel Schlüssel obtained his Mechanical Engineering degree from the ZHAW School of Engineering. He is currently Head of Technique at GRESSEL Spanntechnik AG, based in Aadorf, Switzerland.

Konrad Wegener holds M.Sc.Eng. and PhD degrees in Mechanical Engineering from the TU Braunschweig. In 2003 he was appointed Professor and Head of the Institute for Machine Tools and Manufacturing (IWF) at the ETH Zurich, Switzerland.

The Design of a 5 Degree of Freedom Parallel Kinematic Manipulator for Machining Applications

W. Dharmalingum[1], J. Padayachee[1], G. Bright[1]
[1]Mechatronics and Robotics Research Group, Department of Mechanical Engineering, University of KwaZulu-Natal, South Africa

Abstract
Robotic machining is a disruptive manufacturing alternative to traditional computer numerically controlled machines. Robotic systems for machining applications have not been fully adopted in industry due mainly to accuracy issues, throughput and rigidity, when compared to computer numerically controlled machines. In comparison to serial robotic systems, parallel kinematic manipulators possess a higher payload to weight ratio, better stiffness and higher accuracy. The aim of this research was to produce a novel parallel kinematic machine that was compact, lightweight, cost-effective and capable of 5 degree of freedom machining. The cost-effective nature of the parallel kinematic manipulator enables small to medium size companies to enter the market. The contribution of the study was to research and develop a new 5 degree of freedom parallel kinematic manipulator. The study presented forward and inverse kinematics including experiments to confirm the validity of the potential to use this robotic platform for computer numerically controlled tasks. The design of a parallel kinematic manipulator was highly influenced by the number of degrees of freedom desired which directly impacted the type of joints used and the complexity of the design of the legs of the machine.

Keywords
Robotic machining, parallel kinematic manipulator, inverse kinematics, forward kinematics

1 INTRODUCTION

The concept of robotic machining has been researched worldwide since the early 1990's. Investigations have been conducted on machining tasks that were once only believed to be capable by computer numerically controlled (CNC) machines [1]. CNC machines are prevalent in the aerospace and automotive industries. CNC machines can accomplish machining tasks with high precision. However, the drawbacks of CNC machines is that they are: expensive, large, and heavy [2].

A serial robot is defined as a robot that is comprised of multiple links attached in series. Generally, the joints comprising the makeup of a serial robot are revolute and prismatic. The fixed link is attached to the ground and commonly referred to as the base and the last link in the chain is free to move in space to which an end-effector is usually attached [3]. Serial industrial robots are available at a low capital investment and possess the flexibility to perform a variety of tasks. The reusability and flexibility of robotic systems for machining applications make them an attractive alternative to CNC machines [4]. Moreover, robotic systems possess a better workspace to installation space ratio [3]. According to the research performed by Karim and Verl [2] and Brüning et al. [5], industrial robots have high economic potential for machining applications in the automotive and aerospace industries.

Drawbacks of serial robotic systems for machining include sensitivity to vibrations, low stiffness and these platforms require complex programming when compared to CNC machines [2, 5]. Field tests have also revealed that the serial robots fail to execute the programed path with high accuracy and repeatability once heavy machining is conducted and position errors occur. This stems from the low stiffness possessed by serial robots [6]. Dynamic elements affecting accuracy includes gear backlash, friction, temperature sensitivity and environmental disturbances [4].

A parallel kinematic manipulator (PKM) is defined as a robotic system that is comprised of two or more closed-loop kinematic chains which connects the end effector to the base of the machine [3]. PKMs present strong advantages in its high stiffness, payload to weight ratio and low inertia, but also suffer drawbacks of relatively small workspace, complicated calibration and contain design difficulties. Furthermore, the forward kinematic analysis for PKMs is highly non-linear and challenging to solve [7].

PKMs and hybrid kinematic architectures have been proposed by various researchers as suitable architectures for robotic machining. Some of these machining architectures have been successfully commercialized but are still large in size and as expensive as CNC machines [3, 8, 9]. The need for flexibility and adaptability of robotic machining platforms cannot be ignored and therefore serves as one of the motivators for this research.

Although there is a limitation on the workspace that can be achieved by PKMs, this research aimed to investigate the potential of machining relatively small workpieces since a desktop prototype was being investigated. The workspace is directly proportional to the size of components. The focus of this paper was to present a new 5 degree of freedom (DOF) desktop PKM for robotic machining and present the unique architecture with its merits. The architecture was also compared to other PKM architectures. The inverse and forward kinematic approach and analyses are presented.

2 CURRENT TRENDS IN ROBOTIC MACHINING

Researchers such as Briot et al. [10] have identified that parallel robots can be used as an alternative to serial robots for machining applications due to their load bearing advantage over serial robots and capability of higher positional accuracy [11]. A notable case is the use of the free leg hexapod by researchers such as Olarra et al. [12] for miniature machining applications. Several programs were written which enabled the working volume to be calculated for different foot positions and for determining optimal foot configuration based on the desired application. Glavonjic et al. [13] developed a new 3 DOF spatial parallel mechanism for a desktop 3-axis parallel kinematic milling machine. The parallel kinematic system was built as a low cost, educational desktop model. The concept was proven but the robotic system can only machine soft materials. Chablat and Wenger [14] developed a 3-DOF Translational Parallel Mechanism to be used for machining applications known as the Orthoglide.

Concerning a non-static location for a PKM, a hexapod developed by Choi et al. [15] was designed to move to the machining location and carry out machining tasks. An advantage of this design was the mobile nature of the machine. The first prototype was inferior to CNC machines in terms of feed and accuracy. Additional research was required for path planning of this hexapod and the inclusion of additional sensors to improve accuracy. A novel PKM based production system for aircraft wing assemblies, named PAW, was designed by Jin et al. [16]. The PAW system was suited for machining tasks such as drilling and trimming. PAW was designed to be mounted onto a gantry-like worktable thus not being confined to a static location. The machine possessed one translation and two rotational degrees of freedom. By performing kinematic analysis and dimensional synthesis, Jin et al. [16] were able to verify that the PAW system exhibits better performance than current mechanisms employed in industry. When compared to a Tricept machine architecture, it possessed a higher accuracy and stiffness due to the removal of the wrist portion which is the weakest part of a Tricept-type PKM.

A study performed by Wu et al. [17] addressed the calibration of a parallel kinematic machine tool through the use of a laser tracker. This was aimed at reducing errors from inaccurate joint motions. The research involved the development of a differential error model, optimized model and a statistical method. This calibration approach was carried out for a 5-DOF hexapod machine tool. Experimental results proved that this method of calibration reduced joint motion errors.

Smirnov et al. [18] researched tool path planning for the generation of energy efficient trajectories. The research exploited redundant DOFs to improve the energy efficiency of tool trajectories during contour forming. Experimental results were based on a 6 DOF industrial PKM-based machine known as KIM-1000. Whilst kinetic energy of the platform is ignored, the total energy consumed by the parallel linear drives is minimized and demonstrated an energy saving. Analytical and numerical solutions are used to address the optimization of the system.

Further research is required to produce a compact parallel kinematic robotic system that is able to compete against CNC machines. A shortcoming of existing research indicates that though novel designs are functional analytically, these platforms are not being commercialised due to future work pending and the designs being carried out as an educational exercise. In addition, the range of potential work pieces has not been described.

3 METHODOLOGY

3.1 Structural Type Synthesis

In a study undertaken by Koseki et al. [19], parallel kinematic manipulators were classified into three basic categories as shown in Table 1.

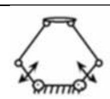
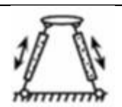
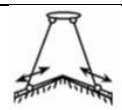
ASPECTS	**STRUCTURAL TYPE**		
	Rotary	**Prismatic**	**Fixed Linear**
			
Force output	Small	Large	Large
Actuator location	Base	On the leg	Base
Moving masses	Small	Large	Small
Speed	Very fast	-	Fast
Rigidity	Poor	Good	Good
Workspace	Large	Small	Large
Structure	Simple and compact	Large moving parts	Large footprint

Table 1 – The different PKM structural types [19]

The rotary type of PKM does not qualify for machining applications because the output force is small and the rigidity is poor. This structure type will be unable to withstand the vibration and chatter from machining tasks. It is desirable that the robotic machining system should occupy a small footprint therefore the fixed linear structural type was omitted from being used to develop conceptual designs.

The most applicable of the three types was the Prismatic type. Although it possesses the drawback of large moving masses it still possesses good rigidity and a smaller footprint than the Fixed Linear type. The Prismatic type is not the fastest of the three types but that is not a critical aspect for machining applications when high quality and high precision is valued. The Prismatic type architecture was therefore selected for further research.

Physical experimentation through rapid prototyping allowed a number of physical conceptual platforms to be built where the number of legs, types of joints, the arrangement of joints and legs and mounting points were varied. In order to match existing 5-axis CNC machines, the desired result was a robotic platform consisting of 3 translational and 2 rotational degrees of freedom. The axis normal to the base of the manipulator was regarded as the Z-axis. Thus rotation about the Z-axis is not permitted.

The novel architecture, named the R(Pa-IQ)R Parallel Manipulator, that was generated can be seen in a SolidWorks model depicted in Figure 1. The main area of innovation is the paired arrangement of the robot's legs. Notable characteristics of the architecture developed in this study are summarised below:

- The prismatic joints are actuated and all revolute joints are passive.
- The position and orientation of the pairs of legs suppresses undesired rotation of the end effector about the Z-axis.
- The layout of the revolute joints restricts each pair of legs to move in the same plane.
- The pairs of legs exhibit a parallelogram (Pa) shape when the end effector is performing translational movements and exhibits an irregular quadrilateral (IQ) shape when the end effector performs rotational movements. This led to the name given to the robotic manipulator.

The Grubler-Kutzbach equation i.e. equation (1) was used to determine the number of degrees of freedom permitted by the machining platform [7]. This equation yields a result of F = -6. The negative result indicates that this robotic manipulator is an over-constrained mechanism [20]. This is held true because the PKM has six actuators and five degrees of freedom.

$$F=(n-j-i)+\sum_i f_i \quad (1)$$

Where F is the number of degrees of freedom; n is the total number of links; j is the number of binary joints; f_i is the degrees of freedom allowed by a particular joint.

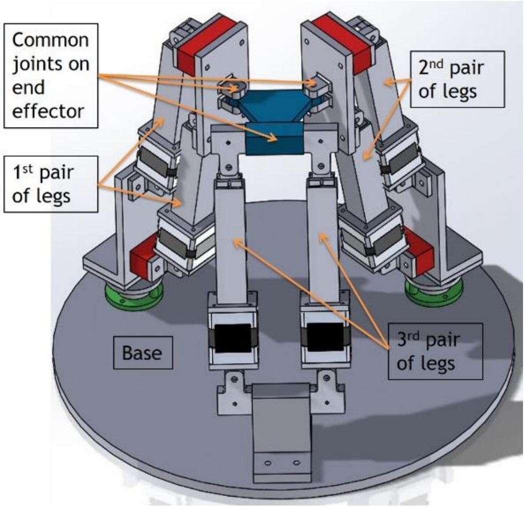


Figure 1 – The SolidWorks model of the PKM developed

The platform designed in this study is similar to the Hexapod (Steward-Gough Platform) in appearance [3]. Differences between the R(Pa-IQ)R and the Hexapod are listed in Table 2.

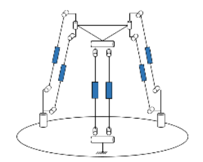
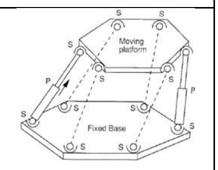
ASPECTS	**TYPE**	
	R(Pa-IQ)R	**Hexapod**
		
Degrees of freedom	5	6
Types of joints	Only prismatic and revolute.	Generally revolute, spherical, universal and prismatic joints
Pairs of legs	Yes	No
Common joints	Yes	Sometimes
Attachment points on base	3	6
Attachment points on end effector	4	6
Some actuators move completely in space	Yes	No

Table 2 – Differences between the R(Pa-IQ)R and a Hexapod

4 INVERSE KINEMATIC ANALYSIS

Inverse kinematics is defined as the known location and orientation of the end effector and the positions of all other desired link lengths and/or joint angles are to be determined [21]. The inverse kinematic solution is more relevant than the forward kinematic solution because the end effector traverses a predefined trajectory. All coordinates along this path are defined therefore the joint angles and/or actuator lengths are to be solved. The Geometric method was preferred over the Denavit-Hartenberg (DH) method to solve the inverse kinematics because it is less complex for PKMs [7].

4.1 Orientation of the end effector

The rotation of the end effector follows the Roll, Pitch and Yaw sequence of rotations. The rotation matrix is therefore given by equation (2):

$$R(\gamma,\beta,\alpha) = \begin{bmatrix} c\gamma\ c\beta & c\gamma\ c\beta\ s\alpha - s\gamma\ c\alpha & c\gamma\ s\beta\ c\alpha + s\gamma\ s\alpha \\ s\gamma\ c\beta & s\gamma\ s\beta\ s\alpha + c\gamma\ c\alpha & s\gamma\ s\beta\ c\alpha - c\gamma\ s\alpha \\ -s\beta & c\beta\ s\alpha & c\beta\ c\alpha \end{bmatrix} \quad (2)$$

Where c and s represent cosine and sine respectively; α, β and γ represent the rotation about the X, Y and Z axes respectively.

4.2 Inverse Kinematic Relationships

Figure 2 shows the outer vector loop across actuators 1 and 2. The vectors were specified in this manner such that passive joint angles need not be known at point A, B, C and D for the solution of the vector loops and simplified the analysis. The outer vector loop is formulated by taking two different paths to point D. The result is shown in equation 3 (which is applicable to other pairs of legs):

$$OA + AD = OP + R(\beta, \alpha)\ PD \quad (3)$$

Where the vector OA is the displacement from the mid-point of the base to the common joint of the leg pair 1 and 2; PD is the displacement vector from the mid-point of the end effector to joint D on the end effector to which the rotational matrix, R, is multiplied. OP is the vector relating the end effector relative to the mid-point of the base. AD is the displacement from the common joint on the base to the common joint on the end effector.

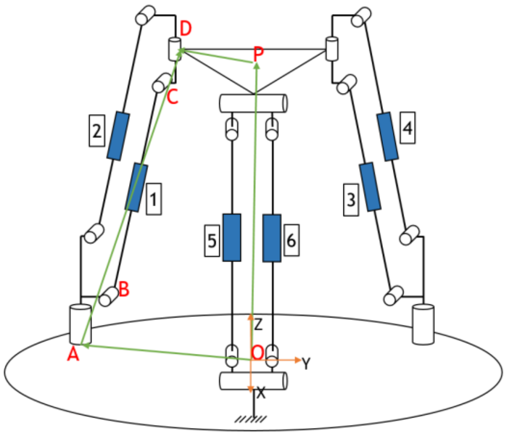


Figure 2 – The outer loop vector diagram

The arrangement of revolute joints constrains each actuator in a pair of legs to move in the same plane relative to each other. This collapses the vector loop into 2 dimensional plane. The inner vector loop equation is seen in equation 4 and also depicted in Figure 3(a) concerning actuator 1 and 2.

$$A_{1,2}B_i + B_iC_i + C_iD_{1,2} = A_{1,2}D_{1,2} \quad (4)$$

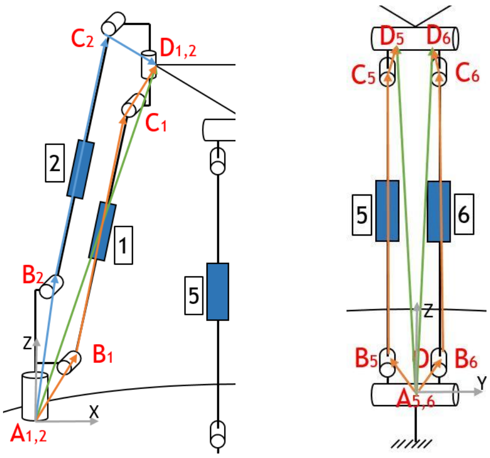
Where the subscripts refer to a leg and a subscript number-pair refers to a common joint. The value for i can be either 1 or 2. Vectors $A_{1,2}B_i$ and $C_iD_{1,2}$ are machine design parameters therefore vector B_iC_i can be solved. Observing the problem in 2 dimensions, a coordinate frame was placed at point A. Point A and point O are on the plane on the base therefore, the Z coordinate of point D is the same in frame O as in frame A. Since the magnitude of vector AD can be found from equation (3), the x coordinate of point D in frame A can be found through the theorem of Pythagoras. The x and z displacements of vectors $A_{1,2}B_i$ and $C_iD_{1,2}$ are known therefore the vector B_iC_i can be solved. The same method is followed for both actuators 1 and 2. The same method is followed to solve the outer and inner vector loops to solve the length of actuators 3 and 4. The subscript j can be either 3 or 4 in equation (5).

$$A_{3,4}B_j + B_jC_j + C_jD_{3,4} = A_{3,4}D_{3,4} \quad (5)$$

The lengths of actuators 5 and 6 is solved in a similar manner except that Point D is not a common point to which each of the vector loops converge. Secondly, since the Y axes of coordinate frame O and coordinate frame A_{56} are parallel, the Y values of the vectors $A_{5,6}D_5$ and $A_{5,6}D_6$ will be used for the analysis when solving the inner loop along the Y, Z plane upon which both legs 5 and 6 lie. The Y value is now used as the reference point. The inner vector loops for actuators 5 and 6 is shown in Figure 3(b) and equation 6 and 7 shows their inner loop equations respectively.

$$A_{5,6}B_5 + B_5C_5 + C_5D_5 = A_{5,6}D_5 \quad (6)$$

$$A_{5,6}B_6 + B_6C_6 + C_6D_6 = A_{5,6}D_6 \quad (7)$$



(a) (b)

Figure 3 – The Inner loop vector diagrams

5 EXPERIMENTS AND RESULTS

5.1 Inverse Kinematic Results

SolidWorks was used in conjunction with Matlab to validate the equations developed for the inverse kinematics analysis. This was aimed at being completed before developing a physical testing system. The PKM was tested for translational movements of the end effector. The end effector was moved to random locations in SolidWorks and then each leg length was measured within SolidWorks. The end effector locations which were measured in SolidWorks was entrered into Matlab and the inverse kinematic formulae were utilised. Figure 4 displays the methodology followed for the inverse kinematic equations to be valiated. The results are depicted in Table 3. It can be seen from Table 3 that the inverse kinematic equations possess high accuracy since the largest error incurred was 0.01 mm.

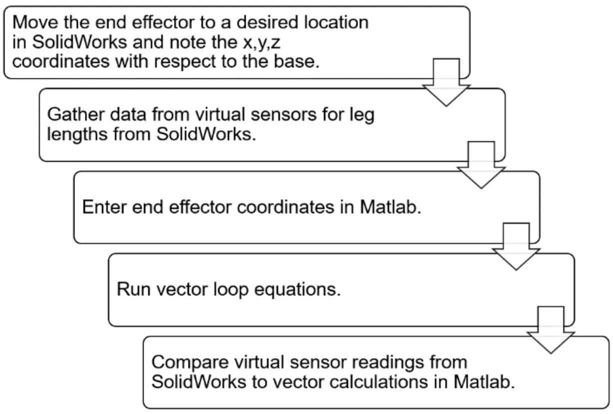


Figure 4 – Methodology for inverse kinematic calculations

END EFFECTOR COORDINATE (mm)			LARGEST ERROR AMONGST THE 6 LEGS (mm)
x	y	z	
74.22	-36.4	296.09	0
116.01	-93.61	275.44	0.01
151.47	-6.99	298.31	0.01
82.52	91.62	292.19	0.01
130.97	104.45	271.42	0
140.22	-19.32	298.14	0.01
138.77	-100.37	290.81	0
174.23	21.43	295.38	0
72.52	-35.9	292.28	0.01
184.18	-6.22	305.47	0

Table 3 – Inverse kinematic errors for translational motion cross-checked between Matlab and SolidWorks

5.2 Forward Kinematic Results

The Newton Raphson (NR) numerical convergence technique makes use of an initial guess which goes thorough a number of iterations until the error between solutions is minimal enough to deem the solution valid. This method was employed to validate the kinematic equations with respect to forward kinematic analysis. This method was selected as it has been undertaken by various authors when performing forward kinematic analysis on PKMs [22, 23]. When the NR method was performed, the error between SolidWorks and Matlab for the end effector location was at a maximum value of 0.0172 mm after 25 tests. It was noted that when the value of zero is guessed as a solution for any coordinate, the NR solution breaks down and produces large errors. The results conform to the NR iterative pattern such that the larger the guess devation from the true value, the greater the number of iterations that is required to be performed and is validated in Figure 5.

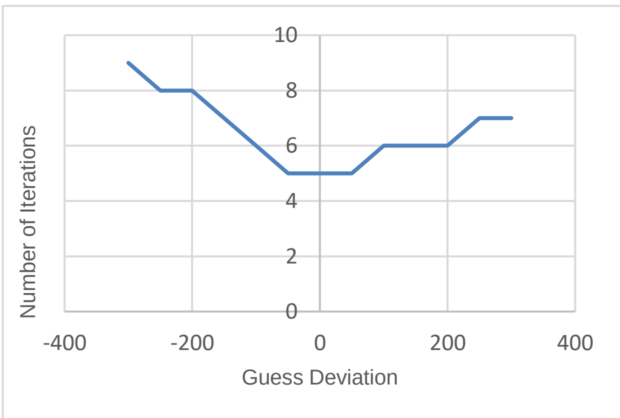


Figure 5 – Graph of guess deviation versus number of iterations

6 CONCLUSIONS

The R(Pa-IQ)R parallel manipulator developed through this research study was an apt robotic machining platform due to the six legs it possessed thus increasing its stiffness. The triangulation configuration of the legs added to the overall stiffness of the machine. The rotational analysis of the end effector is still to be conducted. A physical testing system was constructed to validate simulations. Current schemes of analysis for the forward and inverse kinematic analyses resulted in high accuracy and was therefore carried forward in this study. Through the forward and inverse kinematic analysis, it was discovered that the difficulty of solving the kinematic equations was directly proportional to the complexity of the design of the legs of the machine.

Research showed that machining tasks require accuracy that ranged from low to high depending, on the type of product being produced. For disruptive robotic machining to start being adopted by industry, robotic systems need to be able to perform machining tasks with higher accuracy and throughput. Future research on PKMs could include aspects related to machine calibration, trajectory planning, stiffness analysis, open architecture control and machine reconfigurability.

7 ACKNOWLEDGEMENTS

We would like to thank the University of KwaZulu-Natal for the Reino Stegen Scholarship and for the provision of infrastructure and resources that enabled this research to take place. We also thank the National Research Foundation (NRF) of South Africa for the scholarship received for this research.

8 REFERENCES

[1] Chen, Y., Dong F., 2013, Robot machining: recent development and future research issues, The International Journal of Advanced Manufacturing Technology, 66/9, 1489-1497.

[2] Karim, A., Verl, A., 2013, Challenges and obstacles in robot-machining, 44th International Symposium on Robotics (ISR) IEEE, 1-4.

[3] Pandilov, Z., Dukovski, V., 2014, Comparison of the characteristics between serial and parallel robots, Acta Technica Corviniensis-Bulletin of Engineering, 7/1, 143-160.

[4] Schneider, U., Drust, M., Diaz Posada, J., Verl, A., 2013, Position control of an industrial robot using an optical measurement system for machining purposes, in Proceedings of the 11th International Conference on Manufacturing Research (ICMR2013), 19, 307-312.

[5] Brüning, J., Denkena, B., Dittrich, M.A., Park, H.S., 2016, Simulation Based Planning of Machining Processes with Industrial Robots, Procedia Manufacturing, 6, 17-24.

[6] Wang, J., Zhang, H., Fuhlbrigge, T., 2009, Improving machining accuracy with robot deformation compensation, in 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, 3826-3831.

[7] Tsai, L.-W., 1999, Robot analysis: the mechanics of serial and parallel manipulators, 1st Edition ed. John Wiley & Sons.

[8] Pandilov, Z., Rall, K., 2006, Parallel kinematics machine tools: History, present, future, Ana Lazarevska, 3, 3-20.

[9] Borràs, J., Thomas, F., 2010, Singularity-invariant leg substitutions in pentapods, in 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2766-2771.

[10] Briot, S., Pashkevich, A., Chablat, D., 2010, Optimal technology-oriented design of parallel robots for high-speed machining applications, in 2010 IEEE International Conference on Robotics and Automation, 1155-1161.

[11] Plitea, N., Lese, D., Pisla, D., Vaida, C., 2013, Structural design and kinematics of a new parallel reconfigurable robot, Robotics and Computer-Integrated Manufacturing, 29/1, 219-235.

[12] Olarra, A., Allen, J.M., Axinte, D.A., 2014, Experimental evaluation of a special purpose miniature machine tool with parallel kinematics architecture: Free leg hexapod, Precision Engineering, 38/3, 589-604.

[13] Glavonjic, M., Milutinovic, D., Zivanovic, S., Dimic, Z., Kvrgic, V., 2010, Desktop 3-axis parallel kinematic milling machine, The International Journal of Advanced Manufacturing Technology, 46/1, 51-60.

[14] Chablat, D., Wenger, P., 2003, Architecture optimization of a 3-DOF translational parallel mechanism for machining applications, the orthoglide, IEEE Transactions on Robotics and Automation, 19/3, 403-410.

[15] Choi, S., Cho, C.N., Kim, H.J., 2015, Development of hexapod robot for machining, in 2015 15th International Conference on Control, Automation and Systems (ICCAS), 738-740.

[16] Jin, Y., Kong, X., Higgins, C., Price, M., 2012, Kinematic design of a new parallel kinematic machine for aircraft wing assembly, in IEEE 10th International Conference on Industrial Informatics, 669-674.

[17] Wu, J.-F., Zhang, R., Wang, R, -H., Yao, Y.-X., 2014, A systematic optimization approach for the calibration of parallel kinematics machine tools by a laser tracker, International Journal of Machine Tools and Manufacture, 86, 1-11.

[18] Smirnov, V., Plyusnin, V., Mirzaeva, G., 2013, Energy efficient trajectories of industrial machine tools with parallel kinematics, in 2013

IEEE International Conference on Industrial Technology (ICIT), 1267-1272.

[19] Koseki, Y., Arai, T., Sugimoto, K., Takatuji, T., Goto, M., 1998, Design and accuracy evaluation of high-speed and high precision parallel mechanism, in 1998 IEEE International Conference on Robotics and Automation Proceedings (Cat. No.98CH36146), 2, 1340-1345.

[20] Merlet, J.-P., 2006, Parallel Robots, 2nd Edition ed. Springer Science & Business Media.

[21] Raj, N.J., Iyer, K., Dash, A.K., 2016, Design, fabrication, kinematic analysis and control of a 3-DOF serial manipulator, in 2016 International Conference on Next Generation Intelligent Systems (ICNGIS), 1-6.

[22] Abo-Shanab, R.F., An Efficient Method for Solving the Direct Kinematics of Parallel Manipulators Following a Trajectory, Journal of Automation and Control Engineering, 2/ 3, 228-233.

[23] Jakobović, D., Budin, L., 2002, Forward kinematics of a Stewart platform mechanism, Faculty of Electrical Engineering and Computing, Unska, Zagreb, Croatia

9 BIOGRAPHY

Wesley Dharmalingum obtained his BSc.Eng degree in Mechanical Engineering from the University of KwaZulu-Natal in 2013. He is currently pursuing a MSc.Eng degree at the University of KwaZulu-Natal, South Africa.

Jared Padayachee holds MSc.Eng and PhD degrees in Mechanical Engineering from the University of KwaZulu-Natal. He is currently a senior lecturer in the Department of Mechanical Engineering at the University of KwaZulu-Natal. His area of research interest is Optimisation in Robotics and Manufacturing.

Glen Bright possesses MSc.Eng and PhD degrees in in Mechanical Engineering from the University of KwaZulu-Natal. He is currently the Dean of Engineering at the University of KwaZulu-Natal. His area of research interest is in Mechatronics and Robotics.

ELIC – Teacher as a Medium to Build a New Generation of Skilled Engineers

J. Niemann[1], C. Fussenecker[2], M. Schlösser[3], T. Ahrens[4]
[1,2,3,4]Department of Mechanical and Process Engineering
University of Applied Sciences Düsseldorf, Germany

Abstract
Serious shortages of qualified professionals and technical job-specific skills are hampering Europe's sustainable growth. Even more so, relatively newer technologies, such as digitalization and Industry 4.0, require a new set of qualifications for the future engineers. Deficits in these areas can be critical for innovation and, therefore, also for the well-being of the industry and economy. New skills as well as new training methods are required to be successful and to meet the requirements set by industry and society. ELIC (Engineering Literacy Online - Teachers as Medium for Change) is an ongoing EU project supported by the Erasmus+ program. With this research project, the competence of a teacher's teaching methods and technical didactics are targeted, especially those teaching secondary-level Science Technology Engineering Mathematics (STEM)-related subjects, as well as their ability to incorporate engineering topics in natural sciences classes, motivating more pupils to choose an engineering career.

Keywords
Engineering Education, Skills, Knowledge Management, MOOC, STEM.

1 INTRODUCTION

Many of today's societal, engineering and industrial challenges are too complex and multidimensional that they cannot only be handled with an old-fashioned knowledge of science, technology and soft skills but also need an interdisciplinary view on things as well as socio-economic capabilities. [1] Digitalization and Industry 4.0 are two of those challenges influencing the way we are living and working that today's and future engineers must deal with.

New market developments and values, such as Industry 4.0 and Digitalization, are transforming not only the producing industries but also influencing the way we are living and working. Industry 4.0 is a future vision, first described in the high-tech strategy of the German government. Main content is the information and communication technologies, such as Cyber-Physical Systems, Internet of Things, Physical Internet and Internet of Services, in order to achieve a high degree of flexibility in production. This leads to higher productivity rates through real-time monitoring and diagnosis, and a lower wastage rate of material in production. [2]

In the Future of Jobs Report 2018 by the Centre for the New Economy and Society of the World Economic Forum it is explicitly pointed out, that work tasks and abilities, which were once reserved for humans are now more and more carried out by machines and robots. This is leading to a growing concern about the impact of jobs and a general risk for the wellbeing of the people, industry and government. By 2025 the rate of automation has reached 52% regarding the division of labour as share of hours spent. Moreover, more than 23% of the companies will use humanoid robots in their production process, replacing certain job role, by then. [3]

Apart from that, another challenge is fostering the European economic growth. A serious shortage of qualified professionals and technical job-specific skills are hampering Europe's sustainable growth. The demographic change and an often inadequate engineering education can be named as two reasons of this threat. [4] Not to be left out is the fact that there is an existing gender gap, meaning that a persistent low rate of females in engineering is observed in many European countries. [5] A low birth rate due, as a result of the demographic change, cannot be changed overnight, whereas an adaption of the engineering education, in the direction of motivating more pupils and students to choose the engineering path, could be one suitable solution which can be approached. However, the vision of having enough highly qualified employees being able to meet the challenges of digitalization and Industry 4.0 cannot be met right away.

An adaption of the higher education towards Industry 4.0 and Digitalization, particularly in the engineering education, has to be achieved. [2] As a result, education methods for engineers need to be adapted, appreciating the complex environmental and societal systems, continuously changing technologies, evidenced-based design, and the ability of good communication. [7] Even the current research High-Tech-strategy from Germany for 2015-2020 specifically named the digitalization of economy and society as well as an innovative workspace as two out of six initiatives for this strategy. [6] The way in which the new generation of skilled engineers are trained and educated has changed over the past year. Flexible Learning revolution, e-Learning, Life-Long Learning, Learning Management Systems, Blended Learning, Flipped Classrooms, Learning Analytics, or Massive Open Online Courses (MOOCs) are just a few tools and trends from the educational development. [8] [9]

Together with several universities and schools, ECQA (European Certification and Qualification Association) develops a compliant skill set and exam for teachers, as well as a MOOC platform to train teachers. Leading to this, existing structures in different European countries are analyzed and compared to the requirements of industry and academia. Based on the discovered gaps, a skill card is developed and a curriculum established.

Teachers of Science, Technology, Engineering and Mathematics (STEM) are seen as a medium of change in the way that they can already influence the future engineers during their study period in secondary schools, while teaching the STEM-subjects.

A group of Italian, Czech, Austrian and German universities and schools as well as an E-Learning provider saw the opportunity to develop an online based training for teachers of STEM-subjects to give them tools to improve their way of teaching, leading to a new generation of qualified engineers.

To come up with the right content and training method, the engineering skills required nowadays and in the future are analyzed and compared to the ones which are currently taught. Therefore, a literature research has been conducted. After that, a proposal for a new Education 4.0 model as well as the matching EU-funded project ELIC are described in detail.

2 FUTURE JOB SKILLS

While analysing different literature discussing the changing importance of future skills for successful engineers, the following skills have been mentioned frequently (Figure 1). According to surveys conducted by the authors as well as the frequency of being mentioned, an evaluation of their importance in the past, now and in the future has been carried out.

Due to new technologies and business models and the changing division of labour between workers and machines, the current job profiles are being transformed. It is expected, that by 2022, the skills required to perform most jobs will have shifted significantly. More jobs such as Artificial Intelligence and Machine Learning Specialists, Big Data Specialists, Process Automation Experts or User Experience and Human-Machine Interaction Designers are set in place. Jobs which also require a certain skill set. In order to manage the skills gaps, caused by the adoption of new technologies, the current staff is required to re- and upskill. Furthermore, it is expected that new permanent staff, already possessing skills relevant to new technologies, will be hired. Staff which is now still in the education process. [3]

However, when educating today's and future engineers other challenges have to be met, meaning that they no longer need to be an expert in one field only but need to come up with new ideas and ways of handling or even defining problems. [1] Apart from that the analytical and innovative thinking as well as the active lifelong learning ability belong to the skills of the future and became more important in recent years. Further human' skills such as creativity, originality and initiative, critical thinking, persuasion and negotiation will likewise retain or increase their value, as will attention to detail and flexibility. All this will be taken into consideration, when developing the content for the EU project ELIC and determining the training requirements. [1] [3]

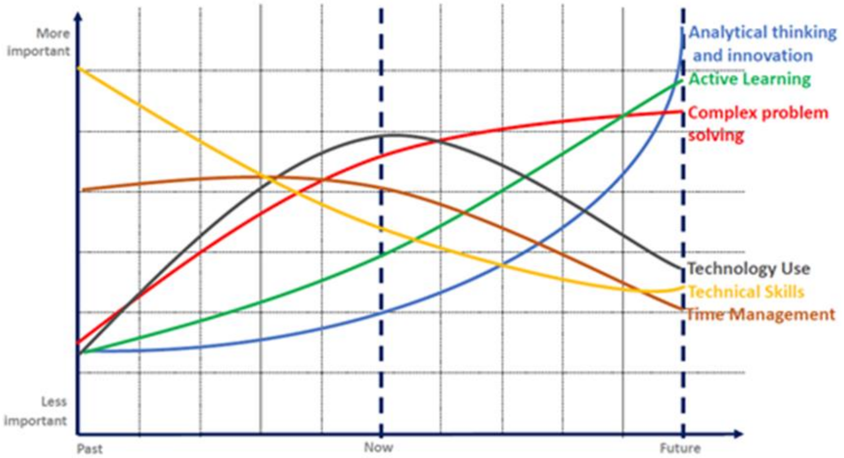


Figure 1 - Engineering Skills in the past, now and in the future [1] [2] [3] [7] [9] [10] [11]

3 EDUCATING FUTURE ENGINEERS: CHALLENGES AND FUTURE TEACHING METHODS

For a long time, the engineering education has not been changed, although changes in society, business and economy took place. [1] One reason for that was, surely, that the old system has been proven extremely successful, making the German “Dipl-Ing.” a No.1 export hit. However, due to a lack of adaption in certain areas, a major renewal has to be undertaken.

However, education has to change with changing times. In the future, an engineer might also be competing with a robot in certain areas or even worse, he might be replaced by it. Being able to combine an excellent knowledge in certain disciplines as well as the ability to think out of the box, can therefore be seen as very important assets for the future engineers. The future engineer needs to define and solve problems at the same time; this will only give him a big advantage in comparison to automated systems and artificial intelligence. Students have to be prepared for jobs which don’t yet exist, using technologies, which have not been invented, solving problems, which right now, do not exist. [12]

Another challenge, which also needs to be met, is the reasons for a low enrolment within the engineering study programs. Recent surveys have shown that the engineering profession itself is becoming less popular. Lopez-Martin (2010) found out that young people often favour a less stressful and easier career, as they already enjoy a high living standard. Moreover, taking the current technological standard as granted leads to a possible underestimation of the importance of engineering. [13]

The under 2 identified important skills for the engineer of the future also pose a challenge to the teachers, schools and universities in how they can be met and adopted.

Therefore, the way engineers are educated, has to be adapted as well. An analysis showed that, in recent times, some engineering curricula have been transformed, including a focus shift from knowledge sharing to the training of skills, such as analytical thinking and innovation or active learning strategies. [1] [3]

However, still many engineering curricula are almost completely filled with technical knowledge. [15] Bit by bit, new courses for engineering students, such as Engineering Conferences - a master’s class at Düsseldorf University of Applied Sciences - are developed. The course uses a learner-centred approach in which students generate, from their undergraduate bachelor thesis, a research paper, take part in a review process and present their paper during a poster presentation. The primary aim is to develop skills and competences needed in the world of modern engineering. [9] [15]

Moreover, projects, which expose students to practical design, systematic thinking and project management as well as real-world experiences, such as internships and research proposals together with industry and their local community, are further ideas. Some universities have already integrated mandatory cooperation and industry visits into the curricula, engaging the industry as an essential partner for the future education of engineering students. [16]

Also, the way the future engineers are educated has changed. An interesting possibility is the introduction of new teaching and learning models, encouraging digitalization and smart interaction between the involved parties. A flexible learning revolution, e-Learning, Life-Long Learning, Learning Management Systems, Blended Learning, Flipped Classrooms, Learning Analytics, or Massive Open Online Courses (MOOCs) are just a few tools and trends from the current educational development. [7] [8] Figure 2 shows an overview of teaching methods, which can not only be used to optimise the education of engineering students.

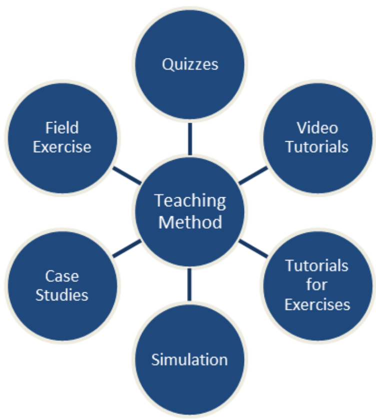


Figure 2 - Teaching and Learning Concept

By analysing the future world of work of an engineer, the potential impact on the engineering education on different levels gives a good overview on what needs to be done in order to meet the challenges of educating the future engineers. The idea/goal of the Erasmus+-Project ELIC is to target teachers, as a medium of change, as they are one of the first points of contact, educating the future engineers and motivating pupils to choose this path.

4 ERASMUS+-PROJECT: ENGINEERING LITERACY ONLINE (ELIC)

4.1 Idea and Status Quo

A serious shortage of qualified professionals and a lack of technical job-specific skills are negatively influencing the sustainable growth of Europe’s economy. The idea is to empower teachers as a medium of change, in order to ward off this threat.

By analyzing the status quo of teachers of STEM-subjects, looking at the learning plans and learning outcomes as well as taking into account the experiences teachers and further responsible authorities have, the aim is to identify current barriers to successfully studying natural sciences in secondary school and later choosing an engineering study program. Engineering Literacy, in the Project, stands for interdisciplinary teaching through practical examples that combine natural sciences subjects and technical sciences input. This should lead to an increased functional thinking among the pupils. The link between basic natural science knowledge and how to apply it when solving challenges in different fields becomes visible and makes secondary school pupils more curious, thereby increasing their interest in engineering.

A team of Austrian, Czech, Italian and German universities, schools and learning institutions is working on that common goal to set up an online MOOC-system, where teachers are trained. Funded by Erasmus+ this project falls under the Lifelong Learning Program, meaning that people should be enabled to stimulate their own learning experience at any stage of their life. [17] This all leads to the lucky coincidence that the engineer of today also finds himself in a lifelong learning process. The project is funded for a duration of 24 months. The starting date was October 1st, 2017. [18]

4.2 Structure and Intellectual Output

Due to the requirements set up by the European Union and in order to receive the funding, measurable goals, such as an Intellectual Output (IO), have to be defined. For the ELIC-research project, Figure 3 shows the seven IOs, which have been determined, structuring the tasks of the two-year research work. After starting with the research on the training needs, a GAP-Analysis was conducted; taking into account what kind of education the learning plan currently specifies, comparing it to what is needed by society and industry. Stakeholder, such as school teacher, industrial partner as well as associations dealing with education matters, are questioned during focus group interviews. Their requirements are then also implemented into a Skill Card, naming the engineering topics and their linkage to the STEM-subjects. This all leads to the development of the training curriculum, where the detailed teaching content combined with the teaching tools, learning outcomes and time management, is defined. Then a MOOC-Platform is set in order to have a suitable platform to disseminate the content. [18]

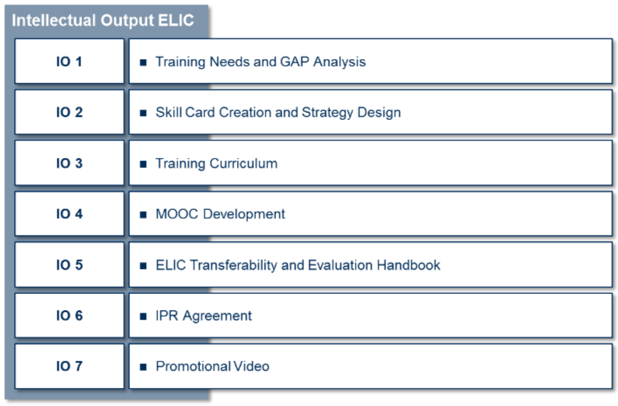


Figure 3 - Intellectual outputs of Erasmus+-project ELIC

4.3 Skill Set & Curriculum

The development of the curriculum is scientific based pedagogic work from all partners. They also make sure that higher education standards are followed, also being tailored to the target group – the teacher. School partners as well as teacher training colleges and school authorities will be constantly included during the development process. This is to make sure that the end product has a high usability, guaranteeing a constant usage and, most important, a long-term effect.

The innovation of the skill card and the curriculum lies in the content. Technical didactics and engineering literacy are currently a hot topic, not only in the participating European countries. [7] [8] The didactics used within the MOOC and in the curriculum will be new and innovative. The generated output will be transferable to different countries, even targeting groups, seeking for additional knowledge on technical didactics and interdisciplinary engineering topics. The chosen content will include the following subjects:

- Engineering Literacy
  - Introduction
- Automotive Case Studies
  - Battery System
  - Light System
  - E-Motor
  - Combustion Engine
- Hot Topics
  - Energy Management
  - Cybersecurity
  - Autonomous Driving
  - Women in Engineering

While the “Automotive Case Studies” generally appear to be more interesting to the male pupils, the content of “Hot Topics” is designed to get the female pupils interested. One draft is illustrated in Table 1, where experiments for the subject “E-Motor” are categorized according to the STEM-subjects taught in secondary school. [18]

Integrated Sciences Teaching							
E-Motor							
Experiments in Schools That Can be Linked							
Mechanics – Parts of motor	Physics – Circuits, Kirchhoff, currents, etc.	Chemistry – Rare Erath Materials, etc.	Technology – Hall Sensors, rpm measurements	Informatics – computer e.g. controlling rpm, controlling phase currents	Mathematics – functions, models, characteristic curves	Biology – avoiding CO2 emission, but CO2 emission, exploiting earth rare earth magnets	Ethics – Impact on Society

Table 1 - Example of interdisciplinary engineering teaching content for secondary schools [18]

5 CONCLUSIONS

Industry 4.0 and Digitalization challenge today's and future engineering education. As a result, educational needs of students as well as industrial requirements are changing.

A more holistic approach on engineering education needs to be found and implemented. The new generation of engineers can no longer only be experts in their technical areas but must face the societal and industrial challenges that the 21st century brings with it. In an attempt to answer the question of which skills are needed to generate a new generation of successful engineers, the still ongoing European research project ELIC was developed. By giving teachers knowledge and tools to motivate more pupils to choose the engineering path, that way pupils are targeted at a relatively early education stage. An important point, which often has been missed out due to a lack of knowledge of the teacher, is to show the pupils a realistic picture on how their future work life as an engineer might look like by giving practical examples and combing the theoretical content of STEM-subjects. If only some of the above described improvements and renewals of the engineering education and the way it is conveyed succeeds, a new generation of well skilled engineers will be able to tackle the challenges globalization brings with it. The European research project ELIC, which will only be finalized in September 2019, is one step into the right direction in meeting the challenges.

However, there is still a long way to go to make sure that the engineers of the future are educated in a suitable way, considering the changing required skills. Besides identifying teachers of STEM-subjects as a medium of change, industry and universities also have to accept their role of responsibility in this process.

6 REFERENCES

[1] Kamp, A.: Engineering education in the rapidly changing world: Rethinking the mission and vision on engineering education at TU Delft, Technical Report, TU Delft, Faculty of Aerospace Engineering, October 2014.

[2] Kayikici, Y., Gencay, E., Coskun, S.: Adapting Engineering Education to Industrie 4.0 Vision, The 16th Production Research Symposium, Istanbul, Turkey, October 2016.

[3] World Economic Forum, Centre for the New Economy and Society: The Future of Jobs Report 2018, Insight Report, Available from: http://www3.weforum.org/docs/WEF_Future_of_Jobs_2018.pdf [accessed Nov 27 2018].

[4] European Centre for the Development of Vocational Training (CEDEFOP): Skill shortages and gaps in European enterprise, Luxembourg 2015.

[5] Abdulwahed, A., Ghani, S., Hamouda, A.: Life in Engineering Program: Impact of an Engineering Outreach Project in K-12, IEEE Global Engineering Education Conference (EDUCON), 2013.

[6] BMBF: Fortschritt durch Forschung und Innovation: Die High-Tech Strategie. Available from: https://www.bmbf.de/pub/Fortschritt_durch_Forschung_und_Innovation.pdf [accessed Jul 24 2018].

[7] Moytl, B., Baronio, G., Uberti, S., Sperzanza, D., Filippi,S.: How will change the future engineers' skills in the Industry 4.0 framework? A questionnaire survey, 27th International Conference on Flexible Automation and Intelligent Manufacturing, FAIM2017, Modena, Italy, June 2017.

[8] Huba, M., Kozák, S.: From e-Learning to Industry 4.0, 2016 International Conference on Emerging eLearning Technologies and Applications (ICETA), November 2016.

[9] Zielke, T., Neef, M., Fussenecker, C.: Teaching Engineering Conferences, Canadian Engineering Education Association (CEEA17) Conference at Toronto, Canada, June 2017.

[10] Lattuca, L.R., Terenzini, P.T., Volkwein, J.F.: Engineering Change – A Study of the Impact of EC2000, Pennsylvania 2006.

[11] Brown, L.: Desired Attributes of a Global Engineer, ASEE Corporate Member Council 2010.

[12] Krueger, N.: Preparing students for jobs that don't exist, ISTE Conference & Expo Education leadership STEM & STEAM, available from: https://www.iste.org/explore/articledetail?articleid=1002 [accessed August 22 2018].

[13] Lopez-Martin, A.J.: Attracting Prospective Engineering Students in the Emerging European Space for Higher Education, IEEE Transaction on Education, vol. 53, no.1, pp.46-52, 2010.

[14] Grasso, D., Burkins, M. B.: Holistic Engineering Education, Beyond Technology, Springer Verlag, New York, 2010.

[15] Neef, M., Zielke, T., Fussenecker, C.: Engineering Conferences – An Innovative Course for Master Students in Engineering, ETALEE 2017 - Exploring Teaching for Active Learning in Engineering Education, May 2017.

[16] Solms, von S., Nel, H.: STEM Project Based Learning: Towards improving secondary school performance in mathematics and science, 2017 IEEE AFRICON, 2017.

[17] European Commission: Life Long Learning Programme, available from: http://ec.europa.eu/education/lifelong-learning-programme_de [accessed Jul 30 2018].

[18] FH Joanneum, ELIC – Engineering Literacy Online, in: Strategic Partnership for School Education, KA201, ELIC, Erasmus+ Programme, 2017-1-AT01-KA201-0305034, 2017.

7 BIOGRAPHY

Claudia Fussenecker holds a M.A. in European Management from the FH Bund in Brühl and a Diploma in Industrial Management from EUFH in Brühl. She is currently a scientific assistant in the Department of Mechanical and Process Engineering, focusing on Marketing and Services and Life Cycle Management.

Tim Ahrens holds a B.Eng in Industrial Engineering from the University of Applied Sciences Düsseldorf. Currently is he doing his M.Sc. in International Industrial Engineering and a student assistant the Department of Mechanical and Process Engineering, focusing on Marketing and Services and Life Cycle Management.

Jörg Niemann obtained his PhD degree in Manufacturing Engineering from the University of Stuttgart. In 2012 he was appointed Professor in Business Management and Mechanical Engineering at the Düsseldorf University of Applied Sciences, Germany.

Martin Schloesser holds a B.Eng. with specialization in Product Development and Production from the University of Applied Sciences Duesseldorf. He is currently a scientific assistant in the Department of Mechanical and Process Engineering, focusing on production simulation and Life Cycle Management.

VR-Based Design Process of Industrial Products

F. Bellalouna[1]

[1]Faculty of Mechanical Engineering and Mechatronics,
Karlsruhe University of Applied Sciences, Germany

Abstract

This paper introduces the historic development of the virtual reality technology (VR technology) and its current situation in the industrial application. Afterwards the different common types of the VR approach is going to be presented. The architecture of the HMD-based (Head Mounted Display) VR-system, that is relevant for this paper, will be illustrated. In the second part of this paper, the use of the virtual reality technology within the industrial product life-cycle is going to be discussed. Based on two feasibility studies carried out and evaluated within an industrial cooperation project, the potential of the VR technology and the challenges of its application in practice will be outlined. The first feasibility study deals with the use of the VR technology within the requirement-management and design-review process. The second feasibility study presents a new approach for VR-based FMEA (Failure Mode and Effects Analysis) and CA (Criticality Analysis).

Keywords

Virtual Reality (VR), Head Mounted Display (HMD), Product Life-Cycle Processes

1 INTRODUCTION

1.1 History the VR-Technology

Virtual reality is composed of an interactive computer simulation, which senses the user's state and operation and replaces or augments sensory feedback information to one or more senses in a way that the user gets a sense of being immersed in the simulation [1].

The vision of virtual reality arose already in the thirties. Aldous Huxley introduced 1937 in his book Brave New World the concept of feeling movies that involve touch in addition to sight and sound [1]. The first virtual reality solution named Sensorama was developed and realized by Morton Heilig in 1957. Sensorama is a machine that offered a virtual bicycle riding experience [1]. 1968 Ivan Sutherland developed a head-mounted display connected to a virtual environment allowed a user to experience virtual environments using different senses, but did not allow any interaction with the environment. Myron Krueger developed 1970 the first environments that reacted to the user's actions [1]. In the eighties and nineties the VR technology saw rise in public awareness, through VR entertainment productions (e.g. Last Action Hero, Brainscan) and the gaming industry (e.g. Virtual Boy by Nintendo). The first practical VR-applications in the industry appeared during the 2000s, especially in the defense, aerospace and automotive industry [2]. Using the 360° cave concept, VR based applications for e.g. cars and airplanes design review, training simulation for pilots, were developed. The real breakthrough of the VR technology began in 2016 with the launch of the VR-HMD (Head Mounted Display) solutions e.g. Oculus Rift, HTC Vive. Due to the high and faster improvement in performance of PC, mobile devices, and specialized hardware, and software the VR technology becomes more practicable in use. This leads to huge increase of VR based applications mainly in the consumer industry, e.g. gaming, entertainment, film, tourism, (figure 1) and to huge demand by end-consumers, which makes the VR equipment (Hardware and Software) affordable and easy to use.

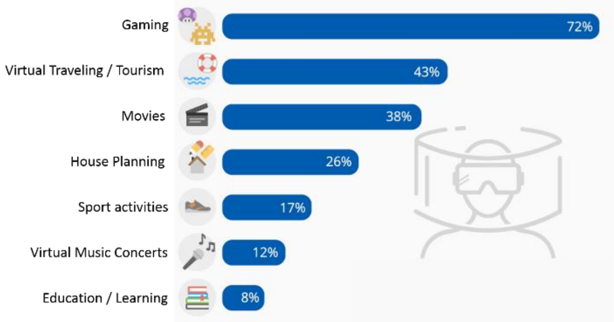


Figure 1 - Virtual reality is especially attractive in gaming [6].

1.2 Current situation in the Industry and Future potential of the VR-Technology

Currently, the consumer industry presents the main growth's driver of the VR technology in terms of performance and applications increase and in terms of cost reduction. Quite the contrary is the situation in industry. The use of the VR technology within design process of industrial products with few exceptions in the automotive and aerospace industry is still sobering. The most of the companies are either in waiting position or in feasibility study's phase. The most developed industrial VR applications have a character of a prototype. The reticence of the industry regarding the use of the VR technology along the product life-cycle processes is mainly due to the following issues:

- Negative experiences with the old generations of the VR solutions (e.g. VR power wall, VR cave) in terms of cost-performance ration.
- Missing of VR technology know-how and competencies related to the performance and to the potential of the new VR solutions.
- Missing of adequate engineering-oriented VR solutions and standards particularly with regard to the handling and the visualisation of CAD- and CAE-Data. Actually, the preparation of the CAD- und Simulations-Data for a VR scene is time-consuming and cost-intensive.
- Missing of specialized VR engineers for the development and the implementation of industrial-oriented VR applications.

However, a several trend studies and forecasts show that the interest and the investment of the industrial and engineering companies in the VR technology will increase considerably in the near future (figure 2). It is to be expected that this trend will reinforce the development efforts of industrial-oriented VR solutions and standards especially for the CAD- and CAE-data transformation and visualization.

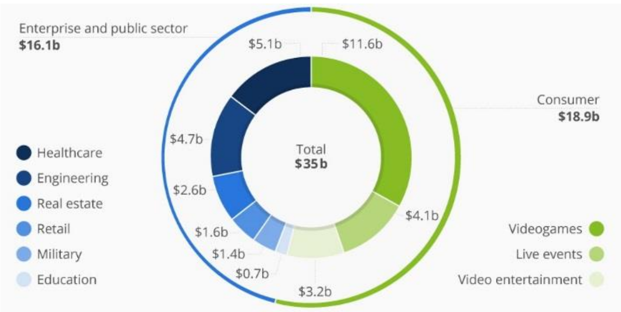


Figure 2 - Predicted market size of VR applications for different use cases in 2025 [7].

2 TYPES OF THE VR-SYSTEMS

Nowadays, there are basically three main types of VR solutions that can be found in the industry and in the universities and research institutes:

- VR-Powerwall: The VR-Powerwall represents a simple one-surface stereoscopic projection, which allows a three-dimensional visualization of objects. The VR-Powerwall is consisting of: a Tracking-System to capture the user's 3D-motion and 3D-position (e.g. Cameras); Interaction-System for interconnectivity between user and 3D-objects (e.g. Flystick); Visualization-System for image projection (e.g. projector and screen); VR-Application for stereoscopic 3D-presentation and Stereoscopic LCD shutter glasses to convey a 3D image [3].
- Cave Automatic Virtual Environment: The cave represents a further enhancement of the VR-Powerwall. Instead of a simple one-surface stereoscopic projection, the CAVE is a room consisting of several walls and a floor with 3D stereoscopic projections [3].
- VR-HMD (Head Mounted Display): A VR-HMD is a display device, worn on the head or as part of a helmet that has a small display optic in the front of one or each eye. Additional motion sensors are integrated in the HMD to record the motion and the position of the user in real-time and to adopt the 3D-graphics to this new user's position. In order to enable the interaction between the user and the 3D-objects the HMD-system has a 3D-maus or VR-gloves as user-input-device. HMD-approach represents a big leap forward in advancing the VR-technology. Just within the last couple of years, one could observe enormous advances in the area of HMD technology, which paved the way to a new era of VR and AR [4].

Due to the current high significance and to high potential of the HMD-technology this paper will be dedicated to the HMD-based VR approaches within the product life-cycle processes of industrial products.

3 ARCHITECHTURE OF HMD-SYSTEM

The virtual reality requires the integration of multiple factors: user interface, elements of the virtual world and the user's experiences. Interaction between these factors defines the experience of virtual reality [5]. For the set-up of a virtual reality environment based on HMD-system the following components (figure 3) are needed:

- Virtual content: The Virtual content, is also named as the virtual scene, contains the 3D-obejcts and their descriptions in terms of properties (e.g. shape, weight, color, texture, density, temperature). The interactions between the 3D-obejcts among each other and with the user can be implemented using a 3D-simulation-engine (so-called 3D-Engine). The simulation-engine provides standardized libraries and SDK (Software Development Kit) that allow the creation of rules as well as relationships that govern the 3D-obejcts within a VR-scene. Nowadays, the VR-scene can be created using commercial animation 3D-Engine (e.g. Unity Engine, Unreal Engine). Theses commercial engines are coming originally from the gaming and film sectors, however they have a large set of functions and interfaces that can be useful for the development and implementation of engineering VR-applications. Furthermore a large and high networked developer's community exits worldwide that drives the further development of the VR technology actively. The developers' community represents a valuable know-how resource to introduce the VR technology within the industrial and engineering processes effectively.

- Virtual presence: *Virtual presence* can be roughly divided into physical (sensory) and mental presence. It represents the feeling of actually being in an environment; this can either be a completely psychological state or achieved via some physical medium. Physical virtual presence is the basic characteristic of virtual reality and represents the user's body physically entering the medium. Mental virtual presence represents a state of trance: engagement, expectations, the feeling of being part of the virtual world [5]. To achieve a high virtual presence a high sensible tracking system is needed, which can act and re-act in real-time. For this, two types of sensory systems are needed: the first type is a sensory system to automatically measure the position and the orientation of the user in the real environment and map it in the virtual environment. Based on this measurement data, the VR-scene will be new computed and adapted to the new user's location and orientation in real-time. This process is named "framing". This first type of the sensory system is essential to create a VR environment and requires a high-performance computer with high-performance graphics adapter. To realize this first sensory type the HMD-system provides a Visual Base Stations (e.g. IR, cameras) to create a 360 degree virtual space. The base stations emit timed signals (e.g. infrared pulses) that are then picked up by the sensors (e.g. IR sensors) that integrated in the headset and in the user interaction's devices (e.g. controller) to determine the user's current location in the virtual space [6]. The second type of the sensory system allows a physical feedback from the VR-scene towards the user (e.g. haptic feedback, inertial feedback, sound feedback). This type of sensory system can be realized using e.g. VR-gloves, VR-suits, 3DOF-system that equipped with pressure, vibration sensors and actuators. The development of such kind of sensory systems is currently in its infancy and their integration in the VR-environment can considered as an enhancement.
- Interactivity: If virtual reality is to be realistic, it must respond to the user's actions; in other words, it must be interactive [5]. The ability of the user to affect computer-generated environments represents one form of interaction (e.g. manipulate the 3D-objects and navigate in the VR-scene). For this, multi-functional VR-Input devices can be used (e.g. VR-controller, VR-Flysick, VR-Glove). These systems are equiped with track pads, grip buttons, and a dual-stage trigger that can be configurated according to the user-interaction with the VR-scene [5].

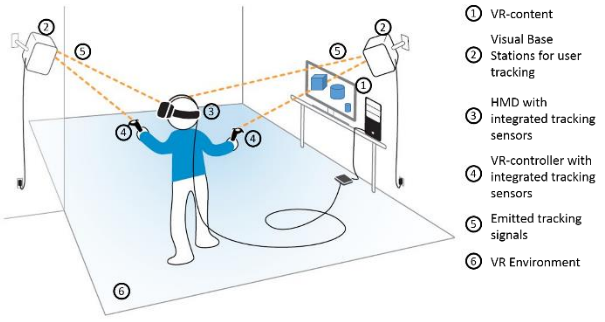


Figure 3 - Main components of HMD-system [based, 8].

4 IDUSTRIAL VR APPLICATIONS

In this chapter the results of two feasibility studies with the focus on "industrial application of the VR-technology" carried out at the faculty of Mechanical Engineering and Mechatronics at the Karlsruhe University of Applied Sciences will be presented and discussed. These feasibility studies were realized in cooperation with a german machine manufacturer, in order to analyse and to show the potential of the VR technology to enhance the product life-cycle processes. Furthermore, the main challenges when introducing such a new technology have to be identified. The feasibilty studies are an essay to analyse and clarify the following issues based on company-related use cases:

- Identification of appropriate processes along the product life-cycle for the VR-application.
- Implementation of VR-applications for the identified processes.
- Presentation and evaluation of the implemented VR-applications in the involved company's business units.
- Recommendation for the companies regarding the use of the VR-technology and the further activities in this area.

4.1 VR-based Mock-Up for Requirement-Management and Design-Review

4.1.1 *Motivation of the use case*

This use case deals with a company developing and producing special machines for automation technology (figure 4). The USP of this company is the design and the production of only customer specific solutions. This ensures the company a great competitive advantage, but at the same time leads to enormous challenges particularly with regard to requirement-management, design-review and -alignment. Misunderstanding and misinterpretation of customer requirements and wishes in the early product life-cycle phases cause elaborate and costly product changes. In order to avoid this situation, the company is actual forced to build up a 1:1 Physical Mock-Up (PMU) in the early design process phase, which can be used as basis for discussion and review meetings with the customers. As already known such PMU are time consuming and costly. In addition, the PMU cannot always be kept up to date during the whole design process.

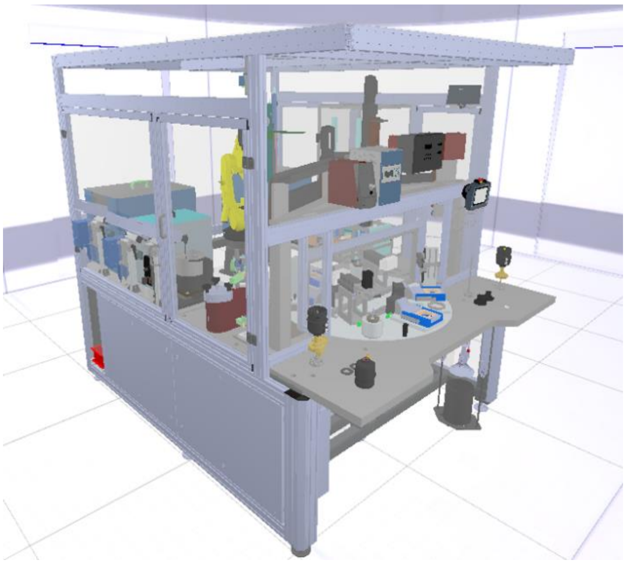


Figure 4 - The automation machine "assembly of valve system" used for the feasibility study.

4.1.2 Approach for VR-based Mock-up

In order to improve the requirement-management and the design-review process and to evite time consuming and costly PMU, a VR-based machine Mock-Up has been developed and implemented in this project (figure 5). The VR-application is containing a VR-library with standardized machine parts and components as 3D-objects. These 3D-objects were created in the native CAD-system during the design process and then imported to the 3D-Engine (in this case Unity). To create a machine VR-Mock-Up, the 3D-data of the specific machine parts had to be transferred from the native CAD-system to the 3D-Engine and then can be completed with the standardized parts from the VR-library to create the whole machine as 3D-assembly. After the completion of the 3D-Model of the machine, the physical properties of the individual parts (e.g. material, texture, weight) were determined. Afterwards the relationships among the machine parts ware defined using either standardized kinematics objects (e.g. pin-joint, pivot-joint, Ball and socket joint) form the kinematics-library of the 3D-Engine, or special implemented kinematics objects (e.g. specific motion of robotic arm) with the SDK of the 3D-Engine. At the end the VR-functions were implemented to enable the interactivity between the user and the machine in the virtual environment. To serve the requirement-management process the following interaction-functions were created:

- User-Teleport around the machine.
- Use and interaction with the MMI-Panel.
- Machine operation.
- Machine equip and retooling.
- Change of color and texture of the machine parts.
- Add notice for changes.

For the design-review process in addition to the functions from the requirement-management the following interaction-functions were implemented:

- Assemble and disassemble of machine parts.
- Hide and show of machine parts.
- Display of the metadata (e.g. material, weight, part-id) for the different machine parts.
- Add of static and dynamic sections.

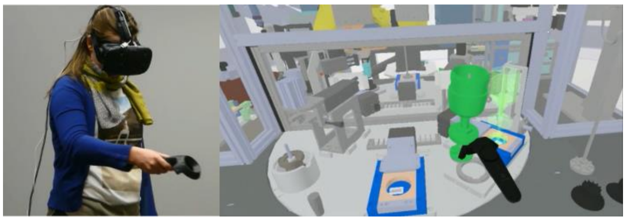


Figure 5 - The Implemented VR-based Mock-Up.

4.1.3 Evaluation of the VR-Application and VR-Technology

In order to evaluate the implemented VR-application, a pilot project was initiated with the involved business units (sales engineering and product design) and the customer (Figure 6). Within the evaluation three main aspects were taken into account:

1. Enhancement of the considered process:

- VR-experience: Both the business units of the company and the customer were impressed with the reality-close VR-environment of the machine regarding the performance of the VR-application and equipment. Especially the uncomfortable feeling known from the old VR technology could not be sensed.
- Interactivity: With the implemented interactivity VR-functions the basic tasks of the processes could be achieved without major restrictions. The main limitations from the perspective of the users are the messing of the feedback from the VR-environment towards the user (e.g. haptic feedback) and the coupling with the native CAD-system to allow an integrated modification of the CAD-models in VR-environment.

2. Implementation of the VR-Application:

- Import of the 3D-CAD Data into the 3D-Engine: Due to the missing of appropriate interfaces between the CAD-systems and the 3D-Engines, the design CAD-data including the design information (e.g. material, structure, kinematics) for the machine created during the design phase cannot be transferred to the 3D-Engines directly and completely. The established 3D-Engines are designed to implement gaming applications, which can only handle with low-content graphic objects (surface geometry data, bounding geometry). On contrary to this, the design CAD-data are high-content 3D-objects with high data volume that cannot be processed by the current 3D-Engines. Thus, bevor loading the design CAD-data in the 3D-Engines, the CAD-data must be treated in third party 3D graphic program, in order to reduce the data content. This procedure leads to loss of important information that is needed for the achievement of diverse engineering tasks. The lost data have to be created once again in the 3D-Engines manually. The current approach causes double work and is time consuming.

- Implementation of the user interaction functions: The common 3D-Engines already have a considerable stock of user interaction functions implemented for gaming applications. These functions are highly advanced and can serve as valuable source and template to develop VR-interaction functions for engineering processes.

3. Maintenance and further development of the VR-Application:

- Change of CAD-data: Small changes of the design CAD-data can be handled in the 3D-Engine easily. Big changes of the design 3D-data usually cause a high modification effort of VR-scene up to the setup of a new VR-environment.
- Change and introduction of new interaction functions: Normally the interaction functions are designed modularly. Hence, the already implemented functions can be re-used and modified easily without a high effort.

4.2 VR-based FMEA (Failure Mode and Effects Analysis)

4.2.1 Motivation of the use case

The FMEA is a design tool used to systematically analyze postulated component failures and identify the resultant effects on product operations. The analysis is sometimes characterized as consisting of two sub-analyses, the first being the failure modes and effects analysis (FMEA), and the second, the criticality analysis (CA). The usefulness of the FMEA as a design tool and in the decision-making process is dependent on the effectiveness and timeliness with which design problems are identified [9]. Timeliness is probably the most important consideration. While the FMEA identifies all part failure modes, its primary benefit is the early identification of all critical and subsystem or system failure modes so they can be eliminated or minimized through design modification at the earliest point in the development effort; therefore, the FMEA should be performed at the system level as soon as preliminary design information is available [9].

The same machine manufacturer from the first use case (chapter 4.1) has to study the problems that might arise due to mal-operations of the machine by the end-customers. The manufactured machine is characterized as Man-Robot-Collaboration system (MRC), where the machine operator and the robot are working together at the same time and in same working area. Therefore, the risk of the accident, that might occur when machine-operator enters in the robot hazard zone unintentionally, must be analyzed and determined accurately. Based on this risk analysis the machine architecture, dimension and functions will be fixed. Currently the machine manufacturer is conducting this analysis based on experience done with similar products and processes and based on customer feedback. For the design of new machine types are these information not reliable to deliver an accurate risk analysis. As consequence, the design of over-engineered and costly machine concepts and in the worst case the overlook of risk sources.

In order to help the machine manufacture to achieve a thorough risk analysis a VR-based FMEA application was developed in this project.

4.2.2 Approach for VR-based FMEA and CA

The developed VR-based application for FMEA and CA allows to setup of a VR-scene including the 3D-data of the whole machine in the early product design phases. The VR-scene contains kinematics objects and interaction function that enable a reality-close machine operation in the VR-environment. By the use of the implemented functions the machine operator can carry out the following process activities:

- Assemble a complete product: In this use case the assembly of valve system.
- Equip the machine with the necessary tools.
- Change of the product type and retooling of the machine.

In order to identify the hazard zone of the machine, the machine room was divided into different area with different severity degree (s). The area with the highest degree corresponds to the most dangerous area: collusion between operator and robot leads to operator injury. The area with lowest degree corresponds to the hazard-free area: the area is not in the working range of the robot. These areas were implemented within the VR-scene as invisible 3D-collider objects with a collision counter (C). Whenever the machine operator touches or enters an area, the area’s counter increases by one. The information about the number of the operator’s collisions with the different areas can be exported as excel-file to analyses and assess the working-mode of the operator and the criticality of the different machine’s areas.

To evaluate the implemented VR-based FMEA and CA application, normal machine operators without any awareness of the existing of the 3D-collider objects were invited to work with the machine in the VR-environment according to a given work instruction. At the end the collected data of all involved operators were exported as excel-file (figure 6) and based on it the criticality factor K (K = S * C) for the diverse areas was calculated (figure 7). Using the value of K, design-countermeasures for the affected areas were defined to ensure the machine safety:

- Areas with a high value of K (very critical): an additional safety wall was designed to prevent operator entry in this area.
- Areas with a medium value of K (critical): an additional safety light barrier was added to the machine, in order to stop the robot whenever the operator enters this critical area.
- Areas with a low value of K (normal): a safety indication was stuck to the machine.

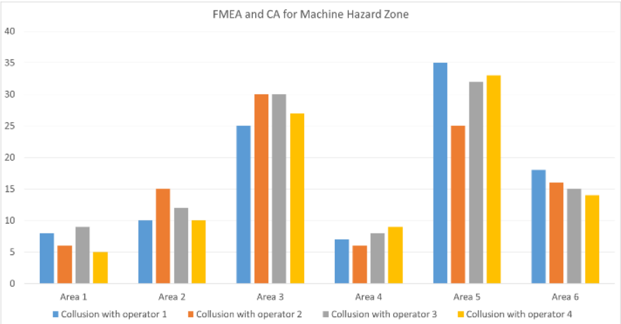


Figure 6 - Output of the FMEA for machine Hazard zone using the VR-application

	Severity degree (S)	Collision counter (C)	Criticality factor K (K = S * C)	Criticality
Area 1	3	28	84	normal
Area 2	2	47	94	normal
Area 3	4	112	448	very critical
Area 4	5	30	150	critical
Area 5	0	125	0	noncritical
Area 6	3	63	189	critical

Figure 7 - Determination of the criticality of the different machine zones.

4.2.3 Evaluation of the VR-Application and VR-Technology

The implemented VR-Application has provided a good basis to investigate a VR-based and reality-close FMEA and CA. Especially in the case of developing of new innovative products, where user's experience and feedback don't exist, the VR-based FMEA and CA approach has a high potential to enable a reliable FMEA and CA.

The implementation of this use case has shown the same difficulties arose in the first use case (chapter 4.1.3). The CAD-data transfer from the native CAD-systems to the 3D-Engine is time consuming and leads to the loss of the process-needed engineering information.

5 CONCLUSIONS

The carried out feasibility studies have demonstrated the potential of the VR-technology as digital transformation's technology to enhance and support the product life-cycle processes. Using the VR technology the reliability, safety and quality engineering can be performed timeliness, efficiently, and reliable. Furthermore, expensive and elaborate PMU can be replaced by cost-, energy-, and material-saving VR-based applications in the early product life-cycle phases. Though, the exhaustion of the potential of this technology within the engineering processes depends on the developing of appropriate solutions for CAD-data transformation including all the needed engineering information (e.g. material, weight, texture, kinematics) from CAD-systems to 3D-Engines. An integrated CAD-data modification in the VR-scene and re-feeding to the native CAD-systems should be also enabled. Without these enhancements, the HMD-based VR-technology would not see a real breakthrough in the industrial application.

6 REFERENCES

[1] Steinicke, F., 2016, Being Really Virtual, Add, Springer.

[2] Ma, M., Jain, L., Anderson, P., 2014, Virtual, Augmented Reality and Serious Games for Healthcare, Springer.

[3] Jeschke, S., Kobbelt, L., Dröge, A., 2014, Exploring Virtuality, Springer.

[4] Mihelj, M., Novak, D., Begus, S., 2014, Virtual Reality Technology and Applications, Springer.

[5] Liu, D., Dede, C., Huang, R., Richards, J., 2017, Virtual, Augmented, and Mixed Realities in Education, Springer.

[6] https://www.statista.com/chart/4602/virtual-and-augmented-reality-software-revenue/, visited on 14th of August 2018.

[7] https://de.statista.com/infografik/7430/virtual-reality-ist-vor-allem-beim-gaming-attraktiv/, visited on 14th of August 2018.

[8] https://www.techradar.com, visited on 14th of August 2018

[9] Rausand, M., Hoylan, A., 2004, System Reliability Theory: Models, Statistical Methods, and Applications.

7 BIOGRAPHY

Fahmi Bellalouna obtained his PhD degree in computer integrated Mechanical Engineering from the Ruhr University Bochum. In 2015 he was appointed Professor in Mechanical Engineering and Mechatronics at Karlsruhe University of Applied Sciences.

Industry 4.0 and its Future Staff. Matching Millennials' Perceptions of a Perfect Job with the Requirements of Digitalization

A.K. Schaar[1], A. Calero Valdez[1], T. Hamann[1], M. Ziefle[1]
[1]Human-Computer Interaction Center, RWTH Aachen University, Germany

Abstract
Against the background of a digitalizing industry (industry 4.0) the question of how companies and future staff are prepared for these changes is becoming increasingly important. That is why the focus of this empirical-experimental study was the investigation of the job choice of Millennials and their requirements for the ideal workplace. The perspective of high potentials was investigated by two focus groups (n=8) which were carried out prior to the main study, a quantitative conjoint study (n = 165). By applying conjoint analysis, the study empirically investigated the relationship between the attributes *job tasks, flexibility, family-friendliness, and salary*.

Keywords
Digitalization, Industry 4.0., Work place design, Millennials,

1 CHALLENGES OF FUTURE WORK

Production systems are not what they used to be. The 21st century will challenge enterprises and manufacturing companies with novel generations of technologies, intelligent data services, and computer based technologies [1]. In order to keep up with the competition on global markets and ensure long-term success, the companies need to adapt to shorter delivery times, increasing product variability, and high market volatility. This adaptivity and resilience enable enterprises to sensitively and timely react to continuous and unexpected changes [2]. In addition, companies also face aggravating challenges that are based on social transformation processes: Triggered by demographic change, work environments are confronted with a decreasing number of employees, and a shortage of skilled laborers. The increasing portion of aged workers have a large amount of process knowledge and work experience, but lack the skills to be able to use novel computer technology. In today's knowledge society, companies heavily rely on their employees' knowledge and skills, but also on their flexibility to tolerate fast changes, to take on responsibilities for work-related decisions, and on a high resilience towards disruptive processes. They therefore need to successfully attract, recruit and keep qualified employees with high potentials. To do so, it is essential to know the prevailing needs and desires of their (future) employees [3]. That is why the next generation of workers, the so-called Millennials, are currently the focus of interest of employers as well as research. It is essential to understand how they will experience the increasing digitalization of work and if their demands and desires for fulfilling work differ compared to previous generations [4]. Thus, the overall challenge is to balance changing demands of workers, industry 4.0, and market pressure (high volatility, high disruptiveness, shortage in workers).

1.1 Industry 4.0 and the digitalization of work

The concept of industry 4.0 describes the digitalization of industry, the technological change of production and the change from a computer-centered world to the so-called Internet of things. In the Internet of things, processes, devices, objects and environments are connected to each other via the Internet and controlled by software [5]. This transformation will fundamentally change the way people will work in the future. From today onwards, work processes will be more networked, digital and flexible than ever before. For the working population and future employees, this means that established working methods are shifting and that skill requirements are subject to change. Overall, work processes are becoming easier and more efficient [6], as are complex technical processes, which industry 4.0 should also make safer. Because Industry 4.0 triggers processes of change in industrial production that are not yet fully foreseeable, it is essential to actively shape these processes and to take all involved actors into account.

1.2 A new generation of workers

In addition to the technical changes within industry 4.0, the work itself and the challenges for employees are also affected by social transformation processes: on one hand demographic change causes a shrinking number of employees, the loss of person-immanent knowledge, and a threatening lack of skilled workers [7]. On the other hand, a new generation of workers, the so-called Millennials, is entering the labor market [8]. Both aspects are addressed in the following sections.

1.2.1 Demographic change

One of the most discussed megatrends in western industry nations with far-reaching effects on the working world is demographic change [9]. The associated changes lead to a decline in the size of Germany's population and to an ageing society. This will strongly impact the composition and the availability of the working population. Today, the 40 to 60-year-olds make up a large proportion of the working population. If this age group is eliminated, the 1970s and 1980s will follow [10]. As a result, the number of employees will fall from 40.6 million to 39.2 million by 2030 and the supply of skilled workers will significantly decrease [11]. If this happens, companies must recruit and retain highly qualified employees from the shrinking number of available employees which will reinforce the hunt for high potentials. This fact leads to an increasing interest in attracting future employees. At present, it is the generation of people born in the 1980s and 1990s, the so-called Millennials, who are increasingly put in the focus of interest. What distinguishes this new generation of workers from their predecessors is outlined in the following section.

1.2.2 A new generation of staff

Since the future of work will be marked by a struggle for high potentials, the next generation of employees will become increasingly important over the next years. That is why Millennials and their demands for, and ideas on, the perfect job are in the focus of interest at this moment. Different studies revealed the existence of generation specific characteristics: Research shows that this generation is characterized by a high degree of individualism, many choices and volatility [12]. Klaffke and Parment (2011) e.g., describe that they are driven by self-realization, which is why frequent changes of employer, flexible working time models, and as much freedom and autonomy at work as possible [13] are prerequisites for achieving their goals. Their strong self-confidence stems from demographic developments and the (imminent) shortage of skilled workers [12]. At the top of their "wish list" is a good work-life balance which means that they expect their employer to provide them with a life-stage oriented personnel policy. Such a life-phase oriented personnel policy is at the same time demographically oriented and ensures a sustainable employability and the compatibility of private and professional life [14]. Also important in the context of the ideal workplace are flexible working time-models, a fair and equitable remuneration, feedback structures, comprehensive education and training opportunities, chances of advancement, challenging tasks, and autonomy. Added to this are aspects like a pleasant working atmosphere, the use of modern ICT, job security [15, 16], fun at work, the company's reputation, the location and an international environment [17,18]. In the context of the discussions on the job preferences of Millennials, it is important to remember that despite the label "generation", this group is not necessarily homogenous. Rather, existing studies show that there are differentiating attributes within this generation that must be considered [e.g. 18].

2 METHODOLOGICAL FRAMEWORK

In this paper we address the demands of Millennials on their (future) job. For this purpose, a mixed-method approach including qualitative and quantitative methods was developed (see figure 1).

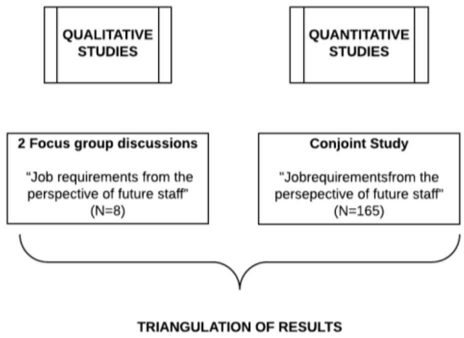


Figure 1 - Overview implemented studies.

As qualitative measure, two focus group discussions were conducted (section 3). For quantitative analysis a conjoint analyses (CA) was set up (section 4). In the following the results of both studies are presented separately. A triangulation of the findings is presented at the end of the paper and is transferred into a fundamental discussion of the results .

3 IDENTIFICATION OF JOB REQUIREMENTS

A qualitative pre-study was carried out to identify the critical job requirements of Millennials.

3.1 Method and sample

In total we ran two focus groups with five participants each who discussed the relevant job characteristics of future work settings. Participants were asked to express their expectations of a satisfying and fulfilling workplace and skills they will provide in return. The discussion was guided by several open questions that addressed job requirements, as e.g., "How important is work-life-balance to you?", "How much flexibility at work do you want?", "How important are career and development opportunities?". We invited both Millennials (young professionals as well as students) with a technical and non-technical background. The non-technical group consisted of 5 people (n=2 men, n=3 women), from the fields of business administration, technical communication, communication science, and political science. The technical group (n=4 men, n=1 women) were from the field of electrical and mechanical engineering.

3.2 Results of the focus group discussions

The results of focus groups are presented using 8 categories that were derived from the discussions:

3.2.1 *Self-confident Millennials*

In general, the Millennials within the focus group came across as being self-confident. They were confident about their future and expect to cope well with future work challenges. At the same time, they had concrete ideas of where they see themselves in the future. They were actively seeking for their dream job, thereby being unwilling to make compromises. The participants characterized this as a typical attitude of their generation:

"[...] *"But I can well imagine is that our generation really pays a lot of attention to that it is right from the start, that it suits us, and that we don't take any job that doesn't really correspond to what we imagine, but (maybe) has the only advantage that you make a lot of money there".*

That is why it is important for them to be "*their own personnel manager"* and to look for a job that suits them and where "*the chemistry is right*".

3.2.2 *Teamwork*

Participants reported that a familiar and social work environment is essential for a good workplace. Only few wished to clearly separate "work friends" from private friends. However, most respondents could well imagine that the private sector and working life would merge. They envisioned to be more successful and happier if the center of life is transferred to work, with friends at work. Participants with a technical background specifically wish flat hierarchies with a clear regulation of responsibilities and competencies within the team.

"[...] *but when it comes to the technical, it was clear that there is one person who sets the tone"*.

According to the participants, team members should have different competencies, so everybody can rely on colleagues with a specific role and responsibility. Each team member should have a unique competence, so everyone is important for the group.

3.2.3 *Working atmosphere*

From the point of view of the interviewed personas, the working atmosphere includes both social and environmental factors: Regarding social aspects, the participants wish for respect, trust and freedom of the superiors, a culture that handles mistakes and errors with care, as well as friendly and competent colleagues. Regarding environmental aspects, the participants expect a kitchen for the employees with free coffee and water as well as modern (daylight) offices. In addition, good and modern technical equipment, office furniture, short waiting times for orders, and laboratory facilities for withdrawal and tinkering were particularly important to people with a technical background. The non-technical group specifically valued the possibility for occasional meetings, e.g. lunch, communication, and exchange.

3.2.4 *Job description*

"Ideal" tasks were described with attributes like interesting, challenging, diversified, and demanding. Both the group with technical and non-technical background were looking for tasks in which they can get involved, with the motivation to learn new things and to be responsible. The participants were aware that there will also be monotonous, and bothersome tasks,

"[...] *but it is always the question of how this outweighs itself and if the challenge prevails, that is enough".*

For the participants, interesting tasks are fun as well and they stressed that they want to enjoy doing them. Additionally, they reported that they are keen to learn new things, as personal development occurs only through interesting and challenging tasks with constant "input". They wish to develop personally and professionally, not only in order to enjoy their work, but also to stay attractive for other companies in case of change of employment.

3.2.5 *Flexibility*

The aspect of flexibility was intensively discussed in both focus groups but was interpreted differently. Among other things, respondents equate flexibility with freely configurable working hours. The participants believed that this is the most effective and productive way to work while still keeping their private lives organized. Flexibility also means being able to come later or leave earlier due to external appointments without this being a problem for supervisors and the company. Moreover, flexibility also means being allowed to work from different locations, using mobile communication technology (skype, email, social media) to stay connected with colleagues.

3.2.6 *Reconciling work and family*

Both focus groups also discussed the family-friendliness of a company and its openness towards the balance of leisure and work in the context of flexibility. In the non-technical focus group, an understanding for private life and family obligations was explicitly desired. For example, it should not be a problem if a male employee wants to take parental leave. The women perceived it as important that they themselves and their partners work in a family-friendly company. The majority of the interviewees stated that they would only choose a company with a family-friendly culture, a compatibility of family and career, and a high work-life balance.

3.2.7 *Loyalty*

In general, the opinions were clear: As long as the employer has something to offer and responds to the employees' needs, there is no reason to change jobs. However, if work gets boring or expectations are not fulfilled, people do not feel obliged to be loyal to the company.

3.2.8 *Salary*

According to the interviewees, salary plays a comparably minor role:

"[...] It's one of many aspects. Good pay is always better, but if it's just good pay that makes a job and somehow the

job itself is stupid or the colleagues are stupid, then I wouldn't take it anyway."

3.3 Key findings of focus group discussions

Major outcomes of the qualitative studies revealed that Millennials attach great importance to "soft" factors of work which are equally, if not more, important as salary. A good flexibility at work, interesting and challenging tasks, and the compatibility of career and family are essential job criteria for the upcoming workforce. At the same time, for the Millennials, salary is only one of many aspects to take into account when deciding in favor of, or against, a job. The importance of a good work-life balance and compatibility does not depend on the professional background. For Millennials with a technical background, development opportunities on the job and demanding tasks were key. For Millennials with a non-technical background, this was the case for the ability and possibility to balance work and family life.

4 JOB DECISION SIMULATION

Based on the results of the focus groups, the following attributes were identified as important and interesting for the subsequent quantitative decision simulation: *(1) tasks* (levels: routine, easy and challenging tasks), *(2) flexibility* (levels: fixed working times, flexitime, flexible work), (*3) family friendliness* (levels: statutory minimum, co-financing, company kindergarten) and (4) *salary* (300€ less per month, average salary and 300€ more per month). All attributes need to be examined in combination with each other, as in reality such decisions reflect a weighing between and across different levels of job-relevant factors.

4.1 Method

To do so, we used the Conjoint Analysis (CA), a quantitative empirical research method which enables the analysis and simulation of consumers' decisions. A so-called Choice-Based Conjoint (CBC) analysis was used, in which participants make preference judgments in the form of selection decisions. The participants were asked to choose the most attractive job offer from several alternatives. This resembles the decisions job-seekers have to make in every-day life [19]. By these trade-offs between different attribute levels can be determined and thus those attributes, which finally influence job decisions most.

4.1.1 Questionnaire design

The questionnaire for the online survey was created using SSI Web Software (Sawtooth Software 2012). The *first part* surveyed demographic data (age, gender, level of education, children, current occupation), job-related information and an assignment of the current job (technical vs. non-technical). The *second and the third part* contained two consecutive choice tasks. The first CBC only varied fictive job alternatives out of three attributes – job task types, flexibility at work and family friendliness of the company. In the second CBC salary was added as a fourth attribute to the decision simulation. This was done because we wanted to quantify the relative importance of job characteristics beyond the salary (CBC 1) for the job decision. Then, in CBC2 salary was added to identify how the decision changes when payments come into play. Each participant finished both choice tasks, with a fixed order (CBC1 before CBC2). At the beginning of the questionnaire a scenario was presented that asked the respondents to imagine that they are in the application phase after finishing their studies and should decide on an employer. The job offers differed in regarding tasks, flexibility and family friendliness (see Figure 2) – those characteristics which had been identified in the focus groups. Each attribute was varied within different levels

Attributes of CBC 1			
Attribute	Specification		
Tasks	routine Tasks	easy tasks	challenging tasks
Flexibility	fixed working times	flexitime	flexibile work
Family friendliness	statutory minimum	co-financing	company kindergarten
Additional attributes of CBC 2			
Salary	less than 300€	avarage salary	more than 300€

Figure 2 - Overview attributes of CBC 1 & 2.

4.2 The sample

Overall, 165 participants took part (41.8% male, 58.2% female). The average age was 24.53, ranging from 17 to 33 years of age (SD = 3.12). 62.4% of the sample already successfully finished their studies. Most of the participants (63%) reported that they did not have any children, but 32,1% stated that they want to become a parent in the next five years. Regarding job orientation and field of training, 52.7% of the participants directed their job searches to the technical sector. Only 2.4% of the respondents reported to have no work experience at all, while 98% had at least some job experience (internships, training, student assistance activities).

4.3 Results

This section depicts the findings regarding relative importance and part worth utilities of the job choice.

4.3.1 Relative importance of attributes

In order to identify the main factors influencing the choice of workplace, the significance of the four attributes was examined with the help of HB analysis. Figure 3 shows the relative importance of the attributes. For CBC1 it is clear that with 44.04%, the nature of the job tasks has the highest importance in the presented scenario, followed by the *flexibility* (30.95%) and the *family friendliness* of the company

(25.01%). In line with recent research, the most important job characteristic for Millennials is the type of work tasks, the least important one the extent of a company's family friendliness, and flexibility ranks in between these two characteristics. When the salary was added as a supplementary job characteristics (CBC2), things changed. Figure 3 shows that salary then became the most important characteristic (39.16%). Nevertheless, the relevance of the other job characteristics remained unchanged.

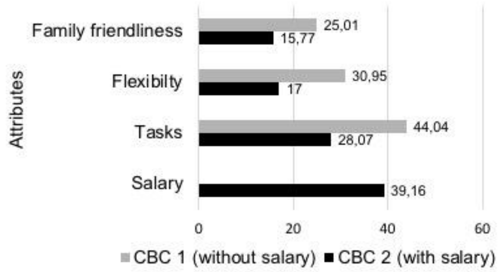


Figure 3 - Relative Importance of attributes.

In second place are the *tasks* (28.07%), followed by *flexibility* (17%) and *family friendliness* (15.77%).

4.3.2 Part worth utility analysis

Now, part worth utilities are analyzed, i.e. the relative impact of the single levels within each attribute. Figure 4 shows the average part utility values (zero-centered differences) of the two CBC analyses across all attributes. Note that the sum of utility scores within each attribute equals zero. This type of scaling makes it possible to compare differences between attribute values. High positive utility values contribute to the scenario selection, while negative scores contribute to the choice of an alternative. Independently of salary as a criterion carrying weight, challenging *tasks* are most attractive (34/44), while routine tasks, in contrast, receive clear negative scores (-53/-46), with easy tasks reaching at least slightly positive scores. With respect to *flexibility*, fixed working times are homogeneously assessed as negative (-31/-36), again, independently of the prevailing salary levels. Flexitime (15/25) and flexible work (16/11.6), in contrast, both receive positive scores. Interestingly, flexitime is still seen more positively compared to completely flexible work, even though the latter has more leeway. When it comes to the company's *family friendliness*, the "no go" is the statutory minimum (-29%/-35%), independently of how much money people earn. Whenever the company co-finances family issues, this leads to positive scores (7%/14%). But the highest support was found when a company kindergarten is available. However, whereas an average salary is still accepted (14.88), a salary of 300 € less than the average is already clearly rejected (-85.68). (22.8/21.3). Finally, for the attribute *salary*, naturally, earning 300 € per month is most attractive (70.8). Thus, in conclusion, the introduction of salary shows clear effects in the participants' decision modelling: Salary becomes the most important criterion, but the order of the assessment of the other attributes stays the same, with work content being most important (task type) and family friendliness least.

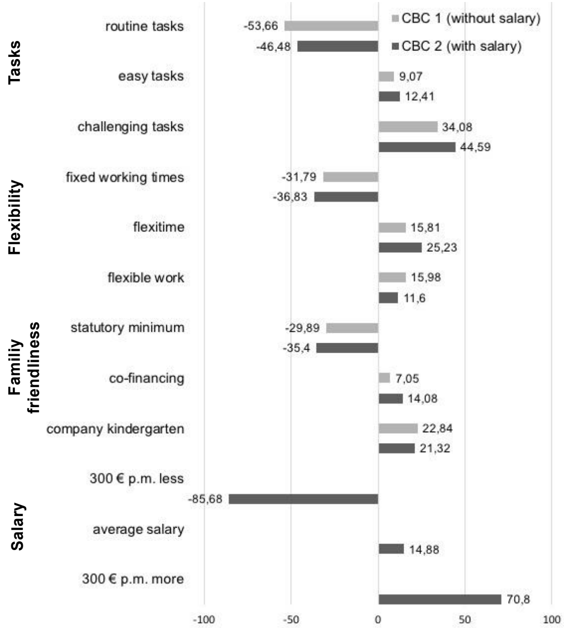


Figure 4 - Part worth utilities.

5 FINAL CONCLUSION

In this paper, it was examined what standards Millennials apply when choosing their employer. The job decisions were explored against the background of a fundamental change in social values and the digitalization of work. Thereby, major key job criteria were used – work content and tasks, organizational flexibility, family friendliness, and salary. The majority of the Millennials indeed preferred work tasks that are challenging and responsible, and thus reject routine tasks. Since the working flexibility of future employees is also a decisive feature, the possibility of flexitime or even completely flexible work times should be offered. When it comes to the importance of family friendliness, the respondents–even though they attach great importance to family friendliness when directly asked (focus group discussion)– "only" rank this job feature in third place. Whenever salary comes into play and serves as additional job decision feature, the old picture of earlier generations is retrieved: Salary is the most important decision criterion in favour of or against a job. Overall, the study shows that choosing a job is a complex process, especially for Millennials, who are starting their career with great expectations. The insights of and about Millennials might help employers, who might not yet be sufficiently informed about the high job expectations of a generation that has the choice, to adjust their company's attractiveness.

6 ACKNOWLEDGMENTS

The German Research Foundation (DFG) funded this project within the Cluster of Excellence

„Integrative Production Technology for High-Wage Countries" (EXC 128).

7 REFERENCES

[1] Schuh, G. and Gottschalk, S. (2008) Production engineering for self-organizing complex systems. *Production Engineering*, **2**, 431–435.

[2] Wiendahl, H.-P., ElMaraghy, H.A., Nyhuis, P., Zäh, M.F., Wiendahl, H.-H., Duffie, N. et al. (2007) Changeable Manufacturing- Classifica-tion, Design and Operation. *CIRP Annals*, Elsevier. **56**, 783–809.

[3] Kowske, B.J., Rasch, R. and Wiley, J. (2010) Millennials' (Lack of) Attitude Problem: An Empirical Examination of Generational Effects on Work Attitudes. *Journal of Busi-ness and Psychology*, Springer US. **25**, 265–79.

[4] Wey Smola, K. and Sutton, C.D. (2002) Generational differences: revisiting genera-tional work values for the new millennium. *Journal of Organizational Behavior*, **23**, 363–82.

[5] Foerster-Metz, U., Marquardt, K., Golowko, N., Kompalla, A. and Hell, C. (2018) Digital Transformation and its Implications on Organizational Behavior, Journal of EU Research in Business.

[6] BMBF. Zukunftsbild "Industrie 4.0" [Internet].

[7] Wolf, M., Kleindienst, M., Ramsauer, C., Zierler, C. and Winter, E. (2018) Current and future industrial challenges: Demographic change and measures for elderly workers in Industry 4.0. Annals of the Faculty of Engineering Hunedoara-International Journal of Enineering 16(1).

[8] Andrade, M. S., & Westover, J. H. (2018) Generational differences in work quality characteristics and job satisfaction. Evidence-based HRM: a Global Forum for Empirical Scholarship. Emerald Publishing Limited.

[9] Bloom, D. and Canning, D. (2004 Oct) Global Demographic Change: Dimensions and Economic Significance. Cambridge, MA. https://doi.org/10.3386/w10817

[10] Statistisches Bundesamt. (2015) Bevölkerung Deutschlands bis 2060.13. Koordinierte Be-völkerungsvorausberechnung. Wiesbaden.

[11] Deissinger, T., and Breuing, K. (2014). Recruitment of skilled employees and work-force development in Germany: Practices, challenges and strategies for the future. *Workforce Development,* 281-301, Springer.

[12] Arakawa, D. (2013)How will leadership style of generation Y impact the traditional standards of law enforcement by 2020. *Command College Futures Project*, 1-15.

[13] Cennamo, L. and Gardner, D. (2011) Generational differences in work values, outcomes and person-organisation values fit. *Engineering Management Review*, **39**, 24–36.

[14] Klammer, U. (2010). The "Rush Hour" of life: insecurities and strains in early life phases as a challenge for a life course-oriented, sustainable ocial policy. In *A Young Generation Under Pressure?* (pp. 155-166). Springer.

[15] Wong, M., Gardiner, E., Lang, W. and Coulon, L. (2008) Generational differences in personality and motivation. Macky K, editor. *Journal of Managerial Psychology*, Emerald Group Publishing Limited. **23**, 878–90.

[16] Guillot-Soulez, C. and Soulez, S. (2014) On the heterogeneity of Generation Y job preferences. *Employee Relations*, **36**, 319–32.

[17] Kultalahti, S. and Viitala, R. (2015) Generation Y – challenging clients for HRM? *Journal of Managerial Psychology*, Emerald Group Publishing Limited, **30**, 101–114.

[18] Acar, A. B. (2014). Do intrinsic and extrinsic motivation factors differ for Generation X and Generation Y. *International Journal of Business and Social Science*, *5*(5), 12-20.

[19] Orme, B.K. (2010) Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research. Res. Publ.

8 BIOGRAPHY

Anne Kathrin Schaar is a Postdoc researcher at the Chair for Communication Science, RWTH Aachen University. Her research focuses on technology acceptance in innovative technologies.

André Calero Valdez is a Post-doc researcher at the Chair for Communication Science RWTH Aachen University. His research interests are directed to understanding data science, the interface of algorithms and people in different areas.

Tatjana Hamann finished her studies in technical communi-cation and language and communication studies at RWTH Aachen University. Currently she works as a usability engineer in industry.

Martina Ziefle is Professor of Communication Science and head of the HCIC of RWTH Aachen University. Her research focuses on the interface between user, computer, media and techno-logy, taking different usage contexts and user requirements into account.

Need and Solution to Transform the Manufacturing Industry in the Age of Industry 4.0 – A Capability Maturity Index Approach

G Gudergan[1], V Stich[1]

[1]Institute for Industrial Management (FIR) at RWTH Aachen University, Germany

Abstract

Digitalization is changing the industrial landscape in a way we did not anticipate. The manufacturing industries worldwide are working to develop strategies and concepts for what is labelled with different terms such as the Industrial Internet of Things in the USA or Industrie 4.0 in Germany. Many industrialized economies are driven by the production sector and this sector needs specific approaches and instruments to take up other than those approaches we know from start-ups and ventures coming from Silicon Valley and other places. In this paper, we demonstrate an appropriate approach to transform producing companies in a systematic and evolutionary approach. In particular, the objective of this paper is to provide results from two initiatives, which conceptually build upon each other and are of particular relevance for the production industry. First, we present a global survey on the state of implementation and the future perspectives of the concept Industrie 4.0 from 2016. Findings from this study have forced parts of the German industry to heavily invest into a common approach to accelerate change towards Industry 4.0 in order to stay competitive in worldwide economy. This approach is presented in a second part.

Keywords

Digital transformation, Industrie 4.0, maturity model, manufacturing industry

1 INTRODUCTION

Concept and vision of Industrie 4.0 is often defined as “real-time, high data volume, multilateral communication and interconnectedness between cyber-physical systems and people”, in order to realize self-optimizing business processes [3]. This depiction of Industrie 4.0 concentrates primarily on a technological understanding with the objective that manufacturing companies achieve a competitive advantage. However, the fundamental economic lever of Industrie 4.0 lies in stimulating business processes through necessary decisions and real-time adaptations [2].

Combined with adequate organizational conditions, companies are able to react faster to growing market dynamics, to develop new products more quickly and precisely [4; 9].

Many results presented in this paper were gained within the context of a conceptual reference framework developed by German Government for Industry 4.0. The initial study is from 2016 and polled 433 industrial manufacturing executives in five regions – China, France, German speaking countries, the United Kingdom and the United States. We applied a specific capability maturity model for our analyses. The results provide an understanding of industry preparedness for Industrie 4.0. The huge potential of Industrie 4.0 is clearly pointed out but there are significant differences between countries. These differences in turn initiated a significant investment of resources by a consortium driven by the German industry and the German academy of science and engineering ACATECH. The later revealed a reference model and open standard and was successfully applied to transform manufacturing companies towards Industry 4.0 [3].

2 BACKGROUND AND MOTIVATION: GLOBAL BENCHMARK OF INDUSTRY 4.0 MATURITY

In 2015 and 2016, a comprehensive global research study assessed industry attitudes towards Industrie 4.0 in the aerospace, automotive, electronics, machinery and process industries in China, France, German speaking countries, the United Kingdom, and the United States. There was a focus defined on asset efficiency because this is considered an important performance indicator for the success of all kind of manufacturing innovation programmes including Industrie 4.0 programs. The study polled 433 executives through an online survey as well as telephone interviews. FIR at RWTH Aachen and the independent research firm Vanson Bourne conducted the study for the leading information systems provider Infosys [4, 5].

For the purpose of analysis, enterprises were categorized as “Early Adopters” or “Followers”. The actual status from 2015 and the aspiration for 2020 was analysed by asset efficiency levers, industry and production type, and country. Respondents were asked to outline their current maturity levels on these levers and their target for 2020 on a four point scale. The research used the Industrie 4.0 framework, conceptualized by the German government and developed by industry leaders.

In 2015, the vast majority (85 percent) of companies were aware of the high potential in implementing Industrie 4.0, Only 15 percent of enterprises surveyed had already implemented dedicated strategies for asset efficiency. An additional 39 percent had partially implemented these strategies.

Nearly half of the respondents surveyed (48 percent) wanted to implement Industrie 4.0 solutions systematically by 2020 (see Figure 1). Conversely, by 2020 still one fifth of the respondents indicated that they will have made at best piecemeal progress.

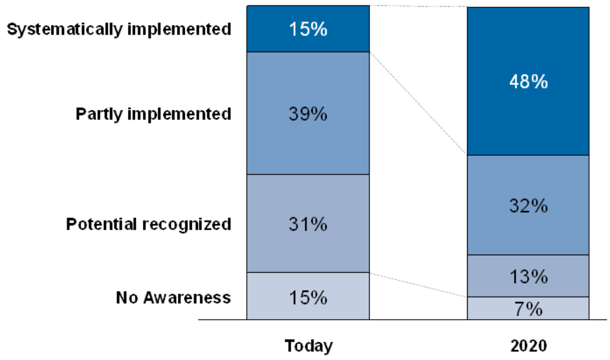


Figure 1 - Use of Industrie 4.0 concepts to manage assets

The survey further found significant variance in the adoption levels in different markets. Figure 2 (see Figure 2) below, shows that in 2015 68 percent of respondents from China had partially or systematically implemented Industrie 4.0 in asset management programs, estimated to increase to 89 percent by 2020. Comparable numbers for France are 27 percent and 58 percent respectively.

Across the five countries surveyed - China, France, Germany, the United Kingdom and the United States – the level of maturity in Industrie 4.0 varied significantly. While no country could claim to be the global early adopter in implementing Industrie 4.0, the percentage of companies in China that claimed to be early adopters was significantly higher than anywhere else (see Figure 4).

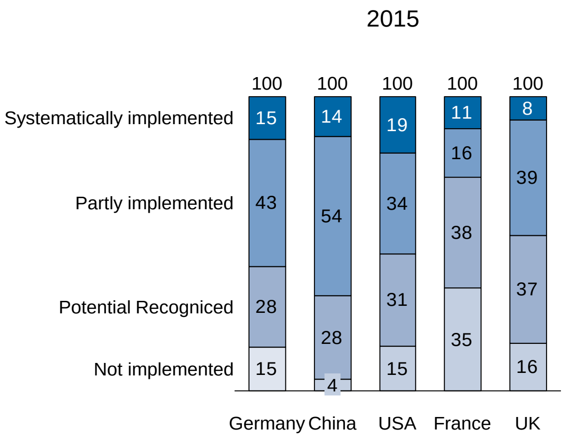


Figure 2 - Country comparison of Industrie 4.0 concepts to manage assets in 2015

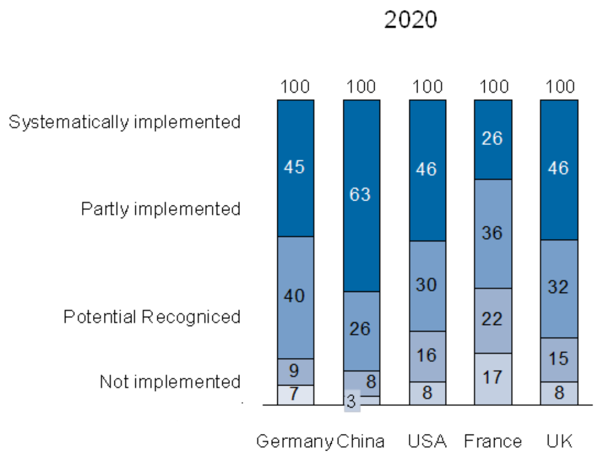


Figure 3 - Country comparison of Industrie 4.0 concepts to manage assets in 2020

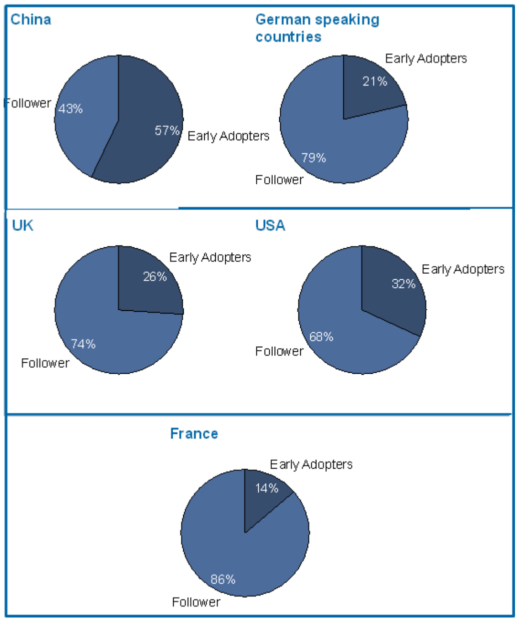


Figure 4 - Early Adopter and Follower analysis by country

It is expected that a number of factors were and still are driving this; notably the focused initiatives and investment from the Chinese Government to develop more sustainable industry growth. In addition, manufacturing is core for China and the market is accustomed to rapidly implementing new technology, especially in green field sites free of legacy infrastructures. Germany (21 percent), the United Kingdom (36 percent) and the United States (32 percent) have similar maturity footprints in terms of both 2015 status and 2020 ambition. This could be because of their historical leadership in manufacturing. In France (14 percent), the Industrie 4.0 implementation was comparatively less mature. The economic downturn phase in 2015 and unsuccessful digitization programs could have been contributing factors.

3 OBJECTIVE AND GOAL: DEVELOPMENT OF THE ACATECH INDUSTRIE 4.0 MATURITY INDEX

The global benchmark on Industrie 4.0 maturity presented earlier clearly indicated an urgent need to act. Before this background, we present the approach and model developed by an industry consortium with the support of acatech in Germany. The model's approach is based on a succession of maturity levels, i.e. *value-based development* stages (see section 3.1) that help companies navigate their way through every stage in the digital transformation. To ensure that all aspects of manufacturing companies are taken into account, the model's structure is based on the "Production and Management Framework" by [6]. The framework's four structural areas enable a comprehensive analysis and set out a number of guiding principles that allow companies to identify which Industrie 4.0 capabilities they still need to develop. In the following, the underlying concepts and model of the acatech Industrie 4.0 Maturity Index are illustrated.

3.1 Value-based Development Stages – Industrie 4.0 Maturity Levels

As a first step, a general awareness is necessary that Industrie 4.0 can be achieved stepwise and according to individual company benefits. The basic structuring of Industrie 4.0 into successive stages is presented below in Figure 5.

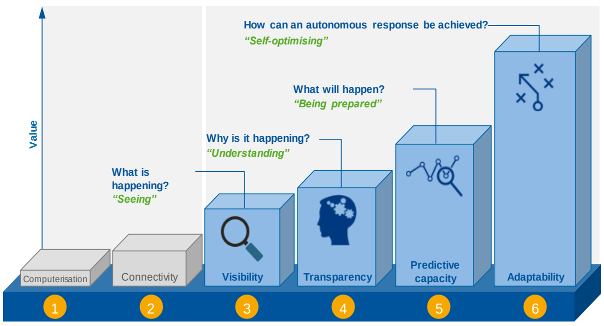


Figure 5 - Value-based development stages of Industrie 4.0 – Industrie 4.0 Maturity Levels [3]

The development path is based on computerization (1), which is the starting point for digitization and refers to the targeted use of information technologies. It enables cost-effective production with low error rates and generates the necessary precision, which is indispensable for the production of many modern products [3; 9].

Achieving the connectivity level (2), the targeted or isolated use of IT is replaced by networked components. A complete integration between IT (information technologies) and OT (operative technologies) levels has not yet taken place; however, interfaces to business IT are provided by parts of implemented OT [3; 11; 12].

Based on this, a digital visibility (3) is established with the help of sensors, which enable recording of processes from start to finish with a high amount of captured data. Processes states are no longer limited to individual areas, such as in a production cell, but can be extended to a production system or the entire company in real time in order to create a digital model, also known as the "digital shadow" [3; 7; 9].

For a better causal understanding of processes, it is necessary to create further transparency (4) about the correlations in data stocks. Process knowledge is more and more required to support more complex decisions, which are based on semantic connections and aggregation of data.

Building up, the predictive capability level (5) enables simulation of different future scenarios and identification of those that are most likely. To this end, the digital shadow is projected into future-based scenarios and evaluated according to probability of occurrence. The ability to adapt (6) can enable an automatic reaction to expected machine failures or delays in delivery through a modified sequence in production planning.

3.2 Required Capabilities for a Company's Structural Areas

The skills that are relevant for the transformation of a manufacturing company into a learning, agile organization are assessed through the four structural areas of resources, information systems, culture and organizational structure (see Figure 6). All of them characterize the structure of an organization and are examined over the six levels of the Industrie 4.0 development path, which is represented by six concentric circles in Figure 6.

Each structural area is divided by two principles, each of which - depending on the benefit-oriented development levels - successively builds up skills. These skills guide the further development of the manufacturing company. The degree to which the abilities are implemented determines the maturity level of each principle. The maturity levels of the two principles are summarized and together they represent the evaluation of the structural area, which is oriented on the development levels.

The structural area resources includes all physical, tangible resources. This contains, for example, employees of a company, machinery and systems, the tools and materials used and the final product.

The two principles dividing this structural area are differentiated into digital competence and structured communication. "Digital competence" (a) characterizes the generation of data and its target-oriented independent processing into information by resources with corresponding technical components. This facilitates an information-driven way of working, based on feedback from the process environments and not on forecast-based planning specifications. The skills of digital competence also include the use of embedded systems and the retention of digital competence, which can only be successful, if attention is paid to promoting interdisciplinary thinking and action by employees - if they are increasingly integrated into the innovation process. Through “structured communication” (b) collected information is linked and creates an overall picture. An efficient communication can be defined and interface designed in order to support decision-makers. [3; 7; 8]

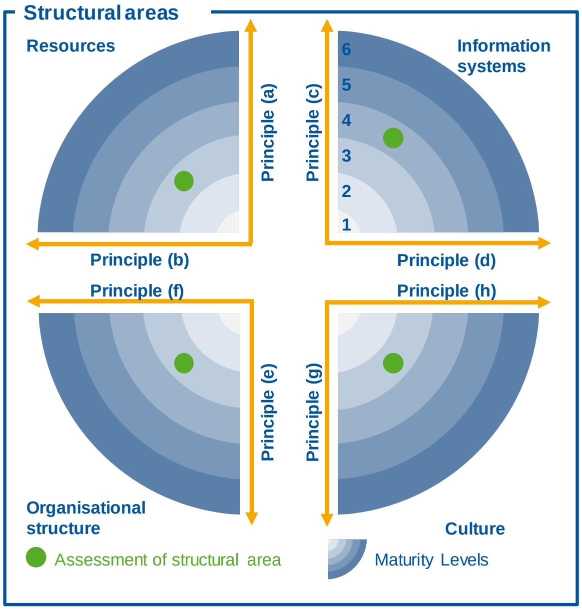


Figure 6 - Structure of the structural areas [3]

With the help of employees, information technologies and data information is available within information systems in accordance to economic criteria. Many manufacturing companies do not make sufficient use of data. The decisive factor is the insufficient processing of the collected data into information and its subsequent provision to the employees, which is why the first principle includes the processing and preparation of data (c) for decision support. This requires, among other things, context-based information provision, data storage and application-oriented interfaces in order to provide a technical infrastructure for real-time use of data and information ultimately [10]. In the context of the second principle, it is a question of integration for optimized data (d) use and increased agility under the primary aspect of data sharing within the value chain. However, this is only possible through the use of one leading information system that eliminates the need to keep duplicates in different IT systems. Standardized interfaces, detailed IT security and vertical and horizontal integration of information systems are required.

The transformation to a learning, agile company is achieved through the technologies explained above and the implementation of an appropriate organizational structure. In this model, the organizational structure refers on the one hand to the internal corporate organization (e) in the form of organizational structures and processes, and on the other hand describes the positioning in the value network (f). In contrast to the structural area culture described below, the organizational structure establishes mandatory rules that organize collaboration both within the company and externally. A high degree of individual responsibility on the part of employees is characteristic of the organic internal organization. In particular, dynamic collaboration requires skills that contribute to a smooth and automated exchange of information between companies [3].

A company’s agility is highly dependent on the behaviour of its employees. In this context, two directions for changing corporate culture can be mentioned: willingness to change (g) and social collaboration (h). Willingness to adapt goes along with the prerequisite of being able to recognize opportunities for change and then initiate appropriate measures. In addition, it is advantageous to see mistakes not as a problem, but as an opportunity for positive change as well as a willingness to undergo continuous further training. The term social collaboration refers to the consideration of knowledge as a decisive guideline for action, which implies that an ideal state is characterized by making decisions based on knowledge [8].

The benefit levels presented allow companies to better assess their own Industrie 4.0 development and determine the next stage on their transformation path [2, 3].

4 APPLICATION OF THE ACATECH INDUSTRIE 4.0 MATURITY INDEX

The application of the acatech Industrie 4.0 Maturity Index consists of three successive phases (see Figure 7) [3] illustrated in the following.

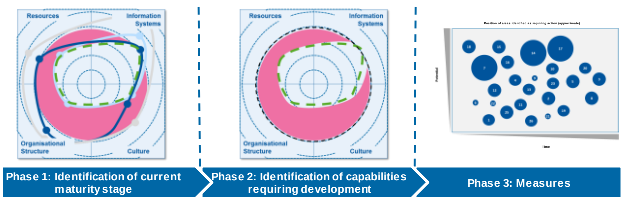


Figure 7 - Application of the acatech Industrie 4.0 Maturity Index [3]

4.1 Phase 1: Identifying the current Industrie 4.0 development stage

The company's location results from the six value-based development stages for Industrie 4.0 (see Section 3.1.) and the skills dedicated to them. For an appropriate assessment, the consortium has developed a questionnaire, with approx. 600 questions for the business processes *engineering, production, logistics, service, sales* and *marketing*. An inspection of a production plant can give a first impression of the processes, followed by a detailed evaluation of the business processes. The asis analysis and the questionnaire is conducted based on the order processing process, which forms the framework situation for the evaluation of existing skills. With the help of a questionnaire evaluation the Industrie 4.0 Maturity Index and Levels can be identified for each structural area.

4.2 Phase 2: Capabilities to be acquired

For the evaluation, the answers to the questionnaire shape the basis for the evaluation of the current situation of the company by the radar image (see Figure 7). The dependencies of the structural areas determine consistent development in all structural areas as an essential goal. This is the basis for the recommendation for companies to approach the resulting areas of action and to strive for a consistent maturity stage across all four structural areas and in this way to use the maturity stage (achieve maturity stage consistency).

4.3 Phase 3: Identifying concrete measures

The next step is to derive measures addressing areas identified as requiring action. Necessary measures can be deduced from the missing capabilities evaluating the four structural areas. By evaluating individual processes, many individual measures can be dedicated, which makes it easier for companies to create a digital roadmap.

In defining strategic objectives for a company, identified measures are worked out precisely. Achieving the targeted stages of development, in turn, aims to support the realization of the strategic objectives formulated at the outset. This enables decisionmakers in manufacturing companies not only to identify at a glance the measures needed to achieve a higher maturity level, but also the interdependencies between identified measures. The purpose of this presentation is also to simplify the creation of a digitization roadmap by determining the order in which measures are implemented in terms of time and costs [3].

5 CONCLUSIONS AND OUTLOOK

The acatech Industrie 4.0 Maturity Index provides companies a supporting tool for transformation into a learning, agile company. This approach has proven in many cases to be of particular relevance for producing companies which has an outstanding role for more traditional structured, manufacturing based economies such as the European economy. It is important to differentiate this against more disruptive perspectives we know from venture capital driven and greenfield innovation in the IT-industry with its focus on end consumer offerings. The methodology describes six benefit-oriented development stages for four key areas. Each stage corresponds with additional, achievable benefits. This approach can be applied to develop a digital roadmap tailored to the needs of individual company in order to help them master the digital transformation if – as mostly in the case of the existing manufacturing and production industry – an evolutionary and structured transformation approach is the best choice. In this cases, the approach presented supports companies to guide their transformation in a very structured and most efficient and fast process.

6 REFERENCES

[1] Schuh, G.; Jordan, F.; Maasem, C.; Zeller, V.: Industrie 4.0: Implikationen für produzierende Unternehmen. In: Digitale Transformation im Unternehmen gestalten. Geschäftsmodelle, Erfolgsfaktoren, Handlungsanweisungen, Fallstudien. Hrsg.: O. Gassmann; P. Sutter. Hanser, München 2016, p. 39–58

[2] Schmitz, S.; Wenger, L.: Acatech Industrie 4.0 Maturity Index: Welche Fähigkeiten sind im Wandel entscheidend? In: IT&Production (2017b) 5, S. 54 – 55.

[3] Schuh, G., Anderl, R., Gausemeier J., ten Hompel, M., Wahlster, W. (Eds.): Industrie 4.0 Maturity Index. Managing the Digital Transformation of Companies (acatech STUDY), Munich: Herbert Utz Verlag 2017, p. 21

[4] Infosys (2015): Industry 4.0. The state of the nations. 1. Aufl. Hg. v. Infosys Ltd. Bangalore, Indien.

[5] Gudergan, G., Stich, V., Schmitz, S., Buschmeyer, A. (2016): The Global Evolution of The Industrial Internet of Things – A Cross Country Comparison Based on an International Study on Industrie 4.0 for Asset Efficiency Management. In: Proceedings of the International Conference on Competitive Manufacturing. Coma ’16. February 2016, Stellenbosch, South Africa.

[6] Boos, W./Völker, M./Schuh, G.: „Grundlagen des Managements produzierender Unternehmen“. In: Schuh, G./Kampker, A. (Hrsg.): Strategie und Management produzierender Unternehmen. Handbuch Produktion und Management 1, Berlin, Heidelberg: Springer Verlag (VDI-Buch) 2011, S. 1–62.

[7] Bauernhansl, Thomas; Krüger, Jörg; Reinhart, Gunther; Schuh, Günther: WGP-Standpunkt Industrie 4.0. Hg. v. Wissenschaftliche Gesellschaft für Produktionstechnik Wgp e. V.

[8] Zühlke, Detlef (2013): Die Cloud ist Voraussetzung für Industrie 4.0. Präsentation. VDI. VDI-Pressegespräch anlässlich des Kongresses „AUTOMATION 2013“. Baden-Baden, 6/25/2013

[9] Schuh, G.; Potente, T.; Thomas, C.; Hauptvogel, A. (2014): Steigerung der Kollaborationsproduktivität durch cyber-physische Systeme. In: Thomas Bauernhansl, Michael ten Hompel und Birgit Vogel-Heuser (Hg.): Industrie 4.0 in Produktion, Automatisierung und Logistik. Wiesbaden: Springer Fachmedien Wiesbaden, S. 277–296.

[10] Hering, Niklas; Brandenburg, Ulrich; Frey, Dominik; Ihne, Manfred; Meißner, Jan; Reschke,

Jan; Schenk, Michael (2015): Smart Operations. Hg. v. FIR an der RWTH Aachen. Aachen.

[11] Vogel-Heuser, Birgit (2014): Herausforderungen und Anforderungen aus Sicht der IT und der Automatisierungstechnik. In: Thomas Bauernhansl, Michael ten Hompel und Birgit Vogel-Heuser (Hg.): Industrie 4.0 in Produktion, Automatisierung und Logistik. Wiesbaden: Springer Fachmedien Wiesbaden, S. 37–48.

[12] Kaufmann, Thomas; Forstner, Lisa (2014): Die horinzontale Integration der Wertschöpfungskette in der Halbleiterindustrie. Chancen und Herausforderungen. In: Thomas Bauernhansl, Michael ten Hompel und Birgit Vogel-Heuser (Hg.): Industrie 4.0 in Produktion, Automatisierung und Logistik. Wiesbaden: Springer Fachmedien Wiesbaden, S. 359–367.

7 BIOGRAPHY

Prof. Dr. Volker Stich – Prof. Dr. Volker Stich has a PhD. in Engineering from RWTH Aachen University. He is Managing Director of FIR Research Institute for Industrial Management at Aachen University. Before being nominated as the Managing Director of FIR, he worked for a global supplier in the automotive sector responsible for the design of the global logistics management.

Dr.-Ing. Gerhard Gudergan – Dr.-Ing. Gerhard Gudergan has a master in mechanical engineering and a PhD from RWTH Aachen University. He is head of the Business Transformation department at FIR at RWTH Aachen University. Since 2009, he is the head of business unit research at FIR. Previously, he established and ran the service management department and is the founding head of the service science innovation lab at RWTH Aachen Campus. His current focus is on business.

Human Digital Twin for Integrating Human Workers in Industry 4.0

D. Sparrow[1], K. Kruger[1], A. Basson[1]
[1]Department of Mechanical and Mechatronic Engineering, University of Stellenbosch, South Africa.

Abstract

The Industry 4.0 vision relies on Cyber-Physical Production Systems to enhance system intelligence, connectedness and responsiveness. In achieving these characteristics, it is necessary to maintain a virtual representation of real world entities, referred to as a Digital Twin. While recent research has predominantly focussed on developing digital twins for machines and systems, the same need exists for connecting humans. A human digital twin can connect humans with the speed and accuracy of other digital entities, while making full use of a human's problem solving and versatility and ensure job opportunities for future factory workers. This paper discusses the need for and value of a human digital twin and presents some important requirements for such a development.

Keywords

Digital Twin, Industry 4.0, Holonic Manufacturing Systems

1 INTRODUCTION

The manufacturing industry has seen three industrial revolutions and the fourth industrial revolution, Industry 4.0 (I4.0), is on the horizon. I4.0 represents a fundamental change to the way systems are managed using digitization and the internet, combined with machine learning and statistical analysis. I4.0 is driven by modern manufacturing requirements of flexibility, robustness to environmental changes and the ability reconfigure and optimize to match new custom requirements [1][2].

While I4.0 research has predominantly focussed on the connecting of machines and systems, it is important to consider the integration of humans within the I4.0 vision as well.

Human workers currently offer irreplaceable, critical skillsets in the manufacturing environment. Human workers and managers contribute "roll-up-your-sleeves-effort, resourcefulness and creativity" to achieve new levels of performance [3]. Furthermore, apart from their autonomy and problem-solving capabilities, humans possess dexterity, sensory and decision-making ability still unmatched by autonomous systems.

For these reasons, industry still recognises the importance of human workers and top producers in the world are trying to maximize their benefits. BMW have declared that full automation is not their goal, since a human is unbeatably flexible [4], and reflection on Tesla's Model 3 production issues revealed that "excessive automation at Tesla was a mistake" [5]. This view is also shared by the composites industry, where the nature of the materials and complexity of manufacturing tasks make automation challenging or even impossible. Here, the skills of human workers are critical to production [6].

Interviews with industry pointed out that manufacturing in developing countries relies on smaller volume production of a larger variety of products. I4.0 can play an important role in this context by reducing the reliance on expensive manufacturing equipment and enhancing the use of information technology for optimizing the shop floor. In effect, it could level the playing field between factories with expensive machines and those that can utilize their human workers' full potential [7].

Exploring the issue of human integration, Rey, Carvalho and Trentesaux [8] report that careful consideration needs to be given to the difference in how artificial systems and humans treat response times, background knowledge, context understanding, information management and other cognitive functions. The authors then propose different ways in which humans and Artificial Self Organising (ASO) systems can exploit mutual advantages and discuss Human-ASO integration issues – specifically considering tight timeframe requirements for decision making on large amounts of data and calculations, and where communication breaks down due to a human's nature to summarize data into abstract ideas.

This paper presents the need for and value of a digital administration shell for human workers to facilitate their integration into Cyber-Physical Production Systems (CPPSs). The proposed solution is a Human Digital Twin (HDT), inspired by holonic manufacturing system concepts, which can handle scheduling, communication to other holons, and metrics analysis on behalf of the human worker.

The paper presents some background on I4.0 and CPPS, digital twin and HMS concepts (sections 2,3, and 4), before discussing the HDT in section 5. The paper concludes with a discussion of future work in section 6.

2 I4.0 AND CPPS

Digital technologies provide the communication and information processing abilities required to enable the manufacturing concepts discussed in section 1. As such, these technologies can revolutionize the integration of physical systems into Cyber-Physical Systems (CPSs).

CPSs are transformative technologies that utilise the increasing number of sensors and data acquisition, storage and processing mechanisms, as well as networking capabilities that modern technology provides. Applying CPS to production, Cyber-Physical Production systems (CPPS) can enable manufacturing entities to reach resilience, self-adaptability and intelligence, and transform factories with significant economic potential by integrating logistics and services [9].

A Reference Architecture Model for Industry 4.0 (RAMI 4.0) has been developed by the German Electrical and Electronic Manufacturer's Association and their 6000 associate companies. Thereby they describe the addition of a digital administration shell around physical or software entities to merge them into an I4.0 environment [10].

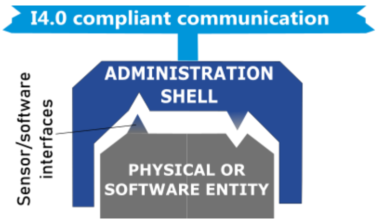


Figure 1 - Admin shell allows Industrie 4.0 compatibility (adapted from [10]).

3 THE DIGITAL TWIN

With the development of the *Internet of Things*, *Big Data* and machine learning, large volumes of data can be collected and aggregated throughout an entity's life-cycle. This requires a structured method for handling the historical state, current state, and predicted future states of that entity. The *Digital Twin* (DT), a concept that was first mentioned by Grieves and Vickers [11], is a technology developed to address this. Aaron and Lane [12] defines the DT as "an evolving digital profile of the historical and current behaviour of a physical object or process that helps optimize business performance". The DT would allow better simulations of entities (or systems of entities) prior to creation and integration, as well as aid in individualised maintenance.

4 HOLONIC MANUFACTURING SYSTEMS

Holonic Manufacturing Systems (HMS) are inspired from the concept of a holon, which is an autonomous, cooperating and goal-driven entity that itself forms part of a larger holon. A HMS then applies this holonic concept to integrate all the manufacturing activities of an enterprise.

HMSs offer robustness, flexibility, scalability and reconfigurability due to their fractal nature, as well as the autonomy and intelligence between communicating holons [13] [14]. The HMS approach is therefore a suitable approach for developing a Human Digital Twin due to the many holonic qualities shared by humans.

HMSs have seen a great deal of research attention in the development of suitable reference architectures, such as PROSA, ADACOR and recently the ARTI reference architecture (which is a generalized version of PROSA) [15][16]. The focus here will be on the use of resource holons described in the PROSA and ARTI reference architectures.

HMS research has considered the integration of humans, e.g. McFarlane and Bussmann [17] mention the integration of humans by a human interface block, and various other authors stress the importance for human integration with industry 4.0. However, the serious communication and data processing differences between humans and digital holons have not yet been adequately addressed.

Humans on a shop floor would fall under the resource holon classification and therefore needs to comply to the responsibilities of the resource holon (as described by Valckenaers and Van Brussel [13]) :

Reflection of reality: Reflect its corresponding physical resource, with synchronized information about state and expected future states, using knowledge about its dynamic behaviour.

Information provision: Provide resource related information to other holons on its processes, local topology and possible constraints.

Maintaining a local schedule: Own an agenda that records future tasks to provide a reservation system and keep track of the availability of the resource over time.

Managing its local schedule: The resource holon has local authority on how it accomplishes the scheduled tasks, which depends on the holon's capabilities.

Virtual execution: This is a service for the order holons (holons responsible for logistics) that can ask for the virtual outcome of an operation (e.g., time or quality). Using its local schedule and reflection of reality, the resource holon should be able to provide this information.

Controlling the resource: The resource holon controls the real-world resource by starting and stopping its scheduled tasks and monitoring the execution.

The internal structure of a holon in Figure 2, as proposed by Nylund [18], takes into account technological developments in virtual modelling, simulation, and digital storage. It has a similar communication shell to the RAMI4.0 architecture of **Figure 1**. For a human resource holon, this communication layer and the digital and virtual components should address the six resource holon responsibilities discussed above. Furthermore, these components should supplement the data processing and storage capabilities of the human to a standard that matches other digital holons.

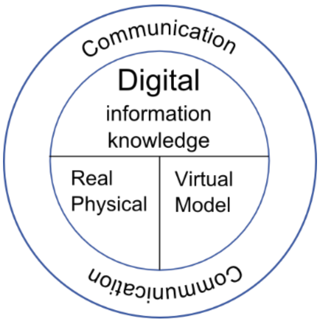


Figure 2 - Holon structure adapted from Nylund [18]

5 THE HUMAN DIGITAL TWIN

The Human Digital Twin (HDT) is proposed as a necessary technology to facilitate human worker integration in an Industry 4.0 environment. The HDT should address the communication, data aggregation, simulation and scheduling requirements that must be satisfied by a human holon.

The realization of the I4.0 vision requires new mechanisms for integrating human workers in their environment. These mechanisms should allow for natural interaction with the human senses and facilitate the capture and digitization of data from human workers, for use by other digital entities. This section describes the considerations involved in the development of such mechanisms.

5.1 Human-to-machine communication

The HDT requires interfaces with the human worker to gather data for analysis and prediction. There exist many proven interfacing technologies apart from the traditional keyboards, touch pads or controllers. Modern human-to-machine communication technologies, such as motion capture, enables accurate tracking of a human's body and actions – as demonstrated in a manufacturing environment by Ferrari [19] (see **Figure 3**) and Cheng [20].

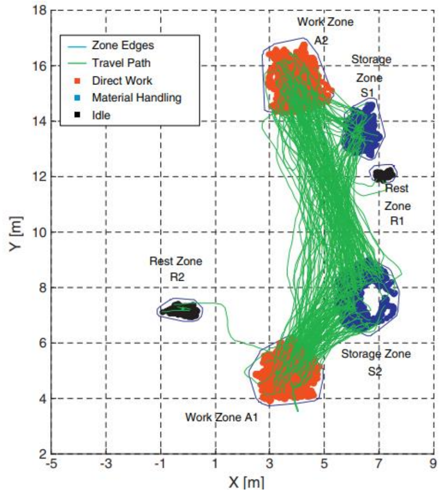


Figure 3 - Work flow capturing using multiple sensors [19].

RFID has seen success in many industries for its durability, rich data capacity, contactless nature and low power to size ratio. Ko developed a working algorithm for 3D location detection using RFID [21] and Valero showed RFID's limitations are reduced when fused with other sensor technologies [22].

Voice Command has been in development ever since computers supported the processing power needed. Microsoft, Apple and Google all have operating systems that can be controlled using only voice commands [23]. Industry examples include the Active Listening Voice Command System [24] and Tabibian's aerospace framework for voice command detection [25].

Eye-tracking allows computers to capture the gaze and focus where Susanto analysed gaze behaviour during assembly of carburettors [26]. An eye-tracking system has also been used to improve safety and performance in a manufacturing setting [27].

Many other technologies exist and many more will still be developed to assist in capturing input and sensed information from humans – however, a centralized access point where all this data is reconciled and processed needs to be developed.

5.2 Machine-to-human communication

For machines to communicate to humans, screens, numerical displays, lights and speakers are matured technologies. However, these technologies require the user to be near the communication device and has restricted screen size in two dimensions.

Other than screens, lights and speakers, a new group of technologies are being developed. Augmented Reality (AR), which is the augmentation of a real-world object or scene by computer perceptual information, has been used successfully in military and medical contexts for some time. However, the technology has not seen widespread adoption in the manufacturing industry, with only a few specific and isolated implementations available [28].

Smart glasses and head mounted displays are a form of visual AR that display computer generated scenes, and have been used in surgery training and maintenance process documentation [29] [30]. AR has also been implemented through work station projection, which allows for the display of information directly onto the work place without the need to wear a headset [31] [32].

In haptic feedback applications, a sense of touch is generated through vibrations or forces. Haptic feedback systems have been used for virtual training [33].

Speech synthesis has seen applications and development ever since the IBM704 sang Daisy Bell for the first time. Modern GPS systems, smart phones, Cortana, Siri, Alexa and various other common examples exist of successful speech synthesis technology.

5.3 Human behavioural modelling

Modelling human behaviour allows partial automation for the prediction component of the HDT. This will enable the HDT to answer simple "what if" questions on behalf of the human, as well as improve safety monitoring. The HDT maintains metrics for processes or events that have occurred, which can be used to more accurately schedule and simulate future events.

5.3.1 Human mental behaviour modelling

Rey, Carvalho and Trentesaux [8] show that *human-in-the-loop* frameworks would benefit from human resources that have tools to support them when interfacing with digital systems. These tools would become decision partners and co-operators that could mitigate the shortcomings of a human (like attention span or calculation capability), but also strengthen their ability to act flexibly in unforeseen circumstances.

Numerous research papers deal with behaviour modelling of humans. Elkosantini describe agent-based models for human-centred systems and present their own model for human behaviour, considering psychological and social backgrounds, and physical capacity [34].

5.3.2 Human physical behaviour modelling

Calzavara discuss the importance of integrating operator fatigue and recovery into the analysis of decision support models in production systems. Studying the exponential trend of fatigue accumulation and recovery, the model uses the energy expenditure and predicted heart rate of the operator [35]. Studies like these show that human behaviour and operations can be digitized into manipulatable data, like *Maximum Endurance Time* or fatigue curves.

5.4 What Industry 4.0 stands to gain from the Human Digital Twin

In terms of the RAMI4.0 architecture, the HDT will be the administration shell to the physical human resource. The HDT can then complete the resource holon responsibilities mentioned in section 4. The HDT would provide a single homogenous point of information of a human holon to any other holon that may require it.

The HDT completes the missing communication component of the human holon. Using an example to demonstrate the HDT's role:

An AGV fleet, conveyor, and human all have the ability to carry a pallet across the shop floor and an order holon needs to choose one of the three resources to do the job. The AGV fleet and conveyor holons can instantly communicate the lead time for the job, the energy cost, the path to be taken and any number of other details. In contrast, the human worker would need to manually enter this data on a keyboard, which not only interferes with his current work, but delays the decision-making process. The human worker is also burdened with calculating the details of accomplishing this simple task in order to give the order holon an estimate [16].

With the data held by the HDT (such as the worker's average walking speed), a simple path-finding calculation and a check of the human's current schedule can be used to estimate a lead time that can be sent to the order holon instantly. This leaves the human worker free to continue with their work, only to receive a notification by their HDT to perform the task if the human holon was chosen.

The HDT is thus a necessary component to complete the human holon and serve as the administration layer of the RAMI4.0 architecture. The following sections give more concrete examples of the benefits of the HDT for industry and its workers.

5.4.1 Analysing, tracking and predicting true standard work

Murthy indicated that the most time-consuming part of implementing lean manufacturing processes on the shop floor is the reliable gathering of data on human workers. Currently, stopwatches are commonly used to gather time data for work study methods and automation of this process, as facilitated by the HDT, would be extremely valuable [7] .

5.4.2 Real time process feedback and updates

The HDT would provide real-time process feedback in manufacturing settings. This feedback would allow for the detection of anomalies before they affect processes down the line, and *Takt* time could be controlled more thoroughly.

5.4.3 The Human Digital Twin enables gamification

Gamification is a term used to describe a task that has elements of game design, such as point scoring, event effects, or any other entertaining elements to improve engagement and productivity.

5.4.4 Quality feedback mechanism for human workers

Factory workers often never see their work down the line. Therefore, they have no feedback on the work they've done or how they could improve. If feedback or process techniques need to be referenced back to a worker, there is either no trace back records or the bureaucracy of getting the information back to the worker in a useful time and manner proves ineffective. If the HDT is informed about a work piece as soon as other holons detect the fault or success, the human worker can receive immediate feedback.

5.4.5 Assisting the resolution of labour disputes

Data from the HDT could be used to analyse the strain or dangers a human worker is subjected to, as well as provide a single source of truth on events. Since the HDT of the worker also has the record of all his work pieces, a worker may be able to prove competence or unfair dismissal.

5.4.6 Faster worker integration

The HDT could hold worker-specific data, such as preferred language, height, walking speed, eye-sight, etc. The HDT could therefore be transferred along with the worker to a new job site or company where training and integration could be customized automatically based on his/her HDT.

5.4.7 Multi-dimensional safety monitoring

The HDT would be able to communicate the digitized state of the human worker to other holons in real time, and vice versa. This communication could enable shop floor safety monitoring to be carried out using multiple data points as received by the HDT and the multiple holons it provides data to, eliminating blind spots.

If the HDT has data on the medical history of a worker, certain oversights can be avoided – such as assigning strenuous work when the worker is on fluoroquinolones [36]. Possibilities like this become potentially infinite.

6 CONCLUSIONS

Humans are still critically important to industry, but there are large information processing and communication differences between humans and machines. The paper proposes a Human Digital Twin (HDT) to facilitate this integration, drawing inspiration from the resource holon concept applied in holonic manufacturing systems.

The HDT needs to supplement the human's ability to store, process, and communicate relevant information of itself or of processes in its control to other digital entities by making use of modern Human-Machine interfaces and Human Behavioural Modelling. The HDT should fulfil the RAMI4.0 administration shell requirements for a human worker connecting to a RAMI4.0 compatible communication channel. The HDT should complete the resource holon requirements as stated by the PROSA and ARTI reference architectures. The HDT architecture should be structured around the ARTI holonic reference architecture to ensure it has a generalised form for any industry that may use it.

Future research will entail detailed requirement and functional analyses for a generic HDT, which will then be used to formulate an HDT architecture. The architecture will be implemented and evaluated in an industrial case study.

7 REFERENCES

[1] H. Lasi, P. Fettke, H. G. Kemper, T. Feld, and M. Hoffmann, "Industry 4.0," *Bus. Inf. Syst. Eng.*, vol. 6, no. 4, pp. 239–242, 2014.

[2] S. S. Dr Reinhard Geissbauer, Jesper Vedso, "Industry 4.0: Building the digital enterprise," *PwC*, pp. 1–36, 2016.

[3] M. Rother, *TOYOTA KATA*. McGraw-Hill Education, 2010.

[4] Steven Loveday, "BMW Comments On Tesla Model 3 Production Woes," 2018. [Online]. Available: https://insideevs.com/bmw-comments-on-tesla-model-3-production-woes/. [Accessed: 23-Jul-2018].

[5] E. Musk, "Elon Musk on Model 3 production issues." 2018.

[6] E. Sparrow, "Interview on I4.0 and the composites industry." Abu Dhabi, 2017.

[7] S. Murthy, "Interview on lean manufacturing and Industry 4.0." 2018.

[8] G. Z. Rey, M. Carvalho, and D. Trentesaux, "Cooperation models between humans and artificial self-organizing systems: Motivations, issues and perspectives," *Proc. - 2013 6th Int. Symp. Resilient Control Syst. ISRCS 2013*, pp. 156–161, 2013.

[9] J. Lee, B. Bagheri, and H. A. Kao, "A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems," *Manuf. Lett.*, vol. 3, pp. 18–23, 2015.

[10] E. Ulrich, "Status Report Industrie 4.0-Technical Assets Basic terminology concepts, life cycles and administration models," Düsseldorf, 2016.

[11] M. Grieves and J. Vickers, "DigitalTwin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems" *Orig. Digit. Twin Concept*, no. August, pp. 1–8, 2016.

[12] P. Aaron and W. Lane, "Industry 4.0 and the digital twin Manufacturing meets its match,"

2017.

[13] P. Valckenaers and H. van Brussel, *Design for the unexpected : from holonic manufacturing systems towards a humane mechatronics society*. Elsevier, 2016.

[14] U. K. Monostori L, Kadar T, Kondoh S, Kumara S, Reinhart G, Sauer O, Schuh G, Sihn W, "CIRP Annals - Manufacturing Technology Cyber-physical systems in manufacturing," vol. 65, pp. 621–641, 2016.

[15] K. Kruger and A. Basson, "An Erlang - based Resource Holon for a Holonic Manufacturing Cell."

[16] H. Van Brussel, J. Wyns, P. Valckenaers, L. Bongaerts, and P. Peeters, "Reference Architecture for Holonic Manufacturing Systems," *Comput. Ind.*, vol. 37, no. 3, pp. 255–274, 1998.

[17] D. C. McFarlane and S. Bussmann, "Holonic Manufacturing Control: Rationales, Developments and Open Issues," *Agent-Based Manuf.*, pp. 303–326, 2003.

[18] H. Nylund, K. Salminen, and P. Andersson, "Digital Virtual Holons — An Approach to Digital Manufacturing Systems," in *Manufacturing Systems and Technologies for the New Frontier*, London: Springer London, 2008, pp. 103–106.

[19] A. R. Emilio Ferrari, Mauro Gamberi, Francesco Pilati, "Motion analysis system for the ergonomic assessment of manufacturing and assembly manual activities," *16th IFAC Symp. Inf. Control Probl. Manuf.*, 2018.

[20] Tao Cheng, Jochen Teizer, Giovanni C. Migliaccio, and Umberto C. Gatti, "Automated task-level activity analysis through fusion of real time location sensors and worker's thoracic posture data," *elsevier*, 2012.

[21] C.-H. Ko, "RFID 3D location sensing algorithms," *Autom. Constr.*, 2010.

[22] E. Valero and A. Adán, "Integration of RFID with other technologies in construction," 2016.

[23] Tom Geller, "Talking to Machines," *Commun. ACM*, 2012.

[24] B. Henry, L. Green, A. Listen-, V. L. H. Controls, E. Alvc, and F. Maleno, "Voice command system turns machines on," no. February, 2006.

[25] S. Tabibian, "A voice command detection system for aerospace applications," *Int J Speech Technol*, vol. 20, pp. 1049–1061, 2017.

[26] N. Susanto, "Analysis of the Gaze Behaviour of the Worker on the Carburetor Assembly Task," *J. Tek. Ind.*, vol. 10, no. 2, 2015.

[27] Tobii, "A Tobii Pro Insight eye tracking study How to reduce accidents and increase efficiencies in an industrial setting Visual Performance in the Foundry," 2018.

[28] M. Quandt, B. Knoke, C. Gorldt, M. Freitag, and K. Thoben, "ScienceDirect General Requirements for Industrial Augmented Reality Applications," *Procedia CIRP*, vol. 00, no. May, pp. 1130–1135, 2018.

[29] R. G. Peden, R. Mercer, and A. J. Tatham, "The use of head-mounted display eyeglasses for teaching surgical skills: A prospective randomised study," *Int. J. Surg.*, 2016.

[30] F. Quint and F. Loch, "Using smart glasses to document maintenance processes," *ensch und Comput. 2015 Work.*, no. September, pp. 203–208, 2015.

[31] P. Bertram, M. Birtel, F. Quint, and M. Ruskowski, "Intelligent Manual Working Station through Assistive Systems," 2018.

[32] A. Doshi, R. T. Smith, B. H. Thomas, and C. Bouras, "Use of projector based augmented reality to improve manual spot-welding precision and accuracy for automotive manufacturing," *Int. J. Adv. Manuf. Technol.*, vol. 89, no. 5–8, pp. 1279–1293, 2017.

[33] F. Loch, U. Ziegler, and B. Vogel-heuser, "Integrating Haptic Interaction into a Virtual Training System for Manual Procedures in Industrial Environments," *Proc. 16th IFAC Symp. Inf. Control Probl. Manuf.*, 2018.

[34] S. Elkosantini, "Toward a new generic behavior model for human centered system simulation," *Simul. Model. Pract. Theory*, vol. 52, pp. 108–122, 2015.

[35] M. Calzavara, A. Persona, F. Sgarbossa, and V. Visentin, "A model for rest allowance estimation to improve tasks assignment to operators," *Int. J. Prod. Res.*, vol. 0, no. 0, pp. 1–15, 2018.

[36] H. Goyal, J. Dennehy, J. Barker, and U. Singla, "Achilles is not alone!!! Ciprofloxacin induced tendinopathy of gluteal tendons Learning Points for Clinicians," *QJM*, 2015.

8 BIOGRAPHY

Dale Sparrow is a Post Graduate Candidate at the University of Stellenbosch. He is currently doing research on Industry 4.0 and developing an architecture for a Human Digital Twin as part of the MAD research group of Stellenbosch University.

Karel Kruger obtained his PhD from Stellenbosch University, South Africa. He is a lecturer at the Department of Mechanical and Mechatronic Engineering and is a co-leader of the MAD research group.

Anton Basson obtained his PhD in Aerospace Engineering at Penn State University in 1992. He is a Professor in Mechanical Engineering and co-leader of the MAD research group.

Cybersecurity Considerations for Industrie 4.0

A.J.H. Redelinghuys[1], A.H. Basson[1], K. Kruger[1]
[1]Department of Mechanical and Mechatronic Engineering, Stellenbosch University.

Abstract
Cyberattacks are increasing with malicious attackers trying to gain unauthorized access to protected data and control over production facilities. In this paper, the risks of cybersecurity in the context of Industrie 4.0 are presented. Further, the main risk management considerations to prevent security breaches are outlined, as well as the best practices that can be adopted when incorporating cybersecurity. Lastly, the paper discusses the cloud computing model formulated by the National Institute of Standards and Technology and the latest Security as a Service cloud service as defined by the Cloud Security Alliance.

Keywords
Industrie 4.0, Cyber-Physical Production Systems, Security

1 INTRODUCTION

The fourth industrial revolution, also known as Industrie 4.0, builds on rapid increases in capabilities in information and communication technologies (ICTs), the ubiquitous internet and the rise of big-data. Businesses will incorporate machinery, warehousing systems and production facilities as internet-enabled Cyber-Physical Systems (CPS) onto global networks [1].

More companies and organisations are moving towards the integration of business models with the Internet of Things (IoT) and cloud services. The integration of the internet into business opportunities and environments increases the demand for safety and security. It is anticipated that by 2021 an amount of $3.1 billion be spent on safety and security, mainly due to the Industrial Internet of Things (IIoT) [2].

The integration of sensors into global supply networks and data transfer over networks may lead to an increase in security breaches. In Industrie 4.0, which is closely related to Cyber-Physical Production Systems (CPPS), the raising interconnectedness between systems and devices, as well as the increasing complexity of systems and processes, enlarge the attack surface and increase risks [3].

About 82% of manufacturers have experienced phishing attacks in 2017 [4] and manufacturing industries are increasingly being targeted, not just by malicious attacks, but also by competing companies and nations. The traditional focus of manufacturing technology has been on performance and safety, but this leads to major security gaps [5].

Even ICT devices that are generally considered to be reliable and robust, such as programmable logic controllers (PLCs), are susceptible to attacks. Stuxnet is an example [6].

This paper is aimed at outlining cybersecurity challenges, solutions and best practises in the context of Industrie 4.0.

A **cyber-attack** is here defined as an attack against a computer system, network or internet-enabled application or devices using various tools, such as malware, ransomware, etc. [7].

2 CYBERSECURITY THREATS

The security threats that Industrie 4.0 may introduce, can broadly be grouped into the following categories [1] [8]:

Data loss or corruption - Cyber-attacks may entail that the data of a CPPS be removed or corrupted. Recent ransomware attacks are examples of these threats.

Intellectual property breaches - Competing companies (and even nations) engaged in corporate espionage may gain unauthorised access to confidential data.

Denial-of-Service (DoS) - It is the process of making a system or application unavailable. Some of these types of attack may include the disabling or destroying of a sensor in a system, corrupting a process with malformed input data to the server, etc. Interconnected systems and processes are vulnerable to DoS attacks, which can cause operational downtime.

A whole production system can be disrupted by hackers exploiting software vulnerabilities in system components and therefore, resulting in system downtime [1]. According to the Business Insider, a global cyber-attack caused widespread disruption at Renault-Nissan manufacturing facilities, initiated by the WannaCry ransomware worm attack [9].

Major cyber-attacks in 2017, as mentioned by [4], include an AWS account hijack, where Uber drivers and customers details have been compromised. In

October 2017, the UK's National Lottery was brought to a hold by a Distributed Denial of Service (DDoS) attack, preventing people from buying lottery tickets.

Another classification of attacks on modern manufacturing technologies is [5]:

Traditional attacks - A hacker gaining unauthorized access to sensitive systems and data.

Advanced malware - A type of attack that is increasingly common in manufacturing and increasingly disruptive. In an era of the ubiquitous internet connectivity, this malicious software infiltrates weak systems and hardware and then spreads itself to other systems.

Internal threats - This type of threats include malicious insiders stealing companies' intellectual property and other confidential information. This may result in a loss of competitive advantage.

3 CYBERSECURITY RISK MANAGEMENT

Risk management is the ongoing process of identifying, assessing and responding to risk. Risk is a combination of the likelihood that an event will occur and the severity of the resulting impact. Organizations can determine the acceptable risk for achieving its organizational objectives and can express this as their risk tolerance.

Cybersecurity should become an integral part of the strategy, design and operations of CPPS, considered from the beginning of any new initiative [10]. This leads to the concept of **smart security**, which implies implementing preventative security policies, rather than responsive procedures.

It will only be possible to implement and adopt Industrie 4.0 if the following two aspects are accepted [11]:

- Security-by-Design needs to be implemented as key design principle. All aspects relating to security in a manufacturing system or process need to be designed and incorporated into new and old systems from the outset.
- IT security strategies, architectures and standards need to be developed and implemented to a high degree of confidentiality, integrity and availability between interconnected systems and processes.

Security provides the basis for information privacy, such as protection of individuals against infringements of personal data rights and it also enables the protection of intellectual property rights. Safety and security issues are currently raised reactively and, if Industrie 4.0 is to be adopted, a more proactive approach to safety and security will need to be considered [11].

The following subsections outline some of the main risk management considerations:

3.1 Security-By-Design and Testing

IT security and privacy protection need to be considered during the design phase of intelligent production plants, processes and services in order to protect Industrie 4.0 from downtimes and attacks. Currently, IT systems are typically only tested after the final design has been developed and security measures are added afterwards.

Several tools and methods for secure software design, development, testing and maintenance already exist, and these tools were developed to identify and avoid vulnerabilities in the complete product lifecycle. It is therefore necessary for the development of standards and test tools to meet the requirements of Industrie 4.0. These standards may include encryption, authentication and authorisation, security monitoring and incident response.

Test alternatives and significant reference numbers are ways to evaluate the protection against cyber-attacks of a system [3]. A challenge with using test alternatives is that attackers may mask error signals with transparent views of the application, presenting a positive status of processes or sensor data.

3.2 Human Safety and Security

One of the main concerns regarding cyber threats is the lack of knowledge about the risks involved in the event of an attack. With the growing demand for interconnected systems and processes, the attack area increases for cyber-attacks and may also increase the risk to the people using or controlling certain systems and processes.

Software-based protection and security controlling solutions need to be implemented in CPPS' processes and executed in real time to protect human life, the production system and the process.

Industrie 4.0 will provide more interesting, flexible and self-determined forms of working for the future worker in manufacturing environments. Personnel should therefore be equipped with the necessary training in IT security.

3.3 Data Security

Organizations will need to consider what data should be shared and how to protect systems and underlying data that may be proprietary or have privacy risks. Data loss prevention solutions using encryption algorithms to protect high value data assets should be considered as a security approach.

Organizations may want to protect certain data to gain competitive advantage. They may also be subject to regulations that limit the type of information able to be shared. Robust cryptologic support, hardware authentication, and attestation should be provided by incorporating trusted platform modules or hardware security modules [10].

3.4 Access control

Security measures such as access control through authentication mechanisms, cryptographic algorithms and behavioural analysis will be required to address these risks.

3.5 Intelligent Cybersecurity

Cyber-attacks are becoming more intelligent, making it more difficult and complex for companies and organisations to detect or prevent cyber threats. There is an urgent need for systems to be able to search out and rectify code error and vulnerabilities, as well as defend against incoming attacks [12]. This section briefly describes some recent developments in this context.

Automatic Exploit Generation (AEG), the first end-to-end system for fully automatic exploit generation [13], was an award-winning bot at the Cyber Grand Challenge in 2016. This bot can automatically find vulnerabilities and generate exploits. If code error is found in a system, AEG is also able to secure the vulnerability.

Researchers from MIT's Computer Science, Artificial Intelligence Laboratory (CSAIL) and machine-learning start-up PatternEx demonstrated an artificial intelligence platform, called AI2, that can predict cyber-attacks [14]. They claim that this artificial intelligence platform is able to predict cyber-attacks significantly better than existing security systems by incorporating human input [12].

AI2 is able to predict cyber-attacks by examining current data and detect suspicious activity by clustering data into patterns using machine-learning algorithms. These patterns are then analysed by security experts and in the event of a confirmed attack, the data are fed back into the AI2 [14]. Active Contextual Modelling is the continuous feedback between human analysis and the AI2 system. AI2 is therefore able to learn in real-time, which will improve accuracy of future cyber-attack predictions.

Another development particularly relevant to CPPS is the combination of a digital twin and industrial control systems (ICS), called a Digital Ghost, to prevent cyber-attacks. This initiative was initiated by General Electric and is set to use physics to prevent attacks on ICSs by sensing anomalies in processes [15].

4 CYBERSECURITY BEST PRACTICES

The National Institute of Standards and Technology (NIST) formulated a "Framework for Improving Critical Infrastructure Cybersecurity" to help organisations to identify and prioritize actions for reducing cybersecurity risk. The framework is a tool for aligning policy, business and technological approaches to managing that risk. The “Framework Core” consists of five concurrent and continuous high-level functions that must be developed and implemented. Similarly, according to Deloitte [5], a cyber-defence must have three key characteristics to be effective and well balanced. The following bullets give the NIST functions, with the corresponding Deloitte key characteristic in italics:

- **Identify** - An organizational understanding to manage cybersecurity risk to systems, assets, data and capabilities.
- **Protect** - Appropriate safeguards to ensure delivery of critical infrastructure services by limiting or containing the impact of a potential cybersecurity event. ***Secure*** *- Focus protection around the risk-sensitive assets at the heart of the organization's mission.*
- **Detect** - Appropriate activities to identify the occurrence of a cybersecurity event. ***Vigilant*** *- Establish threat awareness throughout the organization and developing capacity to detect patterns of behaviour that may indicate or predict compromise of critical assets.*
- **Respond** - Appropriate actions to initiate when a cybersecurity incident is detected.
- **Recover** - Appropriate activities to maintain plans for resilience and to restore any capabilities or services that were impaired due to a cybersecurity incident. ***Resilient*** *- Have the capacity to rapidly contain the damage and mobilize the diverse resources needed to minimize impact.*

The IEC 62443 standard for "Industrial communication networks – Network and system security" provides guidelines for securing design. Schneider Electric [16] mentions in the “Practical Overview of Implementing IEC 62443 Security Levels in Industrial Control Applications”, six cybersecurity best practices for a defence-in-depth strategy to help protect connected manufacturing plants from cyber-attacks:

- **Create a Security Plan:** Organisations should develop a profile of the network by creating a detailed inspection of all the equipment connected to the industrial network by creating network maps. Organisations need to identify how processes and assets are linked on the network map and how critical these assets and processes are according to the organisation's operation ability.
- **Separate Networks:** Dividing a network into different zones such as an enterprise, plant, process, and field zones.
- **Perimeter Protection:** Securing remote access to prevent unauthorized access and control of the devices and sensors in the industrial facility.
- **Network Segmentation:** Splitting the network into smaller segments to separate systems and applications from each other. Network traffic can therefore be isolated and limited. The available attack options are limited by limiting communication through the network.

- **Device Hardening:** Improve ICS's ability to withstand cyberattacks by adding features to harden the device.
- **Monitor and Update:** Monitor the network activity to detect potential threats, and protect devices from code flaws by updating and releasing software to address vulnerabilities or to add security features. Strategies will need to be developed to update software of affected devices within a manufacturing environment.

Andrew Cooke, Head of ICS Consultancy at Airbus CyberSecurity, mentions five cybersecurity best practices to help protect connected manufacturing plants from cyber-attacks (adapted from [17]):

- **Default Credentials**: Default username and passwords set by organisations tend to be major security risks as this allows easy access to attackers. Organisations need to ensure that credentials have been reset before connecting a device onto a network.
- **Patching**: Organisations need to protect devices from code flaws by updating and releasing software to affected devices. They will need to develop strategies to roll out updated software to affected devices within an environment.
- **Network Maps**: Organisations should come to know the profile of the network, including the map of the devices on the specific network. Therefore, defining how operations and the IoT are connected and the risks that are involved within the process.
- **Asset Identification**: Organisations need to determine what processes and assets are critical according to the organisation's operation ability. They also need to compare and correlate detailed processes to a network map. Risk and security can only be managed if the devices are known.
- **Upskilling**: People need to be aware of the risks involved with cyber-attacks and how it may affect business environments. People need to be educated to understand the connectedness of devices on the global network and the risks involved if an attack should occur.

5 CLOUD COMPUTING AND SECURITY AS A SERVICE

The increase of interconnected systems and processes, with the industry shift towards CPPS, requires large data storage space and computational intelligence to manage these large data sets. Cloud-based services are therefore closely associated with Industrie 4.0 and are seen to be the future for the provision of a wide range of IT services [18]. In this section the integration of security using the different cloud service models is considered.

Various definitions of cloud computing exists, but many governing bodies and professional organisations, such as the European Network and Information Security Agency (ENISA), The British Standards Institution (BSI) and the Cloud Security Alliance (CSA), have referred to the definition developed by NIST [18] [19]:

"Cloud Computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management or service provider interaction."

The cloud computing model formulated by NIST [19] and also used by CSA [20] is described by the five essential characteristics, three cloud service models and four cloud deployment models. The different cloud-based deployment models are private, community, public and hybrid. Private cloud infrastructure is provisioned for exclusive use by a single organisation. It may be managed by the organisation or by a third party. Community cloud infrastructure is provisioned for exclusive use by a specific community of consumers that have shared concerns (e.g. mission, security requirements, policy or compliance considerations). Public cloud infrastructure is made available to the general public or a large industry group. Hybrid cloud infrastructure comprises multiple cloud infrastructures that remain unique entities, but are bound together by standardised or proprietary technology that enables data and application portability [18] [19] [20]. Organisations may typically set up their systems with a hybrid cloud infrastructure architecture to protect underlying data from privacy risks.

The cloud service models described by NIST are [19] [20]:

- Software as a Service (**SaaS**) is an application that is managed and hosted by the provider. Consumers can access these applications by web browsers, mobile applications or lightweight client applications. Examples of this service includes Google Apps (e.g. Gmail), Microsoft Office 365, etc.
- Platform as a Service (**PaaS**) provides development or application platforms, such as operating systems, databases, programming language execution environments, web servers or proprietary application processing.
- Infrastructure as a Service (**IaaS**) provides the consumer with fundamental computing infrastructure, such as computation, network or storage. Examples of this service may include Amazon EC2, Windows Azure, Google Compute Engine, etc.

Cloud computing is a shared technology model. Different organisations are responsible for implementing and managing different parts of the stack. Security responsibilities are therefore also distributed across the spectrum as seen in Figure 1.

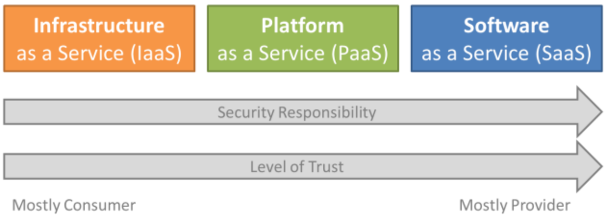


Figure 1 - Figure 2 Security Responsibility Over the Architecture Stack (adapted from [20])

While the three most common service models, SaaS, PaaS and IaaS, are well known and widely used, the latest service defined by the CSA, is Security as a Service (**SecaaS**) [20] [21]. SecaaS security capabilities are provided as a cloud service. The most basic example of SecaaS is using anti-virus software over the internet. Potential benefits (adapted from CSA and [22]) include:

Cloud-computing benefits: Reduced capital costs, agility, redundancy, high availability and resiliency. Security tools or software need to be updated frequently to prevent new cyber threats from entering business and manufacturing environments. The benefit of as-a-service tools is that they are always equipped with the latest software.

Staffing and expertise: Many organisations are not solely focused on security or specific security domains. SecaaS may therefore provide the benefit of extensive domain knowledge and research. This is provided by IT professionals and security experts.

Intelligence sharing: SecaaS providers protect multiple organisations simultaneously and therefore grant the opportunity to share experience across a global network. An example is when malware is detected at one organisation, it can immediately be protected at other organisations who are using the same SecaaS.

Deployment flexibility: Users can have access to these tools instantly. SecaaS offerings are provided on demand and can handle more flexible deployment models, such as supporting distributed locations.

Insulation of clients: SecaaS has the ability to intercept some attacks before they reach the organisation. An example is cloud-based web application firewalls and spam filtering.

Scaling and cost: Consumers do not have to buy hardware or software licenses. Instead, the cloud model provides the consumer with a "Pay as You Grow" model.

The CSA identified the major categories offered by SecaaS. These SecaaS offerings encompass security software that are hosted on the cloud. A selection of these solutions is presented in Table 1.

Domain	Protective	Preventive	Detective	Reactive	SaaS	PaaS	IaaS
Identity and access management	X	X			X	X	
Data loss prevention		X			X	X	
Web security	X		X	X	X	X	
Email security	X		X	X	X		
Security assessment			X		X	X	X
Intrusion management	X		X	X	X	X	X
Encryption	X				X	X	X
Disaster recovery and business continuity	X		X		X	X	
Network security	X		X	X	X	X	X
Security information and event management			X		X	X	

Table 1 - Security as a Service Architecture (adapted from [21])

Some potential concerns may arise with the adoption of SecaaS, such as lack of visibility, handling of confidential data, data leakage, vendor lock-in and migration to SecaaS. The consumer is responsible for evaluating the possibilities that each SecaaS provider may offer and linking the as-a-service tool to their business model.

6 CONCLUSIONS

A major challenge in adopting Industrie 4.0 as business model is not only to protect humans and machines from malicious attacks, but also to deliberately invest in IT security and strive towards a threat prevention through prediction environment. In the context of Industrie 4.0 and big data on cloud computing, with the associated increases in the surface area for malicious attacks, organisations need to evaluate their business models and develop risk management strategies. These strategies should aim to prevent malicious attacks and, in the event of an attack, be able to rectify the situation.

This paper has outlined the main challenges and solutions in the context of Industrie 4.0, and also best practises that can be adopted when incorporating cybersecurity in systems and processes.

Malicious attackers always try to stay one step ahead of security professions, but as industry strives towards the smart factory, digital twins and the IoT, businesses must continuously strive towards integrating smart security into old and new systems and processes. Recent developments in self-learning cybersecurity platforms hold promise to decrease the risk of malicious attacks.

7 REFERENCES

[1] Pereira, T., Barreto, L. & Amaral, A. 2017. Network and information security challenges within Industry 4.0 paradigm. Procedia Manufacturing. 13:1253-1260.

[2] McCaskill, S. 2018. IoT security spend to reach £1bn in 2018. [Online]. Available: https://www.techradar.com/news/iot-security-

spend-to-reach-pound1bn-in-2018. [2018, May 10].

[3] Waidner, M. & Kasper, M. 2016. Security in Industrie 4.0 - Challenges and Solutions for the Fourth Industrial Revolution. In Design, Automation & Test in Europe Conference & Exhibition (DATE). 1303-1308.

[4] Check Point. 2018. Security Report. Tel Aviv-Yafo.

[5] Deloitte. 2018. Global Cyber Executive Briefing. [Online]. Available: https://www2.deloitte.com/global/en/pages/risk/articles/Manufacturing.html#. [2018, May 03].

[6] Falliere, N., Murchu, L.O. & Chien, E. 2011. W32.Stuxnet Dossier. Symantec-Security Response. Version 1. (February 2011):1-69.

[7] Check Point. [S.a.]. What is a Cyber Attack? [Online]. Available: https://www.checkpoint.com/definition/cyber-attack/. [2018, May 16].

[8] BSI. 2012. BSI-Analysen zur Cyber-Sicherheit.

[9] Frost, L., Tajitsu, N. & Reuters. 2017. Renault-Nissan is resuming production after a global cyberattack caused stoppages at 5 plants. [Online]. Available: http://www.businessinsider.com/renault-nissan-production-halt-wannacry-ransomeware-attack-2017-5?IR=T. [2018, May 18].

[10] Waslo, R., Lewis, T., Hajj, R. & Carton, R. 2017. Industry 4.0 and cybersecurity. [Online]. Available: https://www2.deloitte.com/insights/us/en/focus/industry-4-0/cybersecurity-managing-risk-in-age-of-connected-production.html. [2018, May 03].

[11] Kagermann, H., Wahlster, W. & Helbig, J. 2013. Securing the future of German manufacturing industry: Recommendations for implementing the strategic initiative INDUSTRIE 4.0. acatech, Final report of the Industrie 4.0 Working Group.

[12] Fagella, D. 2017. Artificial Intelligence and Security: Current Applications and Tomorrow's Potentials. [Online]. Available: https://www.techemergence.com/artificial-intelligence-and-security-applications/. [2018, May 23].

[13] Avgerinos, T., Kil Cha, S., Lim Tze Hao, B. & Brumley, D. 2011. AEG: Automatic Exploit Generation. Communications of the ACM. 57.

[14] Gupta, A. 2017. Meet AI2, Artificial intelligence based Security System that is 85% accurate. [Online]. Available: https://news.thewindowsclub.com/meet-ai2-artificial-intelligence-based-security-system-that-is-85-accurate-88163/. [2018, May 23].

[15] Dignan, L. 2017. GE aims to replicate Digital Twin success with security-focused Digital Ghost. [Online]. Available: https://www.zdnet.com/article/ge-aims-to-replicate-digital-twin-success-with-security-focused-digital-ghost/. [2018, May 23].

[16] DesRuisseaux, D. 2018. Practical Overview of Implementing IEC 62443 Security Levels in Industrial Control Applications (White Paper). Lake Forest: Schneider Electric Software.

[17] The Engineer. 2018. Rising to the Industry 4.0 cybersecurity challenge. [Online]. Available: https://www.theengineer.co.uk/industry-4-0-cybersecurity/. [2018, May 18].

[18] Schaefer, T., Hofmann, M., Loos, P. & Fettke, P. 2014. Selecting the Right Cloud Operating Model: Privacy and Data Security in the Cloud. ISACA. 3.

[19] Mell, P. & Grance, T. 2011. The NIST Definition of Cloud Computing: Recommendations of the National Institute of Standards and Technology. NIST Special Publication. 145:7.

[20] Mogull, R., Arlen, J., Lane, A., Peterson, G., Rothman, M. & Mortman, D. 2017. Security Guidance: For Critical Areas of Focus In Cloud Computing v4.0.

[21] Wlosinski, L.G. 2013. IT Security Responsibilities Change When Moving to the Cloud. ISACA. 3:1-4.

[22] Aldorisio, J. 2018. What is Security as a Service? A Definition of SECaaS, Benefits, Examples, and More. [Online]. Available: https://digitalguardian.com/blog/what-security-service-definition-secaas-benefits-examples-and-more. [2018, May 22].

8 BIOGRAPHY

Anro Redelinghuys is a master's candidate in Mechatronic Engineering at the University of Stellenbosch, South Africa.

Anton Basson obtained his PhD in Aerospace Engineering at Penn State University in 1992. He is a Professor in Mechanical Engineering and co-leader of the Mechatronics, Automation and Design Research Group at University of Stellenbosch, South Africa.

Karel Kruger obtained his PhD from Stellenbosch University, South Africa. He is a lecturer at the Department of Mechanical and Mechatronic Engineering and is part of the Mechatronics, Automation and Design research group.

Robot Automation in Control Cabinet Assembly

S. Spies[1], M. Bartelt[1], A. Hypki[1], B. Kuhlenkötter[1]
[1]Chair of Production Systems, Faculty of Mechanical Engineering, Ruhr-Universität Bochum, Germany

Abstract
The control modules of machines and systems are housed in control cabinets varying in size and type: the electrical components are mounted on mounting rails and are connected with wires and cables bundled into cable ducts. Currently, control cabinets are predominantly produced and constructed by small and medium-sized enterprises (SMEs) in manual, time-consuming and expensive assembly processes. Available solutions to automatize single production steps are expensive and rarely used by SMEs. Cost-efficient and flexible solutions are still missing particularly for the wiring task, which takes about half of the overall production time. Our approach presented within this paper is to develop a specialized tool that is able to handle pre-configured wires. The tool integrates concepts of cyber-physical systems and can be handled by an affordable industrial robot. Due to the possibility of using pre-configured wires, the solution can be integrated into the production process easily and with low effort, making its application economically viable, especially for SMEs.

Keywords
Control cabinet, robot-based wiring, cabling tool

1 INTRODUCTION

Machines and production systems contain a huge number of electromechanical components, e.g. conveyor belts, industrial robots, or gripping modules. In order to make them act in the desired way, it is necessary to incorporate several control modules. Examples of such modules are simple input/output devices, programmable logic controllers (PLC) or terminal strips. A control cabinet is used to mount and wire these components in a structured way. The configuration of the control cabinet, i.e. its size, the included components, their position and the wiring, depends to a high degree on the corresponding production system. Hence, manufacturers of such control cabinets have to handle a huge variance and small lot sizes. Even a lot size of one is very common in this domain as almost every product is unique and has to be individually designed. As a result, the production process is mostly done manually with only few options for automation [1].

According to Großmann et al., the manufacturing process of control cabinets splits up into nine single steps [2]. The first step is the design and planning of the cabinet. Afterwards, a mechanical preparation of the mounting plate is performed and the mounting rails are prepared. In steps four to six, the components are labelled and mounted to the rails and the accessories are assembled, e.g. bridges and connectors. The seventh step includes the assembly of the rails to the mounting plate and the assembly of the plate into the cabinet. In the last two steps, the wires become pre-configured and the final wiring is done. Analyzing the single steps, it has been found that the wiring requires about 50% of the overall manufacturing time [3]. Hence, a solution for an automatic wiring of pre-configured wires can reduce the manufacturing time significantly.

2 PRODUCTION OF CONTROL CABINETS

Automatizing the manual production is an efficient means to reduce production costs. Correspondingly, the worldwide stock of operational industrial robots increases by about 10% each year [4]. However, the production of control cabinets, which is a central component of almost any automated production system, is still affected by a manual workshop-oriented production. We can identify two major reasons for this fact: the complexity of programming production systems and the high variance and low lot size of the control cabinets.

The first reason for a missing automation is the complexity of programming automated production systems. This complexity is a major hurdle especially for small and medium-sized enterprises (SME) [5], where most manufacturers of control cabinets belong to. Such companies provide a profound knowledge of their manufacturing processes but they do not have the required experts to apply and to program the required robots. They may, of course, use the help of external experts; however, this practice is not economical, e.g. if a production process needs to be changed.

The second reason relates to the high variance and low lot size. As stated above, a control cabinet, i.e. the contained components, is configured particularly for a certain production system or machine. Commonly, there are three different order scenarios:

- Individual orders; such orders are very common for production systems. A production system is designed for a certain product. Hence, the system is highly specialized and the required control cabinet must be individually designed.

- Batch series production; a customer requires a specific number of the same control cabinet. Usually this quantity is up to a few hundreds of cabinets.
- Series production; this kind of order is very common for control cabinets for larger machines. Such machines have a production time of several days or weeks. Hence, the same control cabinet is ordered regularly but in rather low lot sizes.

Even in the two scenarios of series production, a solution for an automated production system for a specific control cabinet is not economical due to the low lot size. As a result, an appropriate production system must be highly flexible and able to produce a huge number of different control cabinet configurations. Of course, there are specific standards for the design of control cabinets. For the shape of a control cabinet, the international standard IEC 60291-3 specifies certain sizes and types of construction. For power transformers, an IEEE standard [6] defines particular design rules. Other examples can be found. Nevertheless, the inner setup, i.e. the number and kind of components, their positions etc. varies, which impedes automated production. A solution for a holistic automated control cabinet production is currently not available on the market. However, mass production of individualized products is a current trend also in other industries. A prominent example is the automotive industry, where customers can configure their cars with a huge number of options. Amongst others, product modularization is a proper means to handle such mass customization [7]. Also concepts summarized in keywords like Industry 4.0 [8], i.e. data exchange in manufacturing technologies and cyber-physical systems, can be adopted to find a solution for an automated production of control cabinets.

The nine single production steps mentioned above can be merged into three phases: the planning, the mechanical processing (e.g. of the mounting plate or rails), and the assembly (e.g. mounting of components and wiring).

In the planning phase, manual actions are predominant. Customers define the required functionalities and requirements of the desired control cabinet. Sometimes, customers even determine the cabinet's complete configuration. With the given information, the control cabinet manufacturer creates appropriate schematics for the mechanical and electrical design of the cabinet. Although this phase is less automated, the creation of a comprehensive data basis for the control cabinet is a main need in order to automate succeeding processes. Because many different software tools are involved within the planning process, it is necessary to establish a certain degree of interoperability between these tools. For this purpose, concepts and demonstrating implementations are being developed within many projects (comp. e.g. [9–11]).

For the different tasks of mechanical processing in the second phase, there are automated solutions available on the market. Examples in this respect are CNC machines, which can be used, for the preparation of the mounting plates and crimping machines, which can produce pre-configured wires fully automatically. The required information, like the type and length of the wire, is provided via a programming interface. Thus, a continuous automation of the cabinet design can be realized for the mechanical processing tasks (comp. [12]). Another example is a machine developed by the Krones company. The machine can construct cable ducts and mounting rails. However, the company does not sell the machine but uses it only for in-house applications [13]. More examples for automated mechanical processing can be found. Nevertheless, due to the specialized focus on the certain automation problem, all these systems have a low flexibility. In addition, the costs for acquisition, operation, and maintenance prevent an economic use for small lot sizes, which are very common for SMEs. With respect to the control cabinet assembly, usually only the preparation of mounting plates and the production of pre-configured wires has sufficient quantities so that a use of such devices would be economical.

In the last phase, the components are mounted and the wiring is done. Even though there is effort spent in the automation of the single tasks, corresponding solutions can only be realized by a considerable technical effort. In addition to a large number of mechanical components, there is need for a sensor system with suitable sensor data processing. Therefore, available automation solutions are expensive and uneconomical for the production of control cabinets with a small lot size. This is particularly true for the wiring process.

One of the most advanced systems for the automated wiring of individual control cabinets is Kiesling's Averex. The necessary development time of five years shows the complexity of the automation task, which has been carried out by a consortium of companies (comp. [14]). The system is susceptible to errors due to the high complexity level and the huge number of moving parts. Only a few prototypes have been built – the system is currently not publicly available on the market. In contrast to that complex and expensive system, we are aiming to develop a cost-efficient and flexible solution by using a sophisticated tool in combination with an industrial robot. Our approach is not intended to realize a full automation of the wiring but it will implement an optimized combination of manual and automated work. Another approach is the utilization of industrial robots to build an assisting system. Such a solution is the aim of the research idea AWeMa [15]. The concept is a digital assistant that may be extended by an adaptive robot system. The system assists a worker in a hybrid environment, e.g. by visualizing the next wiring task or by preparing the wire the worker will installing next. Not only the wiring process but also other assembly processes should be optimized by the approach. However, the idea is currently on a premature level and far from a working solution.

Due to the current manual production process, the use of pre-configured wires is very common. Unless the production process can be automated completely, pre-configured wires are still required during the manufacturing. Hence, an economic automation approach must use these cables instead of a dedicated wire supply. The automated handling of such non-rigid objects has been investigated for several years (comp. e.g. [16, 17]). However, such approaches do consider the complete non-rigid object at all times, which is required for the wiring process in order to prevent errors, e.g. wires stuck to the cabinet.

3 RELATED WORK

In previous works, we have focused on improving the assembly process of components to the mounting rails in order to build terminal strips. Work in this process is related to steps three to six and step nine (comp. section 2). In this context, we have analyzed the current manual assembly process with respect to the potential for automation. The current manual production process of is separated into six tasks. The first task is the terminal assembly. The following tasks are the labeling, circuit bridge and plug assembly, quality inspection, cabling, and finally the packaging. Similar to the control cabinet production, we have implemented a production of individualized configurations with low lot sizes.

Figure 1 – Mounting process of end clamps in a human-robot collaboration scenario.

In the following, we have developed a five-step top-down method [18] that we have used to identify potentials for human-robot collaboration (HRC) in a scenario. As a result, we have realized the assembly of end clamps in a HRC scenario (see Figure 1). These clamps are mounted on either side of a set of components and prevent them to move on the mounting rail. Within the concept, a first step is to mount the end clamps. Afterwards, a worker can mount the corresponding parts between the clamps. We have created a demonstrator that shows the advantages of the automated approach in order to proof the concept [19].

Currently, we are developing a new test bed. Even though it is not a human-robot collaboration, we will use the setup to test and evaluate the tool concept presented in the next section. Next to the industrial robot, we have integrated safety equipment in order to achieve a solution that manufacturers of control cabinets can directly use within their current processes. Figure 2 shows the robot during the mounting process of an end clamp in the new setup.

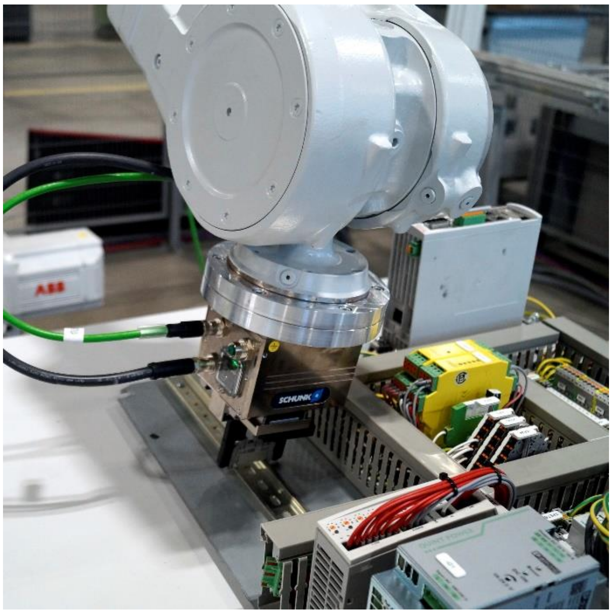


Figure 2 – Image of the current setup.

4 TOOL CONCEPT

A main issue within the wiring process is indeed the use of pre-configured wires. Even though the aim is to automatize almost all wiring, there will remain some wiring tasks that are difficult to automatize economically. Examples are the wiring of components with a high density of plugs or the assembly of wires with special connectors, e.g. connectors that must be mounted in a certain orientation. As a result, the most economical solution will be a mixture of automated and manual wiring. However, within the manual wiring, pre-configured wires are used which the tool has to handle as well in order to avoid additional components.

In order to design a tool that is able to handle pre-configured wires, we have analyzed the assembly process. Related to occurring voltages within the control cabinet, the tool must handle wires with an outer diameter from 2.2 mm (type H05V-K) to 4.6 mm (type H07V-K). The length of the wires is up to 2,000 mm. Wire ferrules are crimped on both ends of the wire. Figure 3 sketches our mechanical concept for the tool design. Two grippers (3) will grip the start and end of a wire. In order to ensure a proper grip, the corresponding claws (4) have a prismatic shape. The gripper that grips the first end of the wire is attached to a wire reel (2) that can coil a wire by a rotational movement. A mechanical structure (1) provides the necessary mechanical connection for the wire reel and the gripper as well as an attachment flange to the industrial robot.

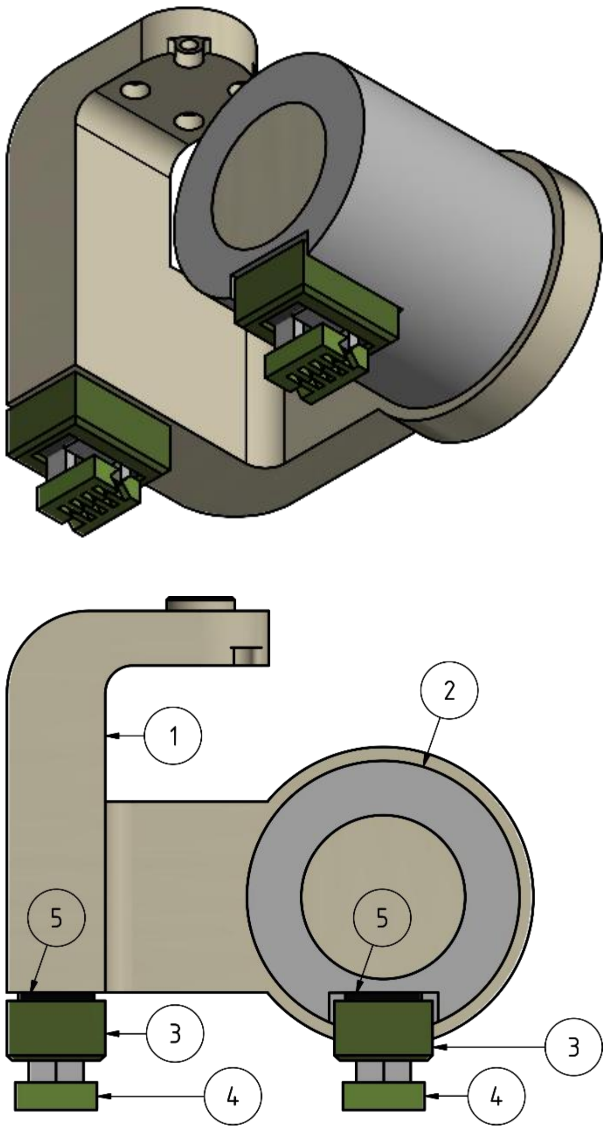


Figure 3 – Sketch of the mechanical tool concept.

What is not shown in the sketch are the required sensor devices for quality assurance as well as additional electronic devices, e.g. a microcontroller that is used to build the cyber-physical system. There are two main aspects to be considered for quality assurance: the assembly of the cable heads and the location of the wire in the duct. For the latter one, a camera in combination with appropriate image processing algorithms are planned to be used. The wire head assembly is checked with force sensors (5) located between the grippers and the mechanical structure or the wire reel. While the assembly requires a force of up to 10 N, the wire must remain mounted when a force of up to 80 N is applied. However, both force values depend on the corresponding wire type.

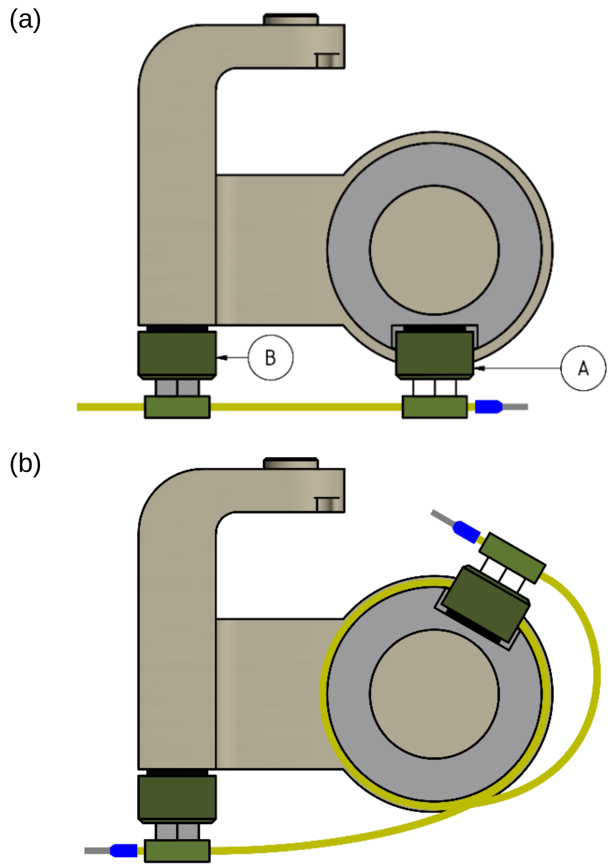


Figure 4 – Take-up sequence to grip and coil a wire.

The pre-configured wires are automatically produced by a crimping machine and then stored in a wire stack. The stack provides a separation of the single pre-configured wires in order to enable the tool to grip and coil a wire. The take-up process consists of four steps:

1. The separated wire is located within the wire stack in such a way that the tool can position its grippers around the wire (comp. Figure 4a).
2. Gripper A will grip the wire just behind the wire ferrules. This is a critical part in the take-up process because too much space between the claws and the ferrule may cause errors within the succeeding wiring process. Gripper B will close in such a way that it will not lose the wire but the wire can be pulled through the claws.

3. The wire reel will coil the wire until the second ferrule reaches gripper B (comp. Figure 4b). As stated above, the type and length of the wire is given within the assembly job. The tool will use this information to coil the wire. However, we might need to include a further sensor device if the remaining space between gripper B's claws and the ferrule is inappropriate.
4. Gripper B can fully grip the wire.

As soon as both ends of the wire are gripped, the robot can start the wiring process. Thereby, the end gripped by gripper B will be handled first. After assembly, the wire can be placed into the duct. Finally, the tool will mount the wire head located in gripper A and will perform a profound quality assurance.

5 CONCLUSIONS

Within this paper we have presented a new approach for a tool that is able to perform a wiring process automatically. Starting with the current manually performed wiring process, we have identified the requirements for such a tool. While information from the planning phase of the control cabinet is provided via an interface, the tool generates an appropriate robot program in order to execute the corresponding wiring task. Thereby, the tool communicates with a wire stack that has provided the necessary pre-configured wires.

The presented approach has already been discussed among experts of manufacturers of control cabinets. Currently, we are planning plug experiments that will provide important input to particular design features, i.e. the gripper claws and the position of the wire within the claws. Afterwards, we will be able to finalize the design and build a first prototype, which will be used to prove the concept viable.

6 ACKNOWLEDGEMENT

This research was funded by the AiF (funding number: ZF 4060720RP7) within the Central Innovation Programme (ZIM) from the Federal Ministry for Economic Affairs and Energy (BMWi) following a decision of the German Bundestag.

7 REFERENCES

[1] P. Boyer, "Automating Control Panel Assembly," *Assembly Magazine*, no. 10, https://www.assemblymag.com/articles/90525-automating-control-panel-assembly, 2012.

[2] C. Großmann, O. Graeser, and A. Schreiber, "ClipX: Auf dem Weg zur Industrialisierung des Schaltschrankbaus," in *Handbuch Industrie 4.0 Bd.2*, B. Vogel-Heuser, T. Bauernhansl, and M. ten Hompel, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2017, pp. 169–187.

[3] P. Tempel, F. Eger, and A. Verl, *Reichlich Potential zur Effizienzsteigerung: Fertigung von Schaltschränken im Wandel moderner Produktionsprozesse.* [Online] Available: https://www.schaltschrankbau-magazin.de/artikel/reichlich-potential-zur-effizienzsteigerung/. Accessed on: Nov. 21 2017.

[4] International Federation of Robotics, "Executive Summary World Robotics 2017 Industrial Robots," 2017. [Online] Available: https://ifr.org/downloads/press/Executive_Summary_WR_2017_Industrial_Robots.pdf. Accessed on: Aug. 12 2018.

[5] Z. Pan, J. Polden, N. Larkin, S. van Duin, and J. Norrish, "Recent progress on programming methods for industrial robots," *Robotics and Computer-Integrated Manufacturing*, vol. 28, no. 2, pp. 87–94, 2012.

[6] *IEEE Standard for Control Cabinets for Power Transformers*, 2011.

[7] V. Modrak and Z. Soltysova, "Process modularity of mass customized manufacturing systems: Principles, measures and assessment," *Procedia CIRP*, vol. 67, pp. 36–40, 2018.

[8] Federal Ministry of Education and Research (BMBF), *The new High-Tech Strategy: Innovations for Germany*. Berlin: trigger.medien gmbh, 2014.

[9] M. Bartelt, A. Schyja, and B. Kuhlenkötter, "More than a Mockup," *Prod. Eng. Res. Devel.*, vol. 8, no. 6, pp. 727–735, 2014.

[10] M. Bartelt, J. Stecken, and B. Kuhlenkötter, "Microservice for deriving digital twins," *VDI Berichte*, no. 2330, pp. 679–690, 2018.

[11] S. Biffl, E. Maetzler, M. Wimmer, A. Lueder, and N. Schmidt, "Linking and versioning support for AutomationML: A model-driven engineering perspective," in *2015 IEEE 13th International Conference on Industrial Informatics (INDIN)*, Cambridge, United Kingdom, 2015, pp. 499–506.

[12] H.-R. Koch and J. von Kiesling, "Automatisches Verdrahten," *Schaltschrankbau Magazin,* https://www.schaltschrankbau-magazin.de/artikel/automatisches-verdrahten/, 2015.

[13] C. Großmann, "Schnell und fehlerfrei fertigen," *IT&Production*, no. 10, pp. 30–31, 2015.

[14] [Online] Available: https://www.rittal.com/de-de/content/de/unternehmen/presse/pressemeldungen/pressemeldung_detail_62336.jsp. Accessed on: Nov. 21 2017.

[15] Kompetenzzentrum Mittelstand GmbH, "AWeMa: Adaptive hybride Werkerassistenz in der Schaltschrank-Montage," 2017. [Online] Available: https://www.kme-mittelstand.de/project-pdf/46.pdf. Accessed on: Aug. 13 2018.

[16] T. Tang, C. Liu, W. Chen, and M. Tomizuka, "Robotic manipulation of deformable objects by tangent space mapping and non-rigid registration," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems*

(IROS), Daejeon, South Korea, 2016, pp. 2689–2696.

[17] D. T. Pham, D. Henrich, and H. Wörn, *Robot Manipulation of Deformable Objects*. London: Springer London, 2000.

[18] M. Linsinger, M. Sudhoff, K. Lemmerz, P. Glogowski, and B. Kuhlenkötter, “Task-based Potential Analysis for Human-Robot Collaboration within Assembly Systems,” in *Tagungsband des 3. Kongresses Montage Handhabung Industrieroboter*, T. Schüppstuhl, K. Tracht, and J. Franke, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2018, pp. 1–12.

[19] M. Linsinger, J. Kutschinski, J. Stecken, and B. Kuhlenkötter, “Mensch-Roboter-Kollaboration im Schaltschrankbau - Konzept zum Setzen von Endhalterklemmen bei der Klemmleistenmontage,” in *Automation 2018*, 2018, pp. 95–107.

8 BIOGRAPHY

Stefanie Spies is research assistant at the Chair of Production Systems (LPS) at Ruhr-Universität Bochum. She is working within the research project RoboSchalt.

Matthias Bartelt obtained his PhD degree from Ruhr-Universität Bochum. He is currently post-doctoral fellow at the Chair of Production Systems. His specific focus of research interest is on the application of cyber-physical systems.

Alfred Hypki is senior engineer at LPS. His specific areas of research are human robot collaboration as well as robot offline programming and simulation.

Bernd Kuhlenkötter is head of the chair of Production Systems. Related to the presented research he is developing certain automation scenarios in the field of control cabinet assembly.

Collaborative Robotics as a Success Factor in Electronics Manufacturing

F. Ranz[1,2], T. Komenda[1,2], W. Sihn[1,2]
[1]Fraunhofer Austria Research GmbH, Vienna, Austria.
[2]Vienna University of Technology, Institute of Management Sciences, Vienna, Austria.

Abstract
Semiconductor electronics is a highly globalized, competitive, high-volume and low-cost manufacturing sector that has experienced continuous outsourcing to specialized Electronic Manufacturing Services (EMS) and offshoring during the last decades. While large portions of the value stream have been automated, EMS struggle to rationalize production of smaller batches in the 5 to 7 digits range due to the immense complexity and invest that comes with equipment for automated PCB assembly, testing and packaging. Thus, these processes are often still carried out manually - implicating high labor cost per piece as well as monotonous, tedious work for operators. This paper outlines the core results of an industry-oriented research project determining the potential of collaborative lightweight robotics for surface-mounted device (SMD) manufacturing. The authors show that collaborative robots are an automation technology meeting the requirements with regard to economics, flexibility and ergonomics. The paper presents not only an analytical approach but also the validation of its results by implementing a pilot application. Coping with the idea of collaboration, technologies available for SMD handling and manipulation had to be adapted to be considered safe for direct human-robot interaction and, at the same time, achieve process velocities that meet the desired customer takt. The authors expound generic fields of applications for human-robot collaboration in SMD manufacturing that could be identified and set out an approach for and the results of the profitability analysis. With regard to the pilot implementation, an exemplary collaborative work system layout is presented. To prove conformity with relevant ISO norms and specifications, effective forces for collisions between human and robot have been evaluated experimentally.

Keywords
Human-Robot-Collaboration, SMD handling, safety, system design

1 INTRODUCTION

In the age of globalisation, competitiveness is becoming the ultimate factor for securing production at a certain location. It requires high quality product manufacturing with state-of-the-art technology. Surface mounted devices (SMD) are high quality products and their production is mostly automated. However, certain conditions such as varying batch sizes, high number product variants and complex processes make it hard to automate lines for specific SMD products. Yet, manual tasks are characterised by high costs, monotonous and tedious work as well as inconsistent quality. By combining the specific capabilities of means of automation and manual operators, these drawbacks can be met while facing the challenges of a variable degree of automation.

This paper describes the approach for designing a human-robot collaborative (HRC) system step-by-step and showcases an implemented industrial application for SMD packaging for an electronics manufacturer (EMS). The presented solution shows high potential of multiplication to further work systems. It is also an example of how collaborative applications secure the existence of manufacturing companies in Austria.

2 STATE OF THE ART

In recent years, power- and force-limited robots were launched, which makes it easier to integrate them into work systems that have so far been operated primarily by humans. When humans and robots share the same workspace, so that they can come into contact with each other, this can be called HRC [1]. By combining strengths of robots such as accuracy and endurance with strengths of humans such as cognition and versatility [2] in a joint environment, a production process that is more flexible than a fully automated one and causes less labour cost and burden than a fully manual is the result. However, a HRC system portrays a challenging and complex planning and design task for the engineer [3] that requires various considerations, trade-offs and decisions - such as the economic *degree of automation*, finding a *task allocation* that matches the requirements of the operators and designing a system that is inherently safe in accordance with safety regulations. Due to the relative novelty of the technology, practical experience and guidelines for the design of HRC systems are rare and in demand [4]. Existing planning approaches focus on certain aspects of the entire HRC planning and design problem. To achieve a more holistic approach and certain industrial relevance at the same time, the authors have condensed the manifold design decisions and their respective options into a morphology [11]. Still, a complete industrial HRC design approach could not be identified. This paper is supposed to close this gap on the example of SMD manufacturing.

3 SYSTEM DESIGN APPROACH

Even though powerful robotic technologies are available, a number of manufacturing tasks are not automated yet [6]. This is not necessarily due to the complexity of the task itself, but rather high investment costs, which are hard to predict since practical experience in the design of HRC systems is still rare as well as normative guidelines and restrictions may lead to iterative design changes.

While the required hardware cost might seem reasonable, engineering, validation and implementation efforts in average quadruplicate the required investment [5] – making the project potentially uneconomical.

The reason for these high costs is the complexity of the robotic cell design and integration. Bouchard states, that there are four reasons for this [7]:

i. Robots evolved to suit high-volume, low-mix applications, but nowadays time and money to set up the robot might be more important. However, easy and low-cost integration might oppose still needed precision and reliability.

ii. There is a lack of standards in the robotics industry. Each robot manufacturer has its own operating system and programming language. Even though, lightweight robots were designed to facilitate programming by implementing hand-guidance, programming skills are still required when the application involves e.g. sensors and logical interdependencies.

iii. Robots deal with the physical world and in times of full automation, the physical world was clear, predictive and logically structured. Today, the physical world we want robots to interact with human beings is rather unstructured. Furthermore, cycle time is not easily predictable anymore, as it depends on the velocity set by the risk analysis. And the risk analysis can again only be conducted with a physical available application. It’s a chicken-and-egg-problem.

iv. Manufacturers lack employees who are skilled in robotics.

While the developed approach might not be able to eliminate these challenges, it at least provides a procedure that helps identifying applications which likely will be economic, reduces implementation time and thereby the risk of misinvestment.

The system design approach developed consists of four crucial phases: (1) Analysis of potentials, (2) technical concept, (3) safety & security concept, (4) validation as well as implementation. One key factor of this approach is that the conceptual work and the implementation of demonstrators is conducted on laboratory level before transfer to the production environment. In this way, the company is not affecting its actual operations.

The Pilot Factory Industry 4.0 at Vienna University of Technology is serving as such lab that allows companies to develop and test their HRC applications with existing hard- and software before real deployment. In this sense, a lab environment is an enabler, facilitating the design process while at the same time simplifying communication between all stakeholders involved and thereby enables true collaboration along the four before mentioned phases of an HRC project.

3.1 Analysis of HRC potential

During the potential analysis, seven different work systems (see example Figure 1) were analysed. All workstations included a manual process from circuit board cutting to container labelling, representing standard processes in SMD manufacturing.

Based on an MTM-analysis (Methods-Time-Measurement) the time share of single activities in the whole packaging process have been determined (Figure 2). With information about batch size, cycle time and remaining life cycle, the maximal investment costs for a collaborative robotic application could be calculated (Table 1).

The next step of the analysis was the determination of options to automate individual processes within the work system. Single sub-processes had to be checked for their ability to be automated. Criteria that aggravate the automation of a task can be found in [8].

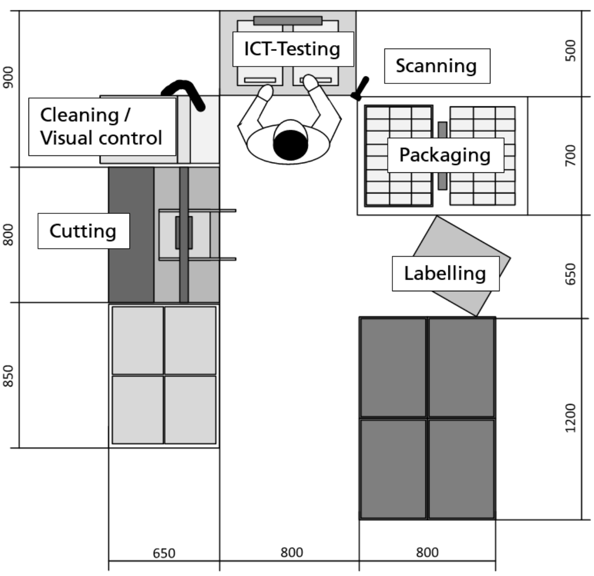


Figure 1 - Example of a manual work system.

No.	Batch size	C/T [s]	Remaining life span [yrs]	Imax [EUR]
1	700 k	28	3	160 k
2	160 k	29	3	80 k
3	10 k	26	3	4 k
4	80 k	26	2	37 k
5	500 k	18	3	120 k
6	160 k	20	3	42 k
7	2.7 M	6	4	290 k

Table 1 - Maximal investment costs for work systems 1-7.

In this example, the operation of the SMD cutting device was identified as unable to be automated using an articulated robot. Also, for visual quality control no suitable automation technology could be identified. Labelling is hard to automate due to the limpness of the label itself. All other sub-processes can be operated by both human and robot (Table 2).

For each sub-process that could be assigned to the robot, the cycle time had to be determined. It can either be estimated, generated from a virtual simulation or derived synthetically using MTM-MRK standard robot motion times. Based on this information, the final task allocation pattern can be created under consideration of cycle times in order to achieve an equal load balancing between human and robot and thereby a high resource utilization [9].

Now, the maximum economic investment cost for this task allocation pattern I_{max}, which is equivalent to the operating cost reduction induced by the robot, can be determined using the formula

Work system 7	Human	C/T h	Robot	C/T r
Cutting	X	9	-	-
Cleaning	X	2	X	6
Visual control	X	1	-	-
ICT testing	X	4	X	9
Scanning	X	3	X	3
Packaging	X	4	X	12
Labelling	X	1	-	-

Table 2 – Possible task assignments and respective cycle times (c/t) for human and robot.

$$I_{max} = (L_{sub} * Q_{ann} * S_{rem})/3600 * R_{hrly}$$

where L_{sub} is the manual labor substituted by the robot per product in seconds, Q_{ann} is the anticipated annual production quantity and S_{rem} is the remaining product lifespan. R_{hrly} represents the hourly rate of an operator.

Table 1 represents the exemplary determination of I_{max} for several analysed manual SMD manufacturing work systems. While in general, larger batch sizes and longer remaining product life spans facilitate amortization, it becomes evident that even batch sizes in the range of 100k-200k per annum may provide sufficient potential for process cost reduction to justify investments into human-robot-collaboration. Also, improved quality and reduced work-related burden may indicate implementation of HRC even if monetary break even might not be reached during the remaining product lifespan. For pilot implementation, system No. 7 was chosen (see. Fig. 2) since it offered the highest operation cost reduction and a relief of very monotonous tasks in packaging at the same time.

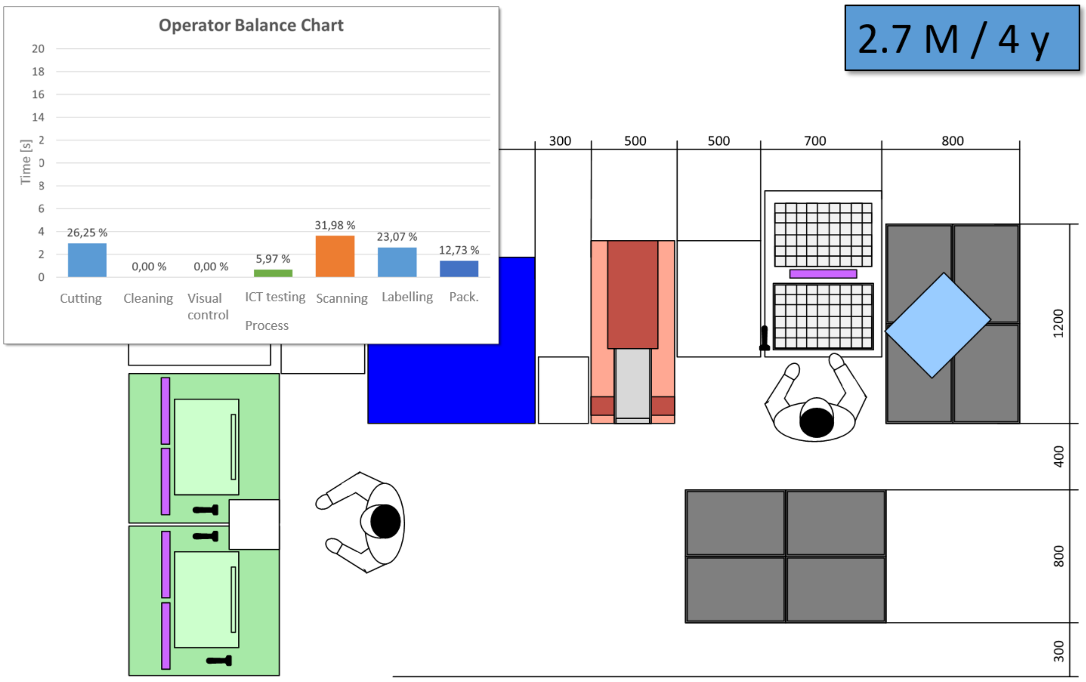


Figure 2 - Potential analysis indicating batch size, remaining life span and operator balance chart.

3.2 Technical concept

After making a decision on HRC implementation, the technical concept needs to be developed. Here, the HRC work system layout is an advisable starting point since it has to take into account the spatial conditions on the shop floor, motion economics of the operators and material flow directions. It should be developed collaboratively between all involved stakeholders such as engineers, operators and works council to achieve high degree of acceptance.

In this use case, the robot was responsible for scanning and packaging, which involves scanning SMD barcodes, picking SMDs up, placing them on a tray, stacking complete trays into a container, and closing the container. The layout, conveyed into a 3D model, facilitates reachability studies and delivers specification requirements that facilitate selecting a suitable manipulator.

General requirements for HRC work systems that reduce investment, implementation efforts and error sources include:

- Low-cost lightweight robot, which is easy to integrate and program.
- Use of gravity conveyor system without any drives, simple layouts.
- Small space requirements, nevertheless stable anchoring of all equipment.
- Ergonomic working height for operators.
- Short walking distances for operators.
- Avoidance of over-engineering that causes excessive implementation workload.
- ESD-protection of the whole system to avoid damage to the SMDs produced.
- Easy adaptability to additional products. This includes end effector design, fixings of trays and containers as well as software design.

Based on available lightweight robots on the market, a Universal Robot UR5 with a reach of 850 mm and a maximum payload of 5 kg was chosen for implementation of the system shown in Figure 3.

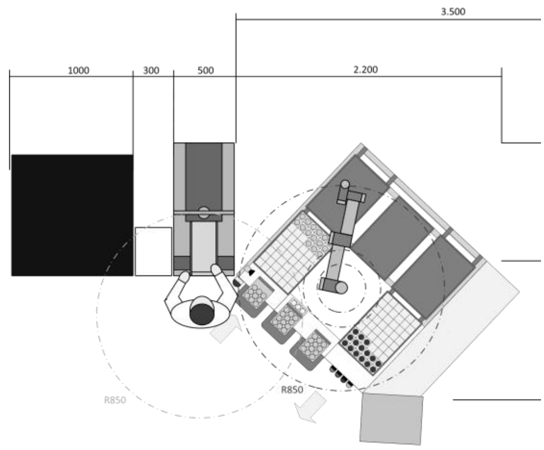


Figure 3 – HRC layout for work system 7.

As initial cycle time comparisons between humans and robots indicated (see Table 2), articulated light-weight robots for HRC move way slower than human operators. To compensate for this, it can be examined whether it would be possible for the robot to e.g. manipulate more than one object at the same time. Another major challenge that comes with SMD handling is their uneven surface with drillings and openings [10]. Hence, a specific end effector was designed (see Figure 4).

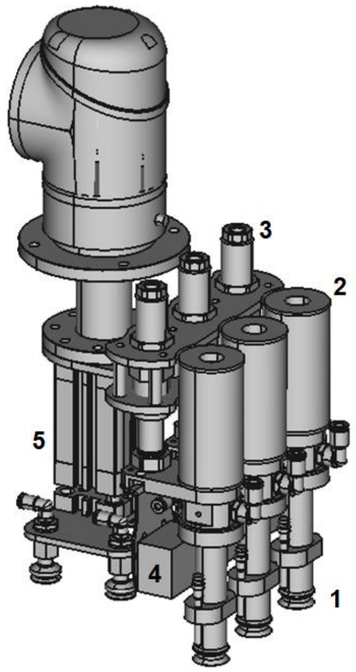


Figure 4 – End effector design in accordance with functional and safety requirements.

The collaborative high-performance end effector is using the Venturi principle and can grip up to three SMDs at the same time (1). Its high suction rate compensates for the leakages that result from the mounted devices and drillings (Figure 5). Silencers reduce the noise level of the employees (2). The entire front part of the gripper is spring-mounted to give way in the event of collisions with employees or the environment (3). Three integrated 2D-scanners identify SMDs through their imprinter barcode (4). A lift cylinder carries conventional suction pads which can be lowered for tray and container handling (5).

Equipment design principles that again ease implementation of HRC are:

- Same distance between products/objects at pick-up and drop position, e.g. within trays and blisters. Thus, a series of three could be grabbed and dropped, without extending or reducing the distance between all three.
- Stable positioning for products/objects, using centering pins or other positioning aids at all pick-up and drop locations allow for robust repeatability and avoid the need of complex vision and sensor systems.

Figure 5 – Uneven surface of circuit boards [12].

3.3 Safety concept

Concurrently to developing a technical concept for the HRC application, a safety and security concept that includes all kinds of safety features and devices should be developed. This is supposed to ensure that the HRC application complies with all relevant guidelines such as ISO 13855, ISO 13857 or ISO/TS 15066. If it does, it can become CE certified and considered safe and does not expose human operators to severe hazards. Most robots meant for HRC are equipped with safety-rated functions to protect operators from collisions and bruises, e.g. by limiting velocities or detecting collisions before they happen. Additional safety devices available on the market include maintenance gates, safety fences, safety covers, software based workspace limitations, extension of contact areas by design, safety mats, light curtains, laser scanners and emergency stop buttons. However, HRC applications should, as far as possible, avoid the use of such devices for the reason of both reducing system complexity and invest and rather rely on robot-inherent safety features such as collision detection, power and force limitation.

The safety of an HRC application can be formally assessed using the analytical-experimental approach described in ISO 12100, which is a standard approach for all kinds of industrial machinery. As a special feature in MRK applications, the acting forces and pressures in representative collision cases between human and robot must be determined by measurements on the real, physical test setup.

Measured forces may not exceed threshold values as defined per ISO/TS 15066 [1]. In the exemplary use case, following restrictions were defined:

- Robot’s speed must be limited to 250 mm/s.
- The robot’s working height must be limited to 2 cm above the working table surface.
- The end effector has to be spring-loaded to absorb quasi-static collisions between the tool and e.g. the work station surface.
- Sixth axis of the robot must be limited so it doesn’t move the gripper under its structure.

3.4 Validation and implementation

For the purpose of validation and optimization of the technical and safety concepts, a demonstrator of the application was realized on lab-level including all peripheral and sensory equipment.

The software framework was designed based on state diagrams to enable a transparent program structure.

In order to be able to determine the current status of the machine at all times and identify sources of error, such as lost parts or imprecise positioning of a container, 37 inductive, capacitive and reflective sensors were implemented, increasing the complexity of the system. To reflect current aspirations in terms of Industry 4.0, the robot was directly linked to the manufacturing execution system via a TCP/IP connection, continuously reporting scanned IDs of successfully packed SMDs for traceability purposes and allowing remote monitoring and maintenance.

The implementation was conducted directly on the EMS factory site (Figure 6). An essential part of the implementation was the involvement of employees working in collaboration or cooperation with the robot. Training of operators took place already during the validation phase in order to collect their feedback. Also, maintenance staff was trained to program the robot, make changes to the program and add new products (and thus parameters) to the application.

Figure 6 - Final design of the HRC workstation.

4 CONCLUSIONS

A step-by-step procedure for realizing human-robot collaborative applications from a broad analysis of potentials to industrial ramp-up has been developed and implemented. The study shows that integrative human robot work systems are indeed a viable approach to achieve efficiency improvements for processes which cannot justify full automation and the technologies available are sufficiently mature to meet industrial requirements. As for many things, the devil is in the detail: As diverse products and processes are, so will be the HRC applications and the solutions and innovations they carry. The presented approach is currently in use by the authors in other industries and will thereby be refined to achieve cross-sector applicability. The provision of a test lab for HRC at the TU Wien has proven itself as an incubator for new HRC applications that accelerates development processes through collaboration and hands-on implementation and testing capabilities.

5 ACKNOWLEDGEMENTS

This work has kindly been sponsored by the Austrian Research Promotion Agency (FFG). The design and analysis approach was developed in the project "MMAssist II – Assistance Systems for Human-Machine Cooperation in Production" (FFG No.: 858623). The development of the technical concept and its implementation were part of a project within the Austrian Center for Digital Production (ACDP), which is part of the Competence Centers for Excellent Technologies (COMET) programme.

6 REFERENCES

[1] ISO/TS 15066 – Collaborative robots & robotic devices.

[2] Michalos, G.; Makris, S.; Spiliotopoulos, J.; Misios, I.; Tsarouchi, P.; Chryssolouris G.: ROBO-PARTNER: Seamless Human-Robot Cooperation for Intelligent, Flexible and Safe Operations in the Assembly Factories of the Future. Procedia CIRP 23 (2014), pp. 71-76.

[3] Wang, X. V.; Kemény, Z.; Váncza, J.; Wang, L.: Human–robot collaborative assembly in cyber-physical production: Classification framework and implementation. CIRP Annals. Volume 66, Issue 1 (2017), pp. 5-8.

[4] Ranz, F.; Hummel, V.; Sihn, W.: Capability-based task allocation in human-robot collaboration. Procedia Manufacturing 9/2017, pp. 182-189.

[5] Bauer, W. (2017): A Smarter World – Leben und Arbeiten in der Zukunft. Presentation. Fürstenfeldbruck, May 23rd 2017.

[6] Zwicker, C.; Reinhart, G.: Human-Robot-Collaboration System for a Universal Packaging Cell for Heavy Electronic Consumer Goods, In: Zaeh, M., Enabling Manufacturing Competitiveness and Economic Sustainability, Springer, 2014, pp. 195-199.

[7] Bouchard, S. [no date], Leanrobotics - A Guide to Making Robots Work in Your Factory, leanrobotics.org

[8] Beumelburg, K.: Fähigkeitsorientierte Montageablaufplanung in der direkten Mensch-Roboter-Kooperation. Approved dissertation, Institut für

Industrielle Fertigung und Fabrikbetrieb (IFF), University of Stuttgart, 2005.

[9] Glogowksi, P.; Lemmerz, K.; Schulte, L.; Barthelmey, A.; Hypki, A.; Kuhlenkötter, B.; Deuse, J.: Task-based Simulation Tool for Human-Robot Collaboration within Assembly Systems, In: Schüppstuhl, T. et al., Tagungsband des 2 Kongresses Montage Handhabung Industrieroboter, Springer, 2017, pp. 155-163,.

[10] Rooker, M.; Wögerer, C.; Angerer, A. et al.: Flexible grasping of electronic consumer goods. Int J Adv Manuf Technol 85 (2016), pp. 71-84.

[11] Ranz, F.; Komenda, T.; Reisinger, G.; Hold, P.; Hummel, V.; Sihn, W.: A Morphology of Human Robot Collaboration Systems for Industrial Assembly. Procedia CIRP 72 (2018), pp. 99-104.

[12] Schmalz SCG product sheet.

7 BIOGRAPHY

Titanilla Komenda holds a MSc from the University of Applied Sciences Technikum Wien and is a researcher in the Industrial Engineering specialist group at Fraunhofer Austria. Her major field of research is human-robot collaboration, where she is also working on her PhD at TU Wien.

Fabian Ranz holds a MSc from Reutlingen University and is a researcher in the Industrial Engineering specialist group at Fraunhofer Austria. His major field of research is human-robot collaboration, where is also working on his PhD at TU Wien.

Wilfried Sihn is Full Professor at the Institute of Management Science at TU Wien, head of Fraunhofer Austria and has been elected a fellow of the International Academy for Production Engineering (CIRP). His area of research is Operations Management.

Elucidation of Influencing Parameters of the Laser Butt Welding Process of Dissimilar Steel to Tungsten Alloy Sheets

M. Schiry[1], P. Plapper[1]

[1] Laser Technology Competence Centre, Research Unit in Engineering Science, University of Luxembourg

Abstract

Laser welded joints of steel and tungsten alloy require deeper understanding of the impact of the laser parameters on the welding process. Due to the dissimilar thermal and metallurgical properties of both materials, a defined energy input and a precise temperature control in the weld interface is required. In the presented work, the influence of different laser parameter (e.g. laser power, focal position, scanner speed etc.) on the joint quality of laser welded steel-tungsten alloy butt-joints were investigated. By using statistical methods, the influencing process parameters were defined and optimized. It is shown that the horizontal focal position is the main parameter for the welding process in a butt-joint configuration. Additionally, manipulation of the heat input leads to modification of the weld bead profile. By obtaining vertical seam boundaries in the sheet metal, a homogenous temperature profile at the interface of the materials was achieved. With reduced amounts of intermetallic or intermediate phases and complete bonding, high joint strengths were achieved.

Keywords

Laser welding, full penetration welding, tungsten hard alloy, steel, design of experiments

1 INTRODUCTION

Resource efficiency receives growing attention in numerous industry segments. Modern manufacturing companies are facing problems regarding material saving, structural strength and reliability of their products. Materials with high strength properties are needed to solve the dichotomy of lightweight construction and high structural performance. Tungsten hard alloys offer favourable mechanical properties like hardness, hot-hardness and wear resistance, which can ensure the requirements of the customer. Therefore, tungsten alloys are commonly used for dies, tools and moulds.

These mechanical advantages are accompanied with disadvantages like high brittleness and a limited machinability of the tungsten alloy part. Because of the thermally sensitive material properties, these tungsten hard alloys are commonly fabricated by sintering processes. This complex manufacturing process leads to higher production costs compared to other materials. In addition, the sintering process causes restrictions in the part design. Thus, dies, tools and moulds, which are completely produced out of tungsten alloy are unfavourable. The resulting economical and technological aspects limit the diversity in the potential product portfolio.

These circumstances lead to a huge demand of joining technologies of tungsten alloy parts to cheaper and more ductile materials (e.g. steel). Wear protection of moulds and dies are usually realized by laser cladding of tungsten alloy on a steel substrate [1]–[6]. By doing so, the tungsten alloy wire or powder is liquefied completely. Due to their thermal sensitive material properties, only specific alloys can be used. Unfortunately, those alloys have limited hardnesses. Therefore, joining solid wear protection parts out of tungsten alloy to a ductile substrate may solve the controversy of hardnesses and joinability. Therefore, different methods have been developed to join tungsten alloy parts to steel.

Brazing offers a solution for joining tungsten alloy parts to dissimilar materials. Compared to mechanical joining, brazing is the cheapest method for joining those materials. Torch brazing, induction brazing and furnace brazing are employed for providing the parts with the required amount of heat in the interface. Due to the specific liquidation temperature of the filler material in the interface, the joint geometry succumb defined restrictions. In addition, the temperature of the final product during the applications must not reach the melting temperature of the filler material [7], [8].Furthermore, the thermal strength decreases significantly with $t_{max}=t_{sold}$ x 0.4. In addition, some filler materials outgas toxic and harmful fumes, which requires expensive fume exhaust and treatment [9].

To compensate the disadvantages of brazing, different welding techniques were developed in the last years. [10] and [11] developed the process of friction stir welding of tungsten alloy parts to steel samples. Diffusion welding was employed as an alternative technique to joint tungsten alloy parts. [12] used interlayer and graphite coatings on the tungsten alloy samples to suppress the formation of brittle intermetallic and intermediate phases. Laser radiation was also used as heat source for welding both materials in few works. Most frequently, the laser parameter and their influence on the process and the joint quality are in the focus of the researchers. [13] analysed the influence of the heat input on the bending strength of tungsten alloys with 15 % and 8 % Co. It was shown that the amount of Co as binder metal has a significant influence on the joint strength and the weld seam metallurgy. By manipulating the welding speed, an improvement of the joint strength was possible. Although a butt-welding process was given, the horizontal focal position kept constant and was not considered as in important parameter in the experiments.

The thermal stress after welding may lead to precipitate failures and fractures in the tungsten alloy part. Therefore, [14] employed a three step process with pre- and post-heating of the weld seam to reduce the thermal stress in the joint after welding. In a second investigation, the author applied nitrating of the steel specimen to avoid the formation of brittle and harmful phases in the weld seam [15]. Tungsten alloy with 20 % Co and steel were used during the experiments of [16]. For all welds, the laser beam had a fixed focus position on the tungsten alloy surface. By changing power, velocity and vertical focal position of the laser beam, a robust process was accomplished. Copper and invar inserts were used to join tungsten alloy grades with 6 % Co to carbon steel (C45) by [17]. Due to the ductile properties and the prevention of the formation of brittle phases, the joint strength was improved. Invar interlayer were also used by [18] for laser welding carbon steel to tungsten alloy with 20 % Co. The influence of different laser sources (CO_2 and Nd:YAG) on the joint quality of steel to tungsten alloy were investigated by [19]–[22]. Besides the laser sources, the laser parameter and the composition of the tungsten alloy parts were changed and their influence on the process analysed. The author showed that the different parameters influence the formation of phases and the joint strength.

The mentioned works define power and velocity as the most important parameters for the laser welding process of steel to tungsten alloy samples. Some authors focus the laser beam on the tungsten alloy side, while others focus on the steel and keep the position constant. It is known that the intermixture of the two partners in the melt bath defines the metallurgy and finally the joint quality. Therefore, the geometry of weld seam has a critical influence on the intermixture and the final joint strength [23]. Thus, the present paper will show the influence of the laser parameters power, velocity, horizontal focal position, laser source and weld seam shape on the joint properties of laser welded dissimilar steel-tungsten alloy joints. The paper is structured as follows. In chapter 2 the experimental set up is outlined. Chapter 3 shows the results and the discussion. The paper finally ends with a conclusion and a scientific outlook.

2 EXPERIMENTAL SET UP

The welding experiments have been carried out with two different laser sources. A 400 W fiber laser from company TRUMPF was employed. The laser radiation of the fiber laser was generated by a Nd:Glas laser medium and had a wavelength of 1070±10 nm. The laser beam was coupled in a 11 µm fiber and guided to a 90 mm collimator, which lead to a focal length of 254 mm and a focal spot diameter of 31 µm. To ensure a sufficient bead width, the laser beam was oscillated in circular trajectory.

The so called “wobble” motion can be described by the following equation:

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} -a \cdot \cos(2\pi f t) + v \cdot t + a_0 \\ -a \cdot \sin(2\pi f t) \end{pmatrix} \tag{1}$$

with

a	amplitude of the spiral [mm]
a_0	starting position at t=0 s [mm]
f	repetition frequency [Hz]
v	beam velocity on the trajectory [mm/s]

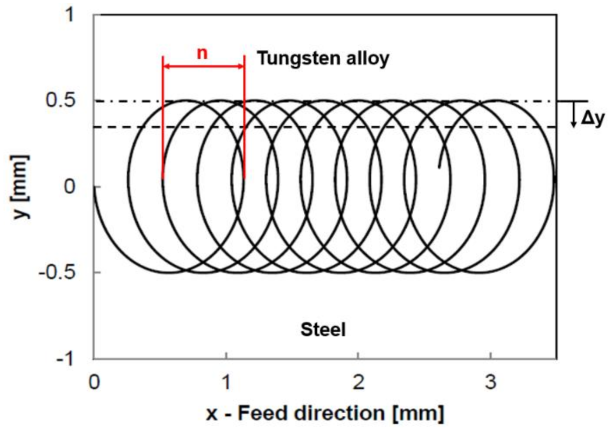


Figure 1 - Circular beam oscillation [18]

The wobble frequency influences the bead shape due to the modified fluid flow behaviour in the melt bath [24], [25]. Therefore, the frequency was kept constant *f*=64 Hz for all welds on the fiber laser. The overlap *n* of the particular spirals is defined by:

$$n = \frac{x \cdot \left(\frac{1}{2f}\right) - x \cdot \left(\frac{2}{f}\right)}{x \cdot \left(\frac{1}{2f}\right) - x \cdot \left(\frac{1}{f}\right)} \tag{2}$$

Thus, the resulting welding speed v_f can be calculated by:

$$v_f = 2 \cdot (a - n) \cdot f \tag{3}$$

The amplitude was set to 0.5 mm and the overlap *n* to 0.091 mm. The focal spot was positioned in the middle of the steel sheet (z-direction) and kept constant throughout the experiments. The remaining laser parameter were varied as described in table 1.

Beside the fiber laser, a 2kW Yd:YAG disk laser also from company TRUMPF was applied for performing the welds. This laser source emits laser radiation with a wavelength of 1030±10 nm. The laser beam was coupled into a dual-core fiber with a first (outer) diameter of 200 µm and a second (inner) diameter of 50 µm.

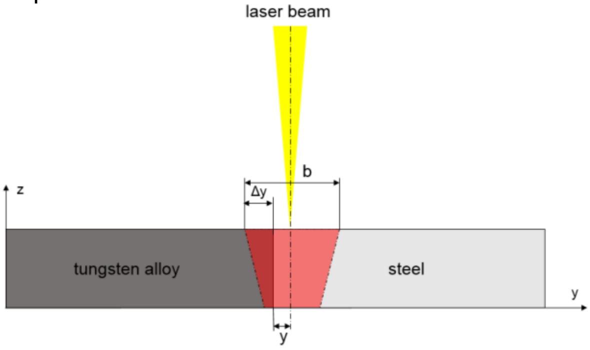


Figure 2 - Welding configuration

In combination with the 90 mm collimator, a focal distance of 160 mm was given. Depending on the used fiber diameter, the laser spot had a diameter of 89 µm for the 50 µm and 356 µm for the 200 µm fiber. Similar to welds with the fiber laser, the laser spot was positioned in the middle of the steel sheet and kept its position during all experiments (z-direction). The welds were performed in a butt weld configuration. Therefore, a full penetration process was necessary to join both parts completely. The laser beam was lateral positioned on the steel sheet (y-direction). Compared to the fiber laser, an expanded melt bath was generated by the larger spot diameters of the disk laser. Therefore, the laser beam was guided in a linear mode. By adjusting the horizontal focal distance of the laser beam to the interface *y* and in addition with the melt bath width *b*, the amount of liquid steel, which interacts with the tungsten alloy part may be modified (figure 2). The resulting width is called overlap *Δy*. The overlap *Δy* is the newly introduced parameter for the process and will be used in the following discussions.

For analysing the influences of the laser parameters on the disk laser, a central composite design of experiments (CCD) was configured (figure 3). The linear parameter behaviours of the 3 parameters power *P* [W], velocity *v* [mm/s] and overlap *Δy* [mm] are represented by the cube points (red). For representing non-linear characteristics, a

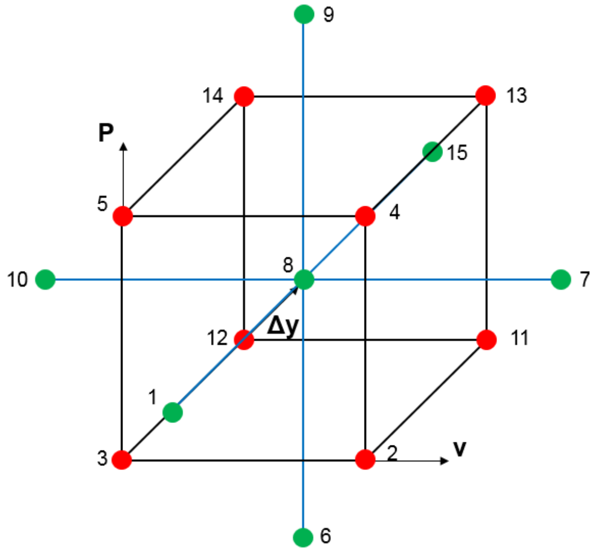


Figure 3 – Design of Experiments on the disk laser

	Disk laser	
Ø spot	**89 µm**	**356 µm**
P [W]	790/ 863/ 1200/ 1537/ 1600	1350/ 1376/ 1500/ 1623/ 1650
v_f [mm/s] (linear)	33/ 36/ 48/ 60/ 63	48/ 50/ 60/ 70/ 72
Δy [mm]	0.078/ 0.1/ 0.2/ 0.3/ 0.322	0.078/ 0.1/ 0.2/ 0.3/ 0.322
	Fiber laser	
P [W]	400	
v_f [mm/s] (wobble)	58	
Δy [mm]	0.1 / 0.2 / 0.3	

Table 1 - Laser parameters

star was set in the middle of the cube, which had a extend factor of $\alpha=\sqrt{2}$ (green).

For all welding experiments, a carbon steel with 0.4 % C, 0.3 % Si, 0.9 % Mn, 1.1 % Cr and 0.15% V (grade SAE 6135) was employed. The steel sheets had the dimensions of 45x35x1.2 mm. The tungsten alloy part consisted out of 82.5 % W, 5.5 % C and 12 % Co. The dimensions of the tungsten alloy part were 1.85x35x1.12 mm. Argon shielding gas with a constant flow rate of 21 l/min was applied to avoid oxidation.

Metallurgical analysis were done by cross section observation. Therefore, the samples were cut via electro discharge machining. Afterwards, the specimen were embedded into thermo-plastic resign, ground and polished with diamond suspension. After etching with Nital- (3 % HNO_3 in C_2H_5OH) and Murakami- (KOH + $K_3[Fe(CN_6)]$ in H_2O) etchant, the microstructure could be analysed by an optical microscope. For a detailed observation of the microstructure, SEM imaging was performed. In addition, the composition of the different intermetallic and intermediate phases were identified with EDS mapping. The mechanical properties of the joint were analysed by shear testing.

3 RESULTS AND DISCUSSION

3.1 Welding with fiber laser

Due to the small laser spot diameter of the fiber laser, the generated seam width lead to problems in the bonding behaviour. Therefore, a circular beam motion was applied (figure 1) to solve this problem. After metallographic weld seam examination, massive degradations in the tungsten alloy part were obtained. Figure 4 shows keyhole artefacts of the laser beam in the tungsten alloy specimen. Due to the circular oscillation, the laser beam alternates along the complete seam width. By adjusting small values of overlap, the laser beam traverses the tungsten alloy part. The high power density of the laser beam accompanied with the low

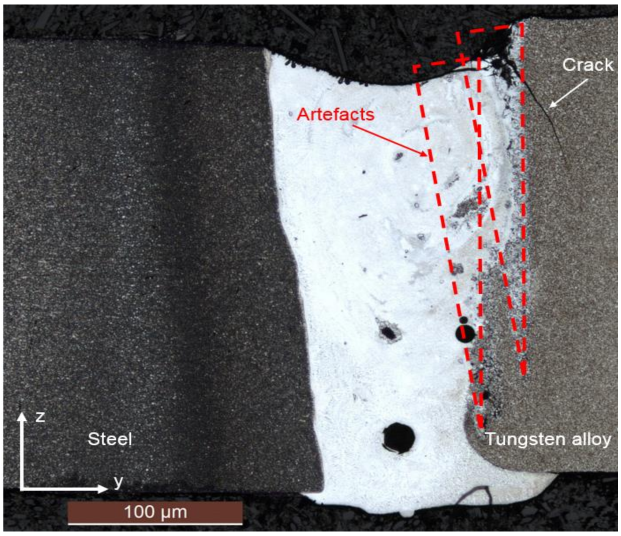


Figure 4- Welding defects with "wobble" motion

melting and vaporisation temperature of cobalt in the alloy results pores and complex intermetallic or intermediate phases. Especially the complex phases generate high thermal stresses and brittle material properties after the welding process. The massive crack in figure 4 documents this issue. In addition, the high power density of the small laser spot results V- and X-shaped weld beads [23]. Those bead geometries lead to local concentrated bonding near to the upper and/or lower surface of the seam, while in the middle of the joint no bonding is achieved. Therefore, complete bonding requires overlap values *Δy* > 0.1 mm, which entails the crossing of the laser beam over the heat sensitive tungsten alloy sample.

In consequence, it was not possible to overcome the discrepancy of sufficient bonding and degeneration free tungsten alloy parts with the "wobble" beam motion. Thus, welding dissimilar heat sensitive materials in a butt-joint configuration by means of circular laser beam oscillation is unfavourable.

3.2 Welding with disk laser

The disk laser source in combination with the employed optical set up generates a lager spot diameter and in consequence a larger melt bath and heat affect zone. Therefore, the previous wobble motion was not necessary and a linear beam guidance was applied. Due to the high power of the disk laser, the induced keyhole was surrounded by a thicker film of liquid metal compared to the smaller spot of the 400 W fiber laser, which enables to join both parts without crossing of the laser beam over the tungsten alloy sheet.

For both optical diameters, statistical design of experiments were used. As an output of the design in figure 3, 15 different parameter sets were defined. These sets can be separated in 5 groups by the overlap and the applied energy. Group I with the parameter set 1 has the minimum overlap of *Δy*=0.078 mm. For the following sections II-V, overlap values *Δy*= 0.1/ 0.2/

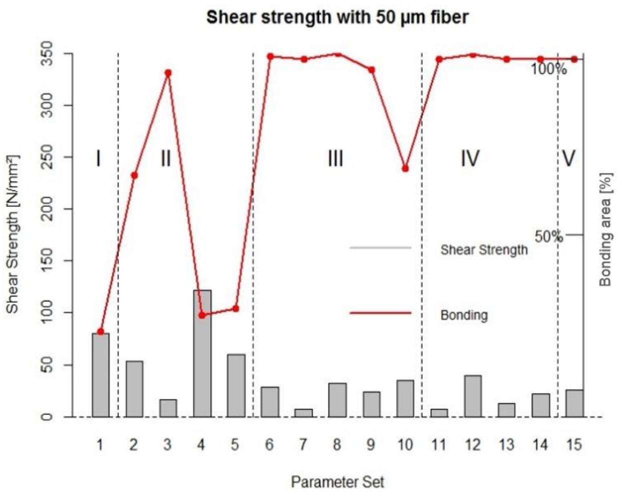


Figure 5 - Shear strength with 50 µm fiber

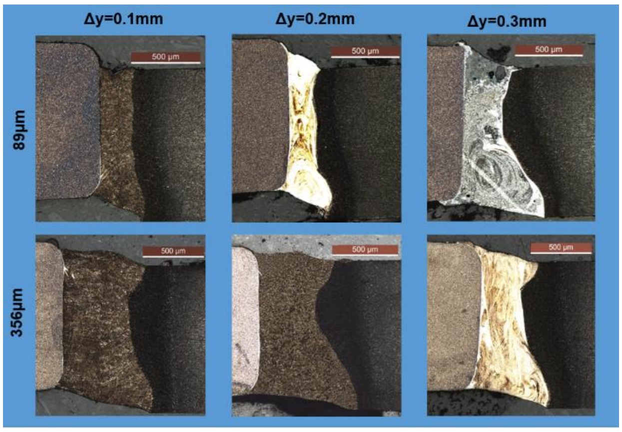


Figure 6 - Cross sections of welded samples

0.3 & 0.322 mm were adjusted respectively. The bars in the particular sections of figure 5 and 8 are listed with increasing energies. E.g. parameter set 2 in group II had an applied energy of 14 J/mm, while set 5 in the same group was welded with an energy of 42 J/mm.

Welds produced with the 89 µm laser spot of the 50 µm fiber result low shear strength. Minimum overlapping of 0.078 µm lead to medium shear strength. An increase of the overlap value up to *Δy*=0.1 mm improves the joint strength in case of higher energies. Welds with *Δy*=0.1 mm and low energies also show low shear strengths (figure 5). This could be explained by the dependency of the weld seam shape on the energy. High energies result in I-shaped seams, while welds produced with lower energies tend to V- or X-shaped seams. Local bonding, which means the area of joined tungsten alloy to the steel sheet, is the consequence. Thus, the percental bonding between both parts also increases when welding with higher *Δy* values. Almost complete bonding was achieved with overlap values of 0.2 mm and more. Although complete bonding was provided, the shear strength of groups III-V decreased noticeably (figure 5).

Metallographical observation of the weld specimen in figure 6 shows distinct

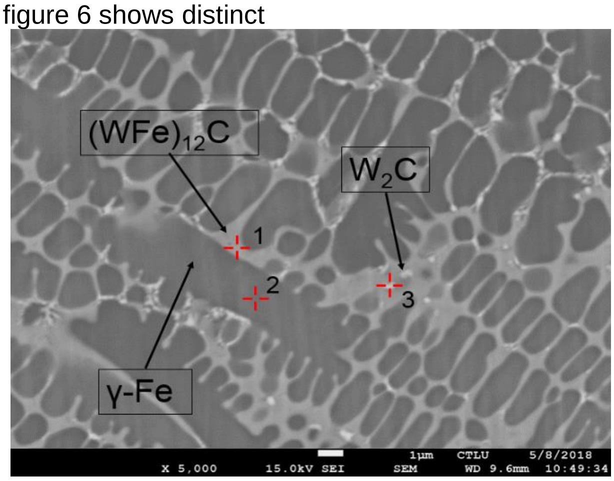


Figure 7 - SEM-Image of the brittle weld seam

EDS-Results			
Point	**W [%]**	**Fe [%]**	**C [%]**
1	40.5	54.7	In Balance
2	10.7	85.1	In Balance
3	91.7	6.4	In Balance

Table 2- EDS-Results

differences in the weld seam metallurgy. The low overlap value of *Δy*=0.1 results incomplete bonding between both specimen. The weld bead does not form a homogeneous geometry with vertical phase boundaries, therefore the wetting of the tungsten alloy part with molten steel only appears in the upper and/or lower area of the seam. This issue lead to the incomplete bonding of both parts. In addition, the weld seam has a martensitic microstructure. This microstructure is a consequence of the high carbon equivalent of 0.62 for the used steel grade. An increasing value of overlap is accompanied by an intensified melting of the tungsten alloy part and an alloying of the melt bath with tungsten and carbon. Because of the higher tungsten and carbon content in the weld seam, a modified seam metallurgy can be seen in figure 6. In the SEM image (figure 7), complex phases are detected. The ferritic-pearlitic microstructure of the raw steel part changed the crystallography to an austenite (γ-Fe) dominated structure.

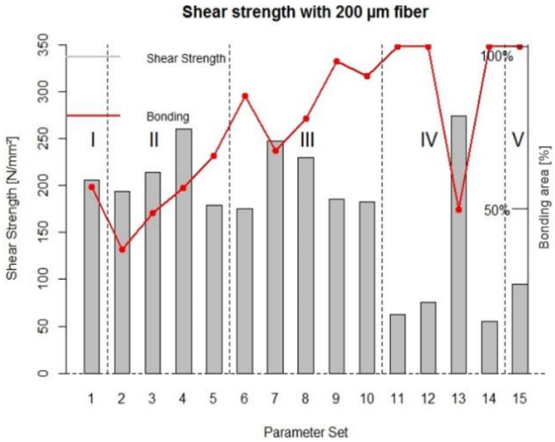


Figure 8 - Shear strength with 200 µm fiber

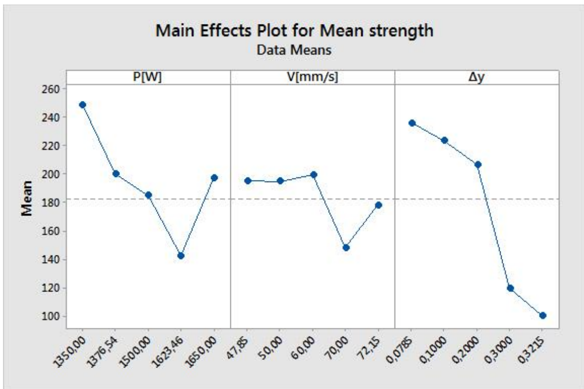


Figure 9 - Main Effect plot (200 µm fiber welds)

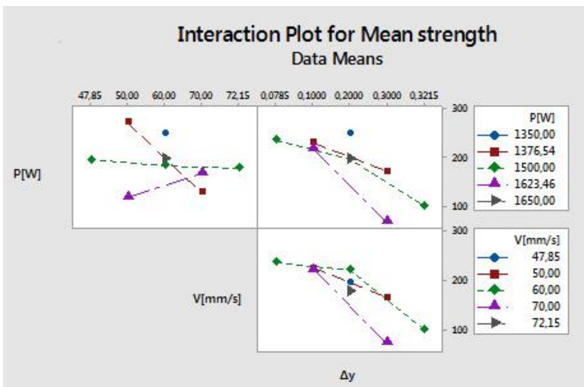


Figure 10- Interaction plot (200 µm fiber welds)

The austenite stabilization took place as a result of the increased carbon content in the melt. The austenite grains are surrounded by a dendritic structure on the grain boundaries, which consist out of the complex double carbide $(WFe)_{12}C$ (eta-phase).

First XRD measurements identified this as Fe_6W_6C. However, this needs to be confirmed in following investigations. Tungsten, which is not dissolved in the eta-phase forms sub stoichiometric W_2C. The performed EDS-analyses in table 2 confirm the alloying effect of the steel bath by the molten tungsten alloy.

The eta-phase and the W_2C are brittle, which lead to the significant reduction of the shear strength (figures 5 and 8). Welding with the larger laser spot diameter of 356 µm for the 200 µm fiber increases the shear strength. In comparison to the welds performed with the 89 µm laser spot, the shear strength of the welds of the groups I-III improved strongly. Due to the lower power density in the laser spot of the 200 µm fiber, the bead profile transformed. As shown in figure 6, the bead profile has a X-shape geometry. This shape leads to local bonding in the upper and lower surfaces of the seam. Therefore, a complete bonding could be obtained with overlap values more than Δy=0.2 mm.

Although, the joint strengths of welds produced with the 356 µm laser spot and lower overlaps were significant higher than samples welded with the smaller spot diameter of 89 µm. With Δy=0.1 and 0.2 mm, the weld seam showed a martensitic microstructure, which can be correlated with higher strengths. Indeed, with *Δy*=0.2 mm the same amount of tungsten alloy was melted during the process compared to the previous welds. Although, the relative concentration (%) of tungsten and carbon was decreased, caused by the larger weld seam volume generated with the 356 µm laser spot. Based on previous investigations, it is known that 6-8 % of the tungsten alloy part can the dissolved in the liquid steel bath without the phase transformation into austenite. Due to the higher seam volume, the critical concentration of tungsten and carbon in the melt bath could be shifted to higher overlap values.

Similar to the samples welded with the 50 µm fiber, overlapping *Δy*≥0.3 mm lead to an intensified melting of the tungsten alloy part and to a carbon and tungsten supersaturated melt bath, which results the harmful phase transformation of the weld seam metallurgy.

The statistical evaluation of the results can be seen in figure 9 and 10. As visualised in the main effect plot, power and velocity influences the joint strength differently. While the power has a strong impact on the joint strength, the influence of the welding speed is more moderate. Thus, the ratio of power and velocity (the energy per unit length [J/mm]) is not appropriate to substitute these two separate parameters, because the interaction of both parameters is not constant. In fact, the energy per unit length has to be mentioned accompanied with the applied value of power and velocity. The interaction plot in figure 10 affirms the strong interaction of power and velocity. Depending on the particular values, limited interaction between power / overlap and velocity / overlap appeared. In addition to the metallographical results, the overlap *Δy* has the strongest impact on the joint strength. For the 356 µm laser spot welds, overlap values of *Δy*=0.078-0.2 generate proper strength. An increase of *Δy* up to 0.3 mm and more lead to a brittle and phase transformed weld seam metallurgy. A significant reduction of the strength is the consequence. The same characteristic can be observed for welds produced with the laser spot of 89 µm.

4 CONCLUSION

A full penetration laser welding process of steel-tungsten alloy sheets has been investigated. The main outcomes can be summarized as follows:

- Using small spot diameter with circular beam oscillation lead to degradation of the tungsten alloy.
- High shear strength can be correlated with martensitic microstructures.
- Increased overlap values lead to intensified melting of the tungsten alloy part and to a phase transformation in the weld seam.
- Due to the higher carbon content, a austenitic stabilisation takes place.
- Additionally, W_2C and $(WFe)_{12}C$ was identified, which lead to brittle material properties and low shear strengths.
- Using of a larger laser spot results a bigger seam volume, which decreases the relative concentration of tungsten and carbon in the melt bath and allows larger overlapping and finally higher shear strengths.
- The impact of power and velocity on the shear strength is unequal and both parameter show strong interaction.
- The overlap *Δy* is the most important parameter for the process.

Future investigation will focus on the improvement of the bonding behaviour of the liquid phase to the joint partner and detailed analysis of the intermetallic and intermediate phases in the seam.

5 REFERENCES

[1] M. Riabkina-Fishman, E. Rabkin, P. Levin, N. Frage, M. Dariel, A. Weisheit, R. Galun, and B. Mordike, “Laser produced functionally graded tungsten carbide coatings on M2 high-speed tool steel,” *Mater. Sci. Eng. A*, vol. 302, no. 1, pp. 106–114, Apr. 2001.

[2] D. Liu, L. Li, F. Li, and Y. Chen, “WCp/Fe metal matrix composites produced by laser melt injection,” *Surf. Coatings Technol.*, vol. 202, no. 9, pp. 1771–1777, Feb. 2008.

[3] A. Liu, M. Guo, M. Zhao, and C. Wang, “Microstructures and wear resistance of large WC particles reinforced surface metal matrix composites produced by plasma melt injection,” *Surf. Coatings Technol.*, vol. 201, no. 18, pp. 7978–7982, Jun. 2007.

[4] C. P. Paul, H. Alemohammad, E. Toyserkani, A. Khajepour, and S. Corbin, “Cladding of WC–12 Co on low carbon steel using a pulsed Nd:YAG laser,” *Mater. Sci. Eng. A*, vol. 464, no. 1–2, pp. 170–176, Aug. 2007.

[5] A. Techel, A. Luft, A. Müller, and S. Nowotny, “Production of hard metal-like wear protection coatings by CO2 laser cladding,” *Opt. Quantum Electron.*, vol. 27, no. 12, pp. 1313–1318, 1995.

[6] M. Piec, L. A. Dobrzański, K. Labisz, E. Jonda, and A. Klimpel, “Laser Alloying with WC Ceramic Powder in Hot Work Tool Steel Using a High Power Diode Laser (HPDL),” *Adv. Mater. Res.*, vol. 15–17, pp. 193–198, 2007.

[7] W.-B. Lee, B.-D. Kwon, and S.-B. Jung, “Effects of Cr3C2 on the microstructure and mechanical properties of the brazed joints between WC–Co and carbon steel,” *Int. J. Refract. Met. Hard Mater.*, vol. 24, no. 3, pp. 215–221, May 2006.

[8] W.-B. Lee, B.-D. Kwon, and S.-B. Jung, “Effect of bonding time on joint properties of vacuum brazed WC – Co hard metal/carbon steel using stacked Cu and Ni alloy as insert metal,” *Mater. Sci. Technol.*, vol. 20, no. 11, pp. 1474–1478, Nov. 2004.

[9] J. S. At, “The joining of tungsten carbide hardmetal to steel,” *J. South African Inst. Min. Metall.*, vol. 87, no. 5, pp. 125–135, 1987.

[10] A. N. Siddiquee and S. Pandey, “Experimental investigation on deformation and wear of WC tool during friction stir welding (FSW) of stainless steel,” *Int. J. Adv. Manuf. Technol.*, vol. 73, no. 1–4, pp. 479–486, Jul. 2014.

[11] M.-N. Avettand-Fènoël, T. Nagaoka, H. Fujii, and R. Taillard, “Characterization of WC/12Co cermet–steel dissimilar friction stir welds,” *J. Manuf. Process.*, vol. 31, pp. 139–155, Jan. 2018.

[12] A. M. Cottenden and E. A. Almond, “Hardmetal interlayered butt-joints made by diffusion bonding and pressure bonding,” June, 1981.

[13] Y. Baohe, “Study on Laser Wwlded Strength of Cemented Carbide,” *Chinese J. Lasers*, no. 5, pp. 2–6, 1998.

[14] C. Barbatti, J. Garcia, G. Liedl, and A. Pyzalla, “Determination of phases and residual stresses in laser beam welded joints of cemented carbide to steel,” *HASYLAB Annu. Rep.*, no. 211, pp. 523–524, 2004.

[15] C. Barbatti, J. Garcia, G. Liedl, and A. Pyzalla, “Joining of cemented carbides to steel by laser beam welding,” *Materwiss. Werksttech.*, vol. 38, no. 11, pp. 907–914, Nov. 2007.

[16] P. Xu, D. Zhou, and L. Li, “Fiber laser welding of WC-Co and carbon steel dissimilar materials,” *Weld. J.*, vol. 96, no. 1, pp. 1–10, 2017.

[17] Z. Mirski, K. Granat, and S. Stano, “Possibilities of laser-beam joining cemented carbides to steel,” *Weld. Int.*, vol. 30, no. 3, pp. 187–191, Mar. 2016.

[18] X. Yu, D. Zhou, D. Yao, F. Lu, and P. Xu, “Fiber laser welding of WC-Co to carbon steel using Fe-Ni Invar as interlayer,” *Int. J. Refract. Met. Hard Mater.*, vol. 56, pp. 76–86, Apr. 2016.

[19] R. M. Miranda, L. Quintino, A. Costa, J. C. P. Pina, T. Rosa, P. Catarino, and J. P. Rodrigues, “Analysis of Different Laser Welding Processes for Joining Hardmetals to Steel,” *Weld. World*, vol. 52, no. 7–8, pp. 42–51, Jul. 2008.

[20] A. Costa, R. M. Miranda, and L. Quintino, “Materials Behavior in Laser Welding of Hardmetals to Steel,” *Mater. Manuf. Process.*, vol. 21, no. 5, pp. 459–465, Aug. 2006.

[21] A. Costa; L: Quintino; R.M. Miranda, “Microstructural aspects of laser dissimilar welds of hard metals to steel,” *J. Laser Applictaions*, vol. 16, no. November, pp. 1–6, 2004.

[22] A. Costa, “Hard Metals Irradiation by a (CW) CO2-Laser beam,” *Int. J. Join. Mater.*, vol. 12, no. 4, pp. 118–123, 2000.

[23] M. Schiry and P. Plapper, “Influence of laser parameter on weld bead shape and process stability of laser welded steel sheets,” *10th Int. Conf. Photonic Technol. LANE 2018*, 2018.

[24] J. Gedicke, A. Olowinsky, J. Artal, and A. Gillner, “Influence of temporal and spatial laser power modulation on melt pool dynamics,” in *ICALEO congress proceedings*, 2007.

[25] O. Meier, “Hochfrequentes Strahlpendeln zur Erhöhung der Prozessstabilität beim Laserstrahlschweißen mit hoher

Schmelzbaddynamik," in *Zwischenbericht AiF, 13.*, 2005.

6 BIOGRAPHY

Marc Schiry holds a M.Eng. degree in General Mechanical Engineering from the University of Applied Sciences Trier. He is currently PhD student at the Laser-Technology-Competence-Center (LTCC) of the University of Luxembourg. His research is focused on laser welding of dissimilar materials.

Peter Plapper obtained his PhD degree in Machine tools and Production Engineering (WZL) from the Technical University of Aachen (RWTH). Since 2010 he is Professor for Tool Machines and Production Technologies at the University of Luxembourg.

Novel Opportunities by Laser Welding of Dissimilar Materials

P. Plapper[1], M. Schiry[1]
[1]Laser Technology Competence Centre, Research Unit in Engineering Science,
University of Luxembourg

Abstract
The Laser Technology Competence Centre (LTCC) of the University of Luxembourg provides skills in joining material combinations, which are considered being non-weld able with traditional methods or at least challenging to join (e.g. Copper and Aluminium, Aluminium and Polyamide (PA), Titanium and PEEK, steel and tungsten hard alloy). Related accomplishments include minimal intermetallic compounds, convincing mechanical and superior electrical properties of the laser-welded specimen. With defined spatial and temporal modulation of the laser beam, an accurately defined temperature profile is created on the lower side of the upper material in the two dimensional directions, which enable joining of dissimilar materials in overlap configurations, despite challenging thermal properties. To weld butt-joint geometries with minimal heat affected volume, the temperature profiles were expanded in the third dimension, which can be achieved through controlled laser energy guidance. The scientific methods to accomplish these convincing results are explained, with selected industrial use cases ranging from automotive industry, energy storage, and medical implants. An outlook with unsolved challenges is intended to ignite discussions about upcoming research topics.

Keywords
Laser welding, dissimilar materials, temperature control

1 INTRODUCTION

Lightweight design is one of the most interesting aspects for many industrial products. Common joining technologies, as clamping, screwing or gluing are accompanied with high part weights and costs. Therefore, substantive joining receives growing interests of engineers of different fields. Laser welding offers the best opportunities for welding dissimilar heat sensitive materials due to the precise heat input [1]. Thus, scientific investigations about the laser welding processes of dissimilar materials have been intensified during the last decades [2].

For the automotive industry, laser welding of polyamide to aluminium is a promising technique. Those joints of metal to polymer enables to reduce the part weight significantly. Thus, several publications deal with the challenging process of overlap laser welding of those materials [3]–[6]. By applying a defined energy density of the laser radiation on the surface of the aluminium sample, it was possible to liquefy the polyamide at the interface without degradation of the lower polymer specimen.

A similar overlap process is used for laser welding of copper to aluminium. Due to the poor dissolubility of copper in aluminium, brittle and electrical high resistant intermetallic phases occur frequently. An accurate heat input in the upper copper part guarantees minimum intermetallics, which is shown by different authors [7]–[10]. Those joints are commonly used for batteries in the automotive sector.

The brilliant energy application of laser radiation also enables to weld sintered materials to metals. Due to the specific material properties of sinter materials, welding, and in addition liquidation, of those metals is accompanied with material damages. However, the precise energy input of the laser allows to minimize the amount of liquefied sinter material in butt-joint welding processes. Thus, a reliable joint of sinter material to a dissimilar metal is accomplishable. This is proven by different studies and investigations [11]–[13].

The presented work points out exemplary the accomplishments of laser radiation for welding these three material combinations. Due to the heat sensitive material properties, the joints of aluminium-polyamide, aluminium-copper and steel-tungsten alloy represent some of the most challenging task for welding dissimilar materials. Thus, the developed processes and generated knowledge is presented in this work.

2 HEAT CONDUCTION WELDING OF ALUMINIUM TO PA 6.6

2.1 Experimental set up

Polyamide 6.6. melts at T_{melt}=255 C and starts to degenerate at a temperatures above 300 °C. In addition, aluminium has a melting temperature of 660 °C. It is obvious that a simultaneous presence of both melt baths is accompanied with degradation of the polymer specimen. Therefore, a heat conduction process is necessary to avoid thermal overload of the polymer and to join both materials successfully. For doing so, the laser beam generates a liquid melt pool on top of the aluminium sheet without full penetration of the metal part (figure1).

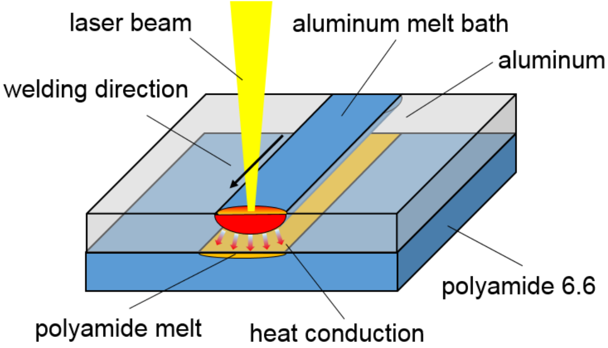


Figure 1 - Heat conduction welding of PA 6.6 to aluminium

Due to the excellent thermal conductivity, the heat of the melt bath disperses in the solid aluminium base material. Depending on the thermal conductivity and capacity, a defined spatial temperature gradient inside the aluminium part is generated. This temperature profile results the required melting temperature at the interface of the two parts. In consequence, a brilliant laser source with a high beam quality is needed to ensure the accurate heat input. Therefore, a fibre laser from company TRUMPF was employed to perform the welding experiments. The used fibre laser had a near-infrared wavelength of 1070 ±10 nm and a maximum peak power of P_{peak} = 400 W. The emitted laser beam was shaped by a SCANLAB HS20 f-θ-scanner with a collimator width of 90 mm. With the 11 µm optical fiber, a focal spot diameter of 31µm was generated. The aluminium surface is covered by an oxide layer. This oxide layer has a significant higher melting temperature as the base aluminium (Al_2O_3: t_{melt}= ca. 2050 °C, Al: t_{melt}: ca. 660 °C). Therefore, a minimum power density on surface of the aluminium sample is necessary to penetrate the oxide layer. An increase of the peak power enables to penetrate the layer, but raises the probability of full penetration of the aluminium sample and finally the risk of polymer degradation. By shaping the pulse time of the laser radiation, it is possible to reduce the average power and to avoid full penetration, but to guarantee a penetration of oxide layer, caused by the high peak power. Therefore, a pulsed laser process was employed, which offers a balance between both conflictive aspects.

The pulse shape is characterized by the pulse time T_{mod} and the pulse period T (figure 2). Modification of the pulse time and the pulse period leads to changes in the average power P_m, which can be calculated as follows:

$$P_m = T_{mod} \cdot {}^{1}/_{T} \cdot P_{peak} \qquad (1)$$

The frequency of the pulsing ($f = {}^{1}/_{T}$) was set to 25 kHz and kept constant for all welds. The laser beam was guided in a circular trajectory, which is shown in figure 2.

For the experiments, 0.5 mm thick aluminium sheets of the grade EN-AW 1050A with a geometry of 150 x 40 mm was employed. The polyamide 6.6 sample

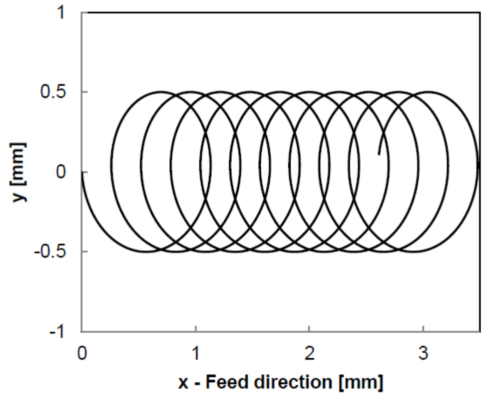


Figure 2 - Beam trajectory [14]

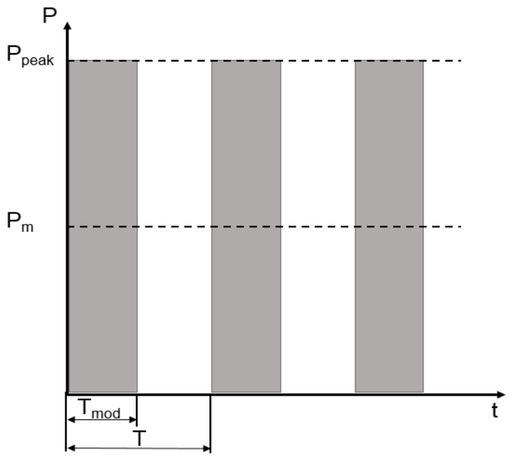


Figure 3 - Pulse modulation

had a thickness of 1 mm and a geometry of 60 x 40 mm. Due to the heat conduction process, the welds were performed in an overlap configuration (figure 1).

2.2 Results

Due to the precise energy input and the excellent beam caustic, a defined melt pool was generated in the aluminium sample. As shown in figure 2a, the resulting bead geometry shows an almost rectangular shape in the metallographical cross section. Depending on the dimension of the melt pool, which can be modulated by laser power, welding speed and pulse frequency, heat conducts out of the melt bath. Due to the high thermal conductivity of aluminium, the required melting temperature of PA6.6 was accomplished at the interface of the joint, which was measured by thermal imaging previously [6]. A thin fusion zone was detected at the interface by means of SEM imaging (figure 4b).

Because a full penetration of the aluminium sample by the laser beam is prevented, the lower surface of the aluminium specimen remains solid. The liquid polymer wets the aluminium and generates a physio-chemical bonding between both materials [5]. Modification of the laser parameters power, feed rate and pulse frequency lead to changes of the heat input

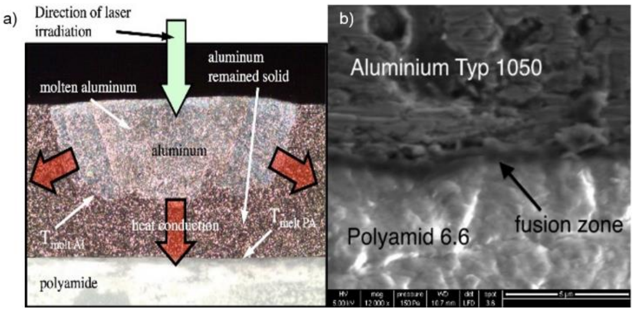


Figure 4 - Cross section (a) and SEM image (b) of fusion zone [5]

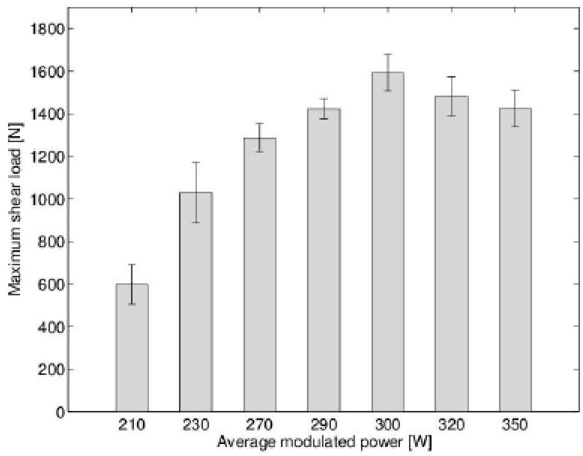


Figure 5 - Shear loads with different average powers [5]

and the dimension of the melt bath. Thus, a larger melt pool intensifies the heat conduction and increases the temperature at the very interface. Temperatures, which exceed the degradation temperature of the polymer, result damaging of the polymer and form gas bubbles, with may brake through the aluminium part and generate tunnels. In addition, the wettability of the liquid polymer to the solid aluminium is reduced by the gas bubbles. A significant decrease of the shear load of the joint is the consequence [6].

As shown in figure 5, small variations of the average power lead to magnificent changes in the shear load of the samples. Therefore, a precise heat input is necessary to ensure high quality. This defined heat provision may be accomplished by modifications of the pulse modulation or the peak power. Therefore, the developed pulsed laser process in addition with the brilliant beam quality of the fibre laser, enables to join polyamide 6.6. to aluminium with high quality.

3 OVERLAP WELDING OF COPPER TO ALUMINIUM

3.1 Experimental set up

Joints of aluminium and cooper are commonly used for batteries in the automotive industry. Those applications require high mechanical strength and low electrical resistance. Laser welding enables to join both materials with high strength and low electrical resistance, which is hard to achieve

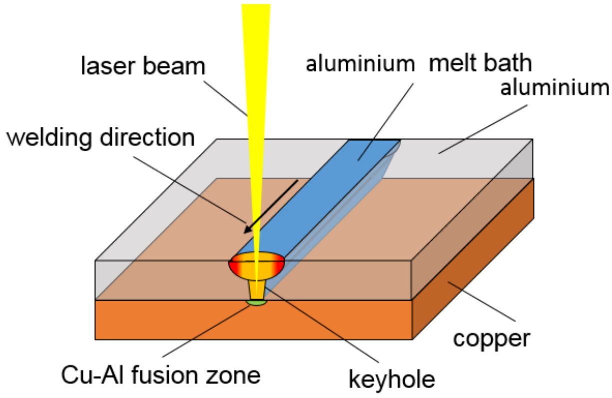


Figure 6 - Overlap welding of copper to aluminium

common techniques like ultrasonic welding.

Similar to welds of PA 6.6. and aluminium, a accurately defined energy input is also needed for welding copper to aluminium. Therefore, the same 400 W fibre laser was used for the Cu-Al welds. In addition, a pulsed laser process with pulse times between 24 µs and 48 µs was employed (figure 1). The laser beam was guided in the same circular beam movement as shown in figure 2 and with a constant feed rate of 500 mm/s. The aluminium and copper sheets were positioned in an overlap configuration (figure 6). Due to the high power density in the focal point on top of the aluminium samples, the laser beam penetrated the upper aluminium sheet completely. In addition, the lower copper part was melted just in a depth of a few µm. Thus, a defined intermixture of both metals in the fusion zone was the result. The mechanical properties of the joints were characterized by means of shear pull testing. The pulling force and the toughness were used to quantify the joint strength. By doing so, the toughness was calculated as a function of the pulling force and the elongation. For analysis of the weld seam metallurgy, metallographic cross sections were perpetrated and observed with an optical microscope. After embedding, grinding and polishing, the samples were etched with Keller reagent.

The electrical resistance of the weld seam was measured and correlated to the mechanical properties like maximum pulling force and toughness. Two oppositely positioned electrodes applied a constant electric current of I = 0.4 A. By measuring the decreasing voltage U, the electric resistance R was calculated regarding to the following equation:

$$R = \frac{U}{I} \tag{2}$$

For a statistical significance, each weld was measured 15 times.

3.2 Results

The joint quality of laser welded Cu-Al samples strongly depends on the induced energy during the laser welding process. As shown in figure 7a, an insufficient heat input leads to reduced local bonding of both materials. Following the wobble trajectory, the lower copper part melts partially at the interface. Most

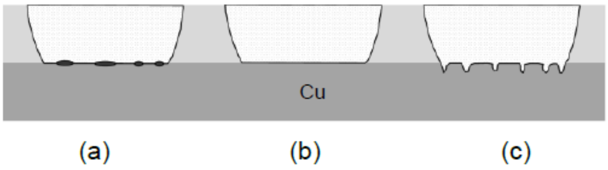


Figure 7 - Metallurgical process window [14]

of the Cu material stays solid at the interface. A small fusion zone accompanied with voids in the interface is the consequence. In contrast, an excessive energy input leads to a deeper penetration of the laser beam into the copper sheet and to an enlarged fusion zone (figure 7c). Due to the limited solvability of Cu in Al, intermetallic phases (IMC) are formed. A defined heat input enables to obtain complete bonding without the formation of a harmful amount of IMC (figure 7b) [14].

Corresponding to figure 8, the intermetallic phases lead to significant deterioration of the electrical and mechanical properties of the weld. Measurements of the electrical resistance showed the correlation between IMC and the electrical performance of the joint (figure 8). The intensity of the formation of IMC correlates with the raising pulse time. Thus, a longer pulse time lead to an intensified melting of copper and an increased formation of IMC. The highest electrical resistance was obtained by bear contact of both samples without a substantive joint. Low heat input accompanied with a metallurgical joint results low electric resistance, because of the low amount of IMC. A constant increase of the heat input results in an intensified formation of the amount of IMC and finally a higher electrical resistance. The optimum electric resistance was accomplished with a pulse time of 24 µs. This point corresponds with figure 7b, where a complete bonding with a low amount of IMC was generated [10].

In addition, higher amounts of IMC also influence the mechanical properties of the joint due to their brittle material properties. Thus, figure 9 shows the dependence of the maximum pulling force and the toughness on the pulse time and in addition on the amount of IMC. Low pulse times correlate with low laser energy input. A pulse time of 18 µs generates joints with inferior pulling strength and toughnesses. A moderate increase of the pulse time up to 21 µs improves the joint strength significantly. A following raise of the pulse time leads to an intensified melting of copper and a growth of IMC. Therefore, the maximum pulling forces decrease. Welds performed with a pulse time of 36 µs and more show brittle material behaviour. Therefore, a significant decrease of the strength and the toughness is the consequence [10].

Comparison of the results of the electric resistance and the pulling forces shows that an optimum heat input with low electrical resistance and high mechanical strength is accomplishable, due to a precise control of the energy input with the fibre laser. Ongoing research yields that detection of the emitted

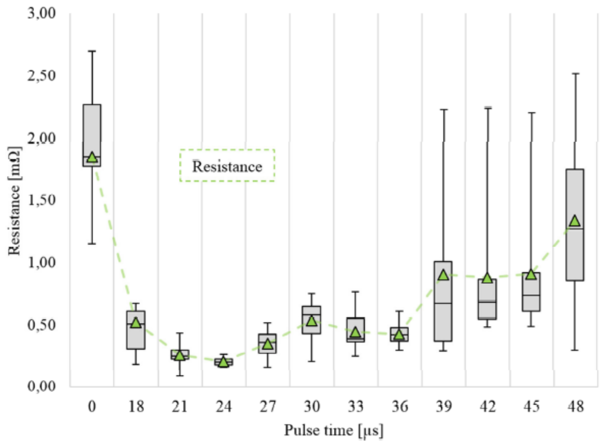


Figure 8 - Electrical resistance of the joints [10]

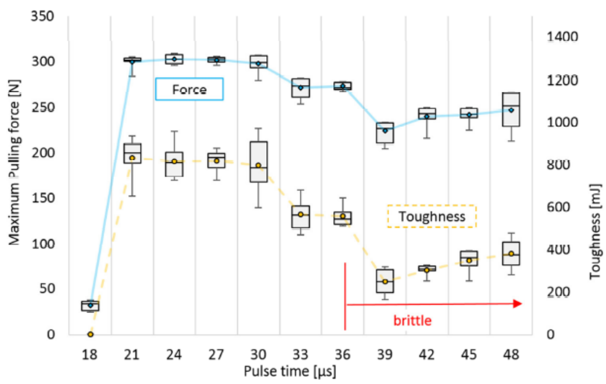


Figure 9 - Mechanical strength of the joints [10]

spectral process radiation is suited to define this process window of optimal joint performance [15].

4 BUTT-JOINT WELDING STEEL TO TUNGSTEN ALLOY

4.1 Experimental set up

In contrast to the other experiments, the tungsten alloy and steel sheets were welded in a butt-joint configuration (figure 10). Complete bonding between both parts was accomplished due to a full penetration process. Because of the heat sensitive material properties of the tungsten alloy part, the laser beam was laterally focused on the steel specimen. By changing the lateral distance of the laser beam to the interface of tungsten alloy and steel samples, the amount of melted tungsten alloy could be manipulated. The melt bath, which is generated by the laser beam, has a width *b*, which is dependent on the laser parameters power P and speed v. In addition, the high power density in the focal spot leads to a keyhole and enables to perform a full penetration weld. The keyhole capillarity is surrounded by a thin film of liquid metal. By adjusting the lateral distance of the laser beam to the interface, an interaction of the liquid steel with the tungsten alloy part is possible without damaging the heat sensitive sample caused

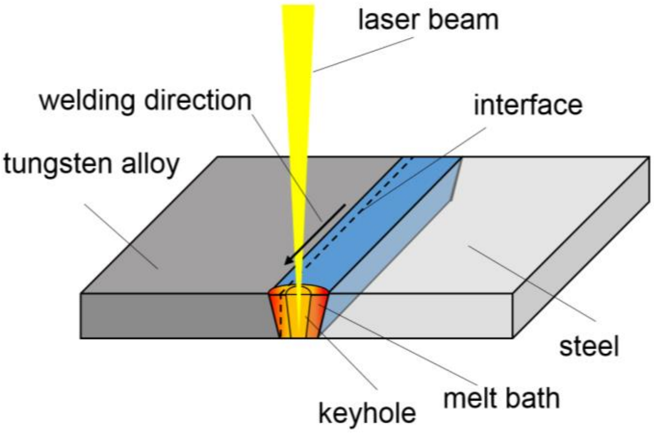


Figure 10 - Butt-joint welding of steel to tungsten alloy

by the laser radiation. The distance of liquid steel, which interacts with the tungsten alloy part, is called overlap Δy and is dependent on the weld seam width b and the lateral distance y [16]. In vertical direction, the laser beam was focused in the middle of the 1.12 mm thick steel sheet. The complete dimension of the steel sample of the grade SAE 6135 was 45x35x1.12 mm. The tungsten alloy specimen, which had a composition of 82.5 % W, 5.5 %C and 12%Co, had a dimension of 1.85x35x1.12 mm.

The welding experiments were carried out with a 2kW disk laser form company TRUMPF. The employed laser source had a wavelength in the near-infrared spectrum of 1030 ±10 nm. The used scanner optic with a collimator width of 90 mm was connect to the laser source by a dual-core fibre. This fibre had a first (inner) diameter of 50 µm and a second (outer) diameter of 200 µm. In combination with the collimator of the scanner head, focal spot diameters of 89 µm and 356 µm were realized.

The mechanical properties of the joint were analysed by shear testing. For the metallurgical investigations, metallographic cross sections were perpetrated. Therefore, the specimen were cut with diamond cutting blades, embedded in thermo-plastic resign, ground and polished with diamond suspension down to 1 µm. The microstructure of the weld seam was obtained by etching with Nital- (3 % HNO_3 in C_2H_5OH) and Murakami-etchant ($KOH+K_3[Fe(CN_6)]$ in H_2O) respectively. Metallographic observation was performed with an optical microscope.

Results

The applied laser spot diameters have a significant influence on the weld seam formation. Welds generated with a smaller laser spot lead to higher power densities on the surface of the specimen. Thus, an intensified heat input is the consequence, which results changes in the density of the liquid metal due to the temperature differences and in addition to changes of the surface tension of the liquid phase. Gradients in the surface tension initializes the Marangoni convection and influences the formation of the weld seam [17]–[19]. The same influence may be obtained due to variations of power

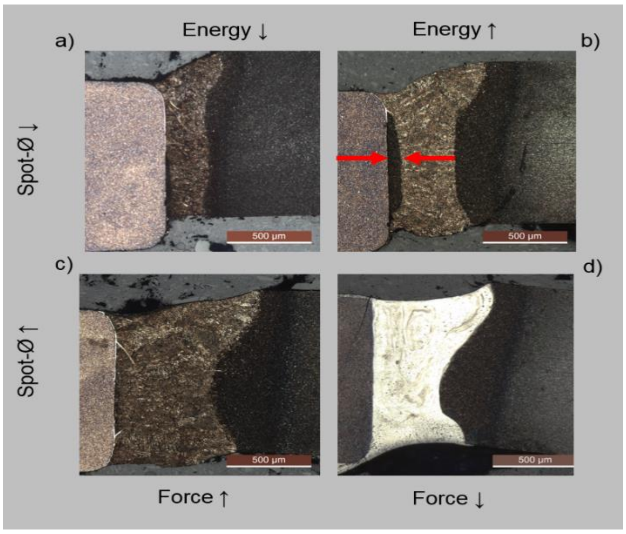


Figure 11 - Correlation between bead shape and strength

and welding speed. Figure 11 shows a distinct influence of the spot diameter on the bead shape and the bonding behaviour of the tungsten alloy sample to the steel specimen. All welds were generated with the same lateral distance *y* to the joint interface.

Low energies accompanied with a small focal diameter of 89 µm lead to an almost I-shaped weld seam geometry. Therefore, a sufficient bonding along the complete height of the weld seam between the tungsten alloy sample and the steel part was obtained. Due to the increased power intensity on the laser spot on the surface of the steel sample, the stimulated Marangoni convection formed a X-shaped bead profile. Thus, bonding between both metals was only accomplished at the upper and lower surface of the bead. An area without metallurgical bonding was observed in the middle of the weld seam (red arrow in figure 11b). The local concentrated joints in the interface decrease the maximum shear load, which is shown in figure 12. The larger focal spot diameter of 356 µm decreases the power density in the focal spot. This change of the heat input results a manipulated seam geometry due to the influenced 2-dimensional temperature profile on the surface of the metal. Complete bonding was accomplished with both energies. However, a significant change in the weld seam metallurgy took place.

A high maximum shear load may be correlated with a martensitic weld seam metallurgy, which is shown in [16]. This microstructure was obtained for the samples welded with a larger laser spot and low energy (figure 11c). Due to the higher energy density in figure 11d, the weld seam geometry and volume changed. In addition, the change in the bead geometry was accompanied with an intensified melting of the tungsten alloy part. This higher ratio of tungsten alloy melting, results a alloying effect of the steel melt bath. Based on previous results in [16] , it is known that a tungsten alloy content in the steel melt bath above 6-8 % causes the formation of complex and brittle intermetallic and intermediate phases. Those following phase transformations of

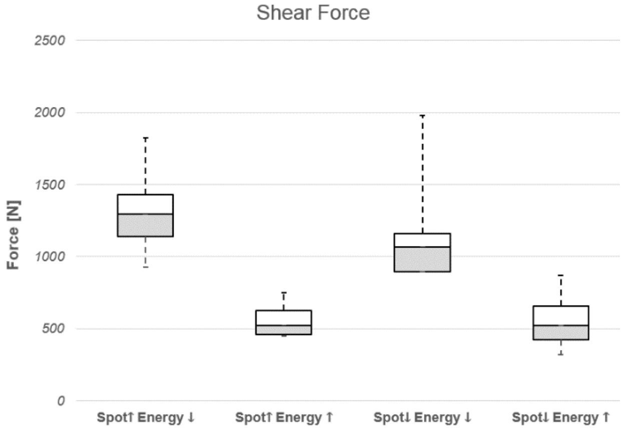


Figure 12 - Shear forces depending on the bead shape

the weld seam metallurgy can be seen due to the white colour of the weld seam in figure 11d. Those intermetallic and intermediate phases are brittle and cause high thermal stress in the weld seam and the heat affected zone. The change of the weld seam metallurgy caused a decrease of the maximum shear load, although a complete bonding between bot materials was given (figure 12).

5 DISCUSSION / CONCLUSION

Different welding methods were presented. Adjustments of the most important laser parameters power, velocity and focal position (in y- and z-direction) enabled to weld different heat sensitive materials in two welding configurations (overlap and butt-joint).

When welding aluminium to polyamide 6.6., an accurate heat input is necessary to melt the polymer without harmful degradation. Therefore, a defined weld pool was generated inside the aluminium sample. The dimension of the melt pool depends on the laser parameters power and velocity. In addition, the width of the melt bath may be manipulated by the width of the wobble path of the laser beam on the surface of the aluminium part. The volume correlates with the conducted heat outside the melt bath. Thus, changes in the laser parameters and in addition in the melt bath dimension, lead to an increase or decrease of the temperature at the interface. Therefore, power and velocity have to be selected carefully. An increased energy may lead to a penetration of the aluminium specimen and to a degradation of the polymer. However, low energy causes too low temperatures at the interface, wherefore the polymer remains solid during the welding procedure.

For the welding copper to aluminium, an overlap welding configuration was applied too. To obtain a joint at the interface of both metals, the upper aluminium layer was completely penetrated by the laser beam. In addition, the lower copper sample melts during the process due to the laser radiation as well. The two dissimilar melts stir and form intermetallic phases. Those phases are necessary for a substantive joint, however, they are accompanied by high brittleness and low electrical conductivity. Therefore, the amount of IMC has to be minimised for obtaining joint with high strength and low electric resistance. The amount of liquefied copper in the melt bath depends on the depth of laser penetration into the copper surface. A high penetration, leads to intensified melting of the copper sample, whereat a low penetration reduces the amount of liquid copper in the aluminium melt. The laser parameters power and velocity influence the penetration depth and finally the amount of IMC. A suited energy input enables a defined adjustment of the penetration depth and the amount of intermetallic phases. While in case of welding aluminium to polyamide the laser parameters power and velocity mainly defines if a joint is given or not, the same parameters allow to optimize the joint properties like mechanical strength and electric resistance for welds of Al-Cu.

The most complex influence of the parameters laser power and velocity on the welding process were obtained for joining steel to tungsten alloy in a butt-joint configuration. Welding in an overlap configuration requires a precise control of the temperatures on the 2-dimensional specimen surface at the interface of both materials. The interaction of the laser beam with the keyhole walls and the fluid flow behaviour around the capillarity play a secondary rule. In contrast, the spatial weld bead formation is a key factor in butt-joint welding processes. Due to specific bulking temperatures, the phase boundary of the liquid to the solid phase and the boundary of the heat affected zone can be correlated with specific temperatures. E.g. at the phase boundary liquid / solid, the melting temperature of the metal is present and follows the geometry of the boundary. Therefore, a variation of the melt bath geometry lead to changes in the temperature profile along the complete height of the weld seam. As shown in the results of the experiments, the weld seam geometry depends on different parameters like focal spot diameter, power, velocity. In addition, the type of shielding gas also influences the weld seam formation significantly [20]. However, the most critical influence was obtained due to variations of the power and welding speed. By

adjusting defined values of power and speed, depending on the thickness of the sheet, the material, the employed laser etc., it was possible to accomplish weld seams with vertical seam boundaries. Those vertical boundaries represent a vertical and homogenous temperature distribution at the interface. Enhanced joint strengths are the result.
Joining of multiple material combination requires comprehensive understanding of the temperatures at the interfaces, which is the key for joining those dissimilar materials thermally. With tailored parameter sets for the different material combination, it is possible to weld heat sensitive and dissimilar materials in an overlap and a butt-joint configuration, due to the defined control of the multidimensional temperature profiles.

6 REFERENCES

[1] J. F. Ready, *LIA Handbook of Laser Materials Processing*, 1st ed. Orlando: Laser Intsitute of America, 2001.

[2] Z. Sun and J. C. Ion, “Review Laser welding of dissimilar metal combinations,” *J. Mater. Sci.*, vol. 30, pp. 4025–4214, 1995.

[3] C. Huang, X. Wang, Y. W. Wu, D. D. Meng, and H. X. Liu, “Experimental Study of Laser Direct Joining of Metal and Carbon Fiber Reinforced Nylon,” *Key Eng. Mater.*, vol. 620, pp. 42–48, 2014.

[4] Y. Cai, C. Huang, H. X. Liu, D. D. Meng, Y. W. Wu, and X. Wang, “Modeling and Optimization of Laser Direct Joining Process Parameters of Titanium Alloy and Carbon Fiber Reinforced Nylon Based on Response Surface Methodology,” *Key Eng. Mater.*, vol. 620, pp. 3–9, Aug. 2014.

[5] C. Lamberti, T. Solchenbach, P. Plapper, and W. Possart, “Laser assisted joining of hybrid polyamide-aluminum structures,” *Phys. Procedia*, vol. 56, no. C, pp. 845–853, 2014.

[6] C. Lamberti, “Optimierung und Charakterisierung einer mittels Laserstrahl gefügten Verbindung zwischen Aluminium und Polyamid 6.6,” PhD-Thesis, University of Luxembourg, 2018.

[7] M. Weigl, F. Albert, and M. Schmidt, “Enhancing the Ductility of Laser-Welded Copper-Aluminum Connections by using Adapted Filler Materials,” *Phys. Procedia*, vol. 12, no. PART 2, pp. 332–338, 2011.

[8] M. M. Hailat, A. Mian, Z. A. Chaudhury, G. Newaz, R. Patwa, and H. J. Herfurth, “Laser micro-welding of aluminum and copper with and without tin foil alloy,” *Microsyst. Technol.*, vol. 18, no. 1, pp. 103–112, Jan. 2012.

[9] T. Solchenbach, P. Plapper, and W. Cai, “Electrical performance of laser braze-welded aluminum–copper interconnects,” *J. Manuf. Process.*, vol. 16, no. 2, pp. 183–189, Apr. 2014.

[10] P. Schmalen and P. Plapper, “Evaluation of Laser Braze-welded Dissimilar Al-Cu Joints,” *Phys. Procedia*, vol. 83, pp. 506–514, 2016.

[11] A. P. Costa, L. Quintino, and M. Greitmann, “Laser beam welding hard metals to steel,” *J. Mater. Process. Technol.*, vol. 141, no. 2, pp. 163–173, Oct. 2003.

[12] C. Barbatti, J. Garcia, G. Liedl, and A. Pyzalla, “Joining of cemented carbides to steel by laser beam welding,” *Materwiss. Werksttech.*, vol. 38, no. 11, pp. 907–914, Nov. 2007.

[13] Z. Mirski, K. Granat, and S. Stano, “Possibilities of laser-beam joining cemented carbides to steel,” *Weld. Int.*, vol. 30, no. 3, pp. 187–191, Mar. 2016.

[14] T. Solchenbach and P. Plapper, “Combined Laser Beam Braze-Welding Process for Fluxless Al-Cu Connections,” in *International Conference on Competetive Manufacturing*, 2012, vol. 3.

[15] P. Schmalen and P. Plapper, “SPECTROSCOPIC STUDIES OF DISSIMILAR AL-CU LASER WELDING,” in *Proceedings of the ASME 2018 13th International Manufacturing Science and Engineering Conference*, 2018, pp. 1–7.

[16] M. Schiry and P. Plapper, “Elucidation of influencing parameters of the laser butt welding process of dissimilar steel to tungsten alloy sheets,” in *International Conference on Competitive Manufacturing COMA’19*, 2019, pp. 1–7.

[17] A. Robert and T. Debroy, “Geometry of laser spot welds from dimensionless numbers,” *Metall. Mater. Trans. B*, vol. 32, no. 5, pp. 941–947, Oct. 2001.

[18] M. Akbari, S. Saedodin, D. Toghraie, R. Shoja-Razavi, and F. Kowsari, “Experimental and numerical investigation of temperature distribution and melt pool geometry during pulsed laser welding of Ti6Al4V alloy,” *Opt. Laser Technol.*, vol. 59, pp. 52–59, Jul. 2014.

[19] K. Kazemi and J. A. Goldak, “Numerical simulation of laser full penetration welding,” *Comput. Mater. Sci.*, vol. 44, no. 3, pp. 841–849, Jan. 2009.

[20] M. Schiry and P. Plapper, “Influence of laser parameter on weld bead shape and process stability of laser welded steel sheets,” *10th Int. Conf. Photonic Technol. LANE 2018*, 2018.

7 BIOGRAPHY

Peter Plapper obtained his PhD degree in Machine tools and Production Engineering (WZL) from the Technical University of Aachen (RWTH). Since 2010 he is Professor for Tool Machines and Production Technologies at the University of Luxembourg.

Marc Schiry holds a M.Eng. degree in General Mechanical Engineering from the University of Applied Sciences Trier. He is currently PhD student at the Laser-Technology-Competence-Center (LTCC) of the University of Luxembourg. His research is focused on laser welding of dissimilar materials.

Machine Learning in Additive and Subtractive Manufacturing as Enabler for Sustainable Manufacturing: A Review

A. du Preez [1], G.A. Oosthuizen [1]
[1]Department of Industrial Engineering, University of Stellenbosch, South Africa,

Abstract
Machine learning is becoming an increasingly popular concept in the modern world since its main goal is to optimize systems by allowing one to make smarter, effective use of resources. The increasing availability of digitalized data, greater computational power and more powerful machine learning algorithms also furthers the attractiveness of machine learning applications. In the manufacturing industry machine learning can lead to cost savings, time savings, increased quality and waste reduction. This study used a systematic review to investigate the different machine learning algorithms applied to additive and subtractive manufacturing within the sustainable manufacturing context.

Keywords
Machine learning, Additive manufacturing, Subtractive manufacturing

1 INTRODUCTION

Machine learning (ML) consists of algorithms, which are sets of rules, that employ mathematical and statistical techniques to aid in the design of computer programs to enable them to study, learn from, identify trends and patterns, and determine similarities in data [1]. Thereafter, predictions and decisions are made based on what was learned, without being explicitly programmed to do so. ML applications are becoming increasingly popular due to the increasing availability of digitalized data, greater computational power and more powerful ML algorithms.

Machine learning have been applied to a variety of additive and subtractive manufacturing (ASM) processes to reduce production time and to increase quality. Quality applications within the additive manufacturing (AM) industry include improved surface finish, minimized support structures, increased structural strength, increased stiffness, reduced warp deformation, increased dimensional accuracy, etc. Quality applications within the subtractive manufacturing (SM) industry include improved surface roughness, tool wear monitoring for improved surface roughness, reduced work piece or tool vibration, reduced operation temperature, improved accuracy, etc.

This paper focuses on ML applications in additive and subtractive manufacturing. Firstly, the methodology is discussed followed by ML techniques and their applications in the industry.

2 METHODOLOGY

2.1 Systematic review

The research methodology used for this study is the systematic review. The systematic review enables the growth of a knowledge base consisting of relevant and useful information, generates information based on research conducted in the areas of study which are of interest and identifies opportunities for further investigation [2]. A systematic review makes use of a pre-specified criteria to collect, evaluate and summarize the collected empirical evidence and research to answer a well-defined research question.

The focus of this paper is to review the different ML techniques which have been applied in ASM, discover the application purposes and analysing the success of the applications. The literature review covers full papers from 2000 to 2018 which are selected according to the criteria provided in Table 1. The template was created by [3] and modifications were added by the author.

3 MACHINE LEARNING ALGORITHMS

In this section the most common ML algorithms used in industry are summarized.

3.1 Neural networks

A neural network (NN) or artificial neural network (ANN) is an arrangement of statistical algorithms which structure is based on the biological brain patterns found in human brains. NNs are used to identify and create the non-linear mathematical relationships between input variables and the output variable(s). NNs are applied to classification, estimation/regression, simulation and prediction problems [4].

Criteria	***Desired value***
Industrial sector of the application	Manufacturing
Specific process	Additive and subtractive manufacturing
Purpose of the study	Quality assurance, optimization, sustainability, waste reduction, cost reduction, time reduction,
Keywords	Machine learning, artificial intelligence, optimization, simulation, additive manufacturing, rapid prototyping, layer manufacturing 3D printing, cutting process, cutting tools design, quality, cost, time, waste, sustainability,
Date of publication	January 2000 – April 2018

Table 1 - The selection criteria for the literature.

A NN is an interconnected parallel network consisting of 3 or more parallel layers: input layer, hidden layer(s) and output layer. Each layer consists of parallel neurons, which uses weights, biases and transfer or activation functions to create a model which best describe the non-linear relationship between the input and output variables. The input layer has a number of neurons equal to the number of input variables and the same holds for the output layer and output variables. The transfer functions (also called neuron functions) are mathematical functions and examples include sigmoid-logistic [5], linear, tangent-sigmoidal [6], hyperbolic trigonometric functions, exponential and Gaussian [7] activation functions. Figure 1 shows the structure of a basic NN.

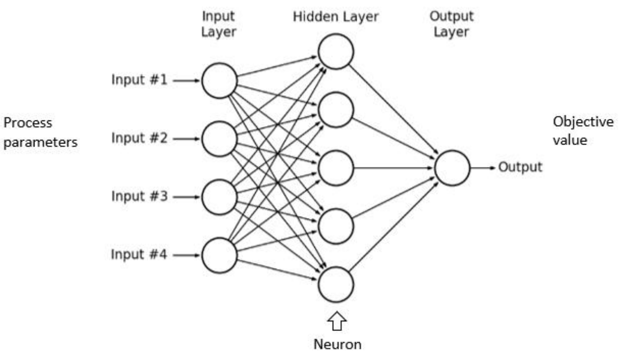


Figure 1 - A basic feed-forward neural network.

3.2 Evolutionary and swarm intelligence-based algorithms

These algorithms are population-based heuristic optimization algorithms and they are derived from natural biological processes. These algorithms include genetic algorithm (GA), non-dominated sorting genetic algorithm (NSGA), particle swarm optimization (PSO), biogeography based optimization algorithm (BBO), firefly algorithm, artificial bee colony (ABC) and ant colony optimization (ACO) [8]. They share common control parameters like population size and number of generations, while each has its own set of algorithm-specific parameters.

The genetic algorithm (GA) is an evolutionary algorithm based on Darwin's theory of natural selection and evolution [7]. At the start an initial population consisting of chromosomes (feasible solutions) is randomly generated. Each individual or solution is evaluated according to the fitness function. The fitness or objective function is a user specified criteria which determines the output variable value of a solution given its input variable characteristics (its genes) [9]. Three primary genetic operations occur to create the next generation or population in order to explore the solution space: reproduction, crossover and mutation. Figure 2 shows the genetic algorithm and the processes of reproduction, crossover and mutation.

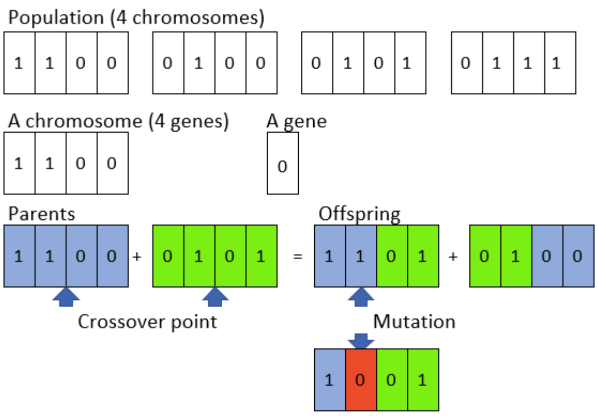


Figure 2 - The genetic algorithm.

3.3 Response surface methodology

Response surface methodology (RSM) is a mathematical and statistical process which creates empirical models which approximates the true functional relationship between the response surface (dependent variable(s)) and a set of experimental input parameters (independent variables). The resulting function is the response function [10] and two types of functions are created: linear (Equation 1) and quadratic (second-order) (Equation 2) [11]. Various methods are available to help determine the RSM model, including the Box bekhen design, ANOVA and design of experiments (DOE), like Taguchi, Central Composite Design (CCD) [12] and full factorial design [13].

$$y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \varepsilon \quad (1)$$

$$y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \beta_i x_i^2 + \sum_{i=1}^{n-1} \sum_{t=i+1}^{n} (\beta_{it} x_i x_t) + \varepsilon \quad (2)$$

3.4 Other machine learning techniques

A wide variety of ML algorithms are used in industry, including finite element analysis (FEA), fuzzy logic, regression analysis, simulated annealing (SA), Solid Isotropic Microstructure with Penalisation (SIMP) method, support vector machines (SVM), hidden Markov model (HMM) and artificial neural fuzzy inference systems (ANFIS).

4 MACHINE LEARNING APPLICATIONS IN ADDITIVE AND SUBTRACTIVE MANUFACTURING

Traditionally process parameters are determined by the operator's experience [4], the conservative technological data provided by the manufacturing equipment manufacturers [5] and trial-and-error operations. This leads to inconsistent manufacturing performance since operator's experience is limited and subjective while the manufacturer data is based on safety-conscious principles and it only includes applications on certain machining materials [14]. New materials are constantly developed, for example titanium alloys, aluminium alloys, advanced plastics, etc. Trail-and-error operations employ post-process techniques to inspect the quality of the finished product. This methodology includes a range of disadvantages: it is costly, time-consuming and it leads to numerous defective and useless products which are only discovered once the process has been completed [15].

Machine learning addresses these resource efficiency challenges by determining the optimal process parameters given an objective(s) through simulations without repeatedly producing physical products. ML also increases sustainability since it leads to the permanent availability of uniform, objective ASM process knowledge (manufacturers do not have to hire costly consultants repeatedly), it enables manufacturers to optimally benefit from their manufacturing equipment without the acquisition of new costly, carbon footprint related equipment and it reduces the usage of valuable resources including time, money, energy and natural resources. ML also enables employee safety as well as product safety, since it can be used proactively to allow one to view the process parameters before application. Thus harmful parameters can be identified before the process has started.

There are complex interrelationships between process parameters, process output, economic factors and environmental factors. The parameters directly affect the production efficiency, production cost, product quality, tool life (in the case of SM), processing time, power energy consumption and the carbon emissions [16]. All the ML applications considered during the systematic review investigated one or combinations of these sustainability-related relationships.

4.1 Additive Manufacturing

During the systematic review 70 AM articles have been randomly selected and inspected. For full details on the process, content and references, please refer to [17]. ML algorithms have been applied to a variety of AM process types including 3D printing (3DP), directed energy deposition (DED) electron beam melting or manufacturing (EBM), fused deposition modelling (FDM), laser engineered net shaping (LENS), laminated object manufacturing (LOM), stereolithography (SLA), selective laser cladding (SLC), selective laser melting (SLM), selective laser sintering (SLS) and wire + arc additive layer manufacture (WAALM). Fig. 3. illustrates the AM processes supported by ML applications. The percentages indicate the relative ratios. It is evident that FDM is the field in which the most applications have been applied, followed by SLC, SLS and AM in general (for example aiding in modelling and design) applications. Fig. 4. illustrates the different types of ML algorithms which have been applied in AM processes. ANNs are the most common application, followed by GA, regression modelling and RSM.

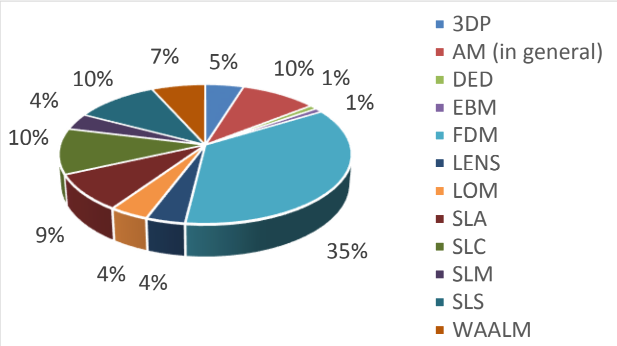


Figure 3 - Additive manufacturing processes supported by machine learning applications.

The articles were further inspected to determine the accuracy or prediction error of the different ML algorithms and to compare them. Table 2 provides a summary. Two measurements of accuracy are used, namely the coefficient of regression, R^2 (Equation 3) [4], and relative error, RE (Equation 4) [18].

$$R^2 = 1 - \left(\frac{\sum(T-P)^2}{\sum P^2}\right) \quad (3)$$

$$RE(\%) = \frac{|P-T|}{T} \times 100 \quad (4)$$

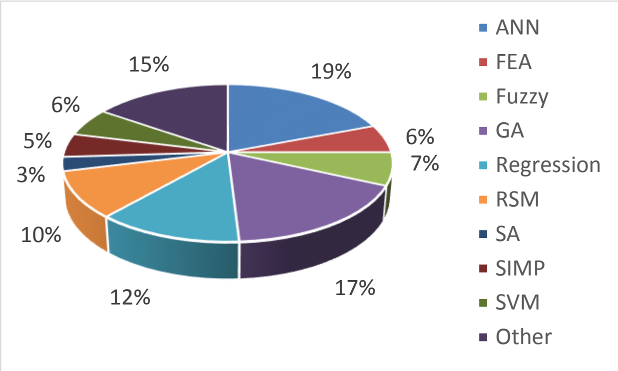


Figure 4 - The different types of machine learning applications applied in additive manufacturing processes.

The objective of the ML applications was included as well. It is interesting to note the small RE for ANNs.

AM Process	Objective	Algorithm & Success
FDM	Surface roughness prediction	ANN - R^2 = 0.996 [4]
	Compressive strength prediction	ANN - RE = 1,2% [18], ANN - RE = 1,7% [7], ANN - RE = 0.2% [19] GA - RE = 4.9% [18], GA - RE = 8,0% [7], GA - RE = 3.5% [20] Fuzzy - RE = 6.1% [20] Regr. - RE = 2,5% [18], Regr. - RE = 1,7% [19] SVM - RE= 10.1% [20]
	Wear strength	ANN - RE =3.4% [5] GA - RE = 5,8% [5] SVM - RE = 5.9% [5]
	Dimensional Accuracy prediction	Fuzzy - RE = 4,1% [21]
SLS	Shrinkage prediction	ANN - R^2 = 0.93 [22] RSM - R^2 = 0.72 [22]
	% porosities prediction	Regr. - RE = 1,5% [23]

Regr. = regression modelling

Table 2 - The ML algorithm successes for AM.

4.2 Subtractive Manufacturing

During the systematic review 70 SM articles have been randomly selected and inspected ([8], [10]. [12 -16], [24-86]). Machine learning algorithms have been applied to a variety of cutting processes including cutting, turning, milling, drilling, boring, grinding, broaching, coroning, electric discharge machining (EDM), ultrasonic-assisted EDM (US/EDM), wire EDM (WEDM), abrasive water jet machining (AWJM), laser cutting process, electro-chemical machining (ECM) and focused ion beam (FIB) micro- milling.

Fig. 5. illustrates the SM supported ML applications. It is evident that turning and milling are the fields in which the most applications have been applied, followed by EDM and drilling. Fig. 6. illustrates the different types of ML algorithms which have been applied in SM. ANNs are the most common application, followed by GA, RSM and PSO.

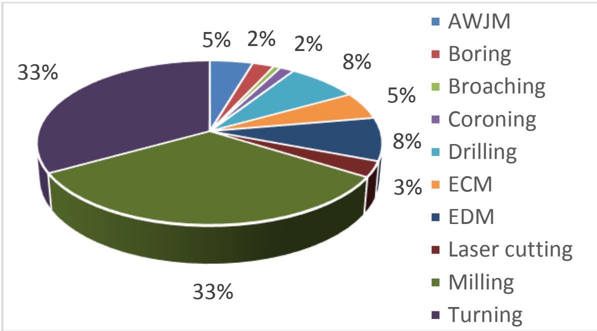


Figure 5 - Subtractive manufacturing processes supported by machine learning applications.

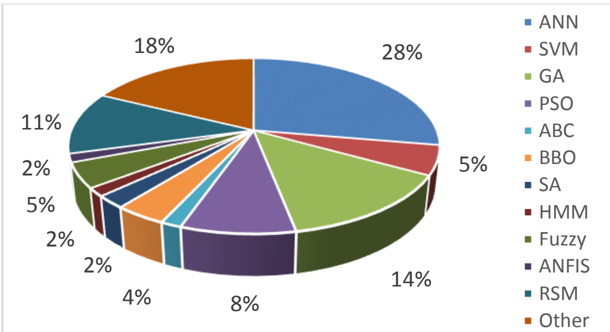


Figure 6 - The different types of machine learning applications applied in subtractive manufacturing processes.

The articles were further inspected to determine the accuracy or prediction error of the different ML algorithms and to compare them. Table 3 provides a summary. Two measurements of accuracy are used, namely the coefficient of regression, R^2 (Equation 3) [4], and relative error, RE (Equation 4) [18].

The objective of the ML applications was included as well. It is interesting to note the small RE for ANNs.

SM Process	Objective	Algorithm & Success
Turning	Surface roughness prediction	ANN (I) - RE = 0.4%, ANN (II) - RE = 0.3%, ANN (III) - RE = 0.4% [12]
		RSM (lin) - RE =14.5%, RSM (q) - RE = 3.5% [12]
		RSM - R^2 (R_a) =0.999, RSM - R^2 (R_t) = 0.964 [13]
		Regr. - RE = 4.9% [12]
	Cutting temperature prediction	ANN -RE(dry) = 0.96%, ANN -RE(hpc) = 0.7% [72]
		RSM - RE(dry) = 0.8%, RSM - RE(hpc) = 0.9% [72]
Milling	Surface roughness prediction	ANN - RE = 2.1% [10], ANN - R^2 = 0.982 [31]
		RSM - (q) RE=2% [10], RSM - R^2 = 0.9713 [31]
	Cutting force prediction	ANN - RE = 3.7% [10]
		RSM (q) - RE = 1.7% [10]
	Tool wear	ANN - RE = 1% [10]
		RSM (q) - RE= 1.4% [10]
	Cutting power prediction	GA - Re = 5.5% [16]
	Tool life prediction	GA - Re = 6.7% [16]

Table 3 - The ML algorithm successes for SM.

5 CONCLUSIONS

In the manufacturing industry machine learning can lead to cost savings, time savings, increased quality and waste reduction. It can also greatly contribute towards sustainability in the manufacturing industry. From the systematic review, trends in ML applications in ASM were identified, investigated and the results presented.

6 REFERENCES

[1] O. Al Sonosy, S. Rady, N. L. Badr, and M. Hashem, ‘Machine Learning Techniques for Mining Location-Based Social Networks for Business Predictions’, *Proc. 10th Int. Conf. Informatics Syst. - INFOS ’16*, pp. 185–190, 2016.

[2] D. Augusto *et al.*, ‘Eco-innovation determinants in manufacturing SMEs : Systematic review and research directions’, vol. 142, 2017.

[3] G. Köksal, I. Batmaz, and M. C. Testik, ‘A review of data mining applications for quality improvement in manufacturing industry’, *Expert Syst. Appl.*, vol. 38, no. 10, pp. 13448–13467, 2011.

[4] A. Boschetto, V. Giordano, and F. Veniali, ‘Surface roughness prediction in fused deposition modelling by neural networks’, pp. 2727–2742, 2013.

[5] V. Vijayaraghavan, A. Garg, J. Siu, L. Lam, B. Panda, and S. S. Mahapatra, ‘Process characterisation of 3D-printed FDM components using improved evolutionary computational approach’, pp. 781–793, 2015.

[6] A. Noriega, D. Blanco, B. J. Alvarez, and A. Garcia, ‘Dimensional accuracy improvement of FDM square cross-section parts using artificial neural networks and an optimization algorithm’, pp. 2301–2313, 2013.

[7] B. N. Panda, M. V. A. R. Bahubalendruni, and B. B. Biswal, ‘A general regression neural network approach for the evaluation of compressive strength of FDM prototypes’, *Neural Comput. Appl.*, vol. 26, pp. 1129–1136, 2015.

[8] R. Shukla and D. Singh, ‘Experimentation investigation of abrasive water jet machining parameters using Taguchi and Evolutionary optimization techniques’, *Swarm Evol. Comput.*, vol. 32, pp. 167–183, 2017.

[9] J. D. Hiller and H. Lipson, ‘DESIGN AUTOMATION FOR MULTI-MATERIAL PRINTING’, pp. 279–287, 2009.

[10] E. Kilickap, A. Yardimeden, and Y. H. Çelik, ‘Mathematical Modelling and Optimization of Cutting Force, Tool Wear and Surface Roughness by Using Artificial Neural Network and Response Surface Methodology in Milling of Ti-6242S’, *Appl. Sci.*, vol. 7, no. 10, p. 1064, 2017.

[11] A. Peng, X. Xiao, and R. Yue, ‘Process parameter optimization for fused deposition modeling using response surface methodology combined with fuzzy inference system’, *Int J Adv Manuf Technol*, pp. 87–100, 2014.

[12] Y. Chen, R. Sun, Y. Gao, and J. Leopold, ‘A nested-ANN prediction model for surface roughness considering the effects of cutting forces and tool vibrations’, *Meas. J. Int. Meas. Confed.*, vol. 98, pp. 25–34, 2017.

[13] Z. Hessainia, A. Belbah, M. A. Yallese, T. Mabrouki, and J. F. Rigal, ‘On the prediction of surface roughness in the hard turning

based on cutting parameters and tool vibrations', *Meas. J. Int. Meas. Confed.*, vol. 46, no. 5, pp. 1671–1681, 2013.

[14] K. M. R. G., R. G., H. R. D., and S. R. M., 'Development of hybrid model and optimization of surface roughness in electric discharge machining using artificial neural networks and genetic algorithm', *J. Mater. Process. Technol.*, vol. 209, no. 3, pp. 1512–1520, 2009.

[15] E. García Plaza and P. J. Núñez López, 'Surface roughness monitoring by singular spectrum analysis of vibration signals', *Mech. Syst. Signal Process.*, vol. 84, pp. 516–530, 2017.

[16] H. Zhang, Z. Deng, Y. Fu, L. Lv, and C. Yan, 'A process parameters optimization method of multi-pass dry milling for high efficiency, low energy and low carbon emissions', *J. Clean. Prod.*, vol. 148, pp. 174–184, 2017.

[17] A. Du Preez and G. A. Oosthuizen, 'MACHINE LEARNING IN ADDITIVE MANUFACTURING AS ENABLER FOR SMART SUSTAINABLE MANUFACTURING: A REVIEW', in *Proceedings - 2018 RAPDASA International Conference on Additive manufacturing, 6-9 Nov, South Africa*, 2018, vol. 19, pp. 12–21.

[18] M. M. S. A. Garg, K. Tai, 'State-of-the-art in empirical modelling of rapid prototyping processes', *Rapid Prototyp. J.*, vol. 20, no. 2, pp. 164–178, 2014.

[19] A. K. Sood, R. K. Ohdar, and S. S. Mahapatra, 'Experimental investigation and empirical modelling of FDM process for compressive strength improvement', *J. Adv. Res.*, vol. 3, no. 1, pp. 81–90, 2012.

[20] A. G. K. Tai and C. H. L. M. M. Savalani, 'A hybrid M5 -genetic programming approach for ensuring greater trustworthiness of prediction ability in modelling of FDM process', *J. Intell. Manuf.*, vol. 25, pp. 1349–1365, 2014.

[21] R. K. Sahu, S. S. Mahapatra, and A. K. Sood, 'A Study on Dimensional Accuracy of Fused Deposition Modeling (FDM) Processed Parts using Fuzzy Logic', *J. Manuf. Sci. Prod.*, vol. 13, no. 3, pp. 183–197, 2013.

[22] S. Negi and R. K. Sharma, 'Study on shrinkage behaviour of laser sintered PA 3200GF specimens using RSM and ANN', *Rapid Prototyp. J.*, vol. 22, no. 4, pp. 645–659, 2016.

[23] H. Liao and L. M. Trung, 'Optimization on Selective Fiber Laser Sintering of Metallic Powder', *Adv. Mater. Res.*, vol. 472–475, pp. 2519–2530, 2012.

[24] J. Wang, J. Xie, R. Zhao, L. Zhang, and L. Duan, 'Multisensory fusion based virtual tool wear sensing for ubiquitous manufacturing', *Robot. Comput. Integr. Manuf.*, vol. 45, pp. 47–58, 2017.

[25] A. Conde, A. Arriandiaga, J. A. Sanchez, E. Portillo, S. Plaza, and I. Cabanes, 'High-accuracy wire electrical discharge machining using artificial neural networks and optimization techniques', *Robot. Comput. Integr. Manuf.*, vol. 49, pp. 24–38, 2017.

[26] R. Corne, C. Nath, M. El, and T. Kurfess, 'Enhancing Spindle Power Data Application with Neural Network for Real - Time Tool Wear / Breakage Prediction during Inconel Drilling', *Procedia Manuf.*, vol. 5, pp. 1–14, 2016.

[27] C. L. Yen, M. C. Lu, and J. L. Chen, 'Applying the self-organization feature map (SOM) algorithm to AE-based tool wear monitoring in micro-cutting', *Mech. Syst. Signal Process.*, vol. 34, no. 1–2, pp. 353–366, 2013.

[28] P. V. S. Suresh, P. V. Rao, and S. G. Deshmukh, 'A genetic algorithmic approach for optimization of surface roughness prediction model', *Int. J. Mach. Tools Manuf.*, vol. 42, pp. 675–680, 2002.

[29] R. Q. Sardiñas, M. R. Santana, and E. Alfonso Brindis, 'Genetic algorithm-based multi-objective optimization of cutting parameters in turning processes', *Eng. Appl. Artif. Intell.*, vol. 19, no. 2, pp. 127–133, 2006.

[30] R. V. Rao, D. P. Rai, and J. Balic, 'A multi-objective algorithm for optimization of modern machining processes', *Eng. Appl. Artif. Intell.*, vol. 61, no. March, pp. 103–125, 2017.

[31] N. E. Karkalos, N. I. Galanis, and A. P. Markopoulos, 'Surface roughness prediction for the milling of Ti-6Al-4V ELI alloy with the use of statistical and soft computing techniques', *Meas. J. Int. Meas. Confed.*, vol. 90, pp. 25–35, 2016.

[32] J. Yan and L. Li, 'Multi-objective optimization of milling parameters - the trade-offs between energy , production rate and cutting quality', *J. Clean. Prod.*, vol. 52, pp. 462–471, 2013.

[33] G. Kant and K. S. Sangwan, 'Prediction and optimization of machining parameters for minimizing power consumption and surface roughness in machining', *J. Clean. Prod.*, vol. 83, pp. 151–164, 2014.

[34] A. K. Gupta, 'Predictive modelling of turning operations using response surface methodology, artificial neural networks and support vector regression', *Int. J. Prod. Res.*, vol. 48, no. 3, pp. 763–778, 2010.

[35] D. Shi and N. N. Gindy, 'Tool wear predictive model based on least squares support vector

machines', *Mech. Syst. Signal Process.*, vol. 21, no. 4, pp. 1799–1814, 2007.

[36] E. Kannatey-Asibu, J. Yum, and T. H. Kim, 'Monitoring tool wear using classifier fusion', *Mech. Syst. Signal Process.*, vol. 85, pp. 651–661, 2017.

[37] K. Venkata Rao, B. S. N. Murthy, and N. Mohan Rao, 'Prediction of cutting tool wear, surface roughness and vibration of work piece in boring of AISI 316 steel with artificial neural network', *Meas. J. Int. Meas. Confed.*, vol. 51, no. 1, pp. 63–70, 2014.

[38] J. Zhou, Z. Duan, Y. Li, J. Deng, and D. Yu, 'PSO-based neural network optimization and its utilization in a boring machine', *J. Mater. Process. Technol.*, vol. 178, pp. 19–23, 2006.

[39] R. Corne, C. Nath, M. El, and T. Kurfess, 'Study of spindle power data with neural network for predicting real-time tool wear / breakage during inconel drilling', *J. Manuf. Syst.*, vol. 43, pp. 287–295, 2017.

[40] K. Patra, A. K. Jha, T. Szalay, J. Ranjan, and L. Monostori, 'Artificial neural network based tool condition monitoring in micro mechanical peck drilling using thrust force signals', *Precis. Eng.*, vol. 48, pp. 279–291, 2017.

[41] M. Sen and H. S. Ã. Shan, 'Electro jet drilling using hybrid NNGA approach', *Robot. Comput. Integr. Manuf.*, vol. 23, pp. 17–24, 2007.

[42] X. Li, S. Dong, and P. K. Venuvinod, 'Hybrid Learning for Tool Wear Monitoring', *Int J Adv Manuf Technol*, pp. 303–307, 2000.

[43] M. Sen, H. S. Shan, M. Sen, and H. S. Shan, 'Optimal Selection of Machining Conditions in the Electrojet Drilling Process Using Hybrid NN-DF-GA Approach Optimal Selection of Machining Conditions in the Electrojet Drilling Process Using Hybrid NN-DF-GA Approach', *Mater. Manuf. Process.*, vol. 6914, pp. 349–356, 2007.

[44] P. Asokan, R. R. Kumar, R. Jeyapaul, and M. Santhi, 'Development of multi-objective optimization models for electrochemical machining process', pp. 55–63, 2008.

[45] M. C. Panda and V. Yadava, 'Intelligent Modeling and Multiobjective Optimization of Die Sinking Electrochemical Spark Machining Process Intelligent Modeling and Multiobjective Optimization of Die Sinking Electrochemical Spark Machining Process', *Mater. Manuf. Process.*, vol. 6914, pp. 10–25, 2012.

[46] D. Devarasiddappa, J. George, M. Chandrasekaran, and N. Teyi, 'Application of Artificial Intelligence Approach in Modeling Surface Quality of Aerospace Alloys in WEDM Process', *Procedia Technol.*, vol. 25, no. Raerest, pp. 1199–1208, 2016.

[47] M. R. Shabgard, M. A. Badamchizadeh, G. Ranjbary, and K. Amini, 'Fuzzy approach to select machining parameters in electrical discharge machining (EDM) and ultrasonic-assisted EDM processes', *J. Manuf. Syst.*, vol. 32, no. 1, pp. 32–39, 2013.

[48] U. Zuperl and F. Cus, 'Surface roughness fuzzy inference system within the control simulation of end milling', *Precis. Eng.*, vol. 43, pp. 530–543, 2016.

[49] P. . Benardos and G. . Vosniakos, 'Prediction of surface roughness in CNC face milling using neural networks and Taguchi's design of experiments', *Robot. Comput. Integr. Manuf.*, vol. 18, no. 5, pp. 343–354, 2002.

[50] H. Saglam and A. Unuvar, 'Tool condition monitoring in milling based on cutting forces by a neural network', *Int. J. Prod. Res.*, vol. 41, no. 7, pp. 1519–1532, 2003.

[51] F. Arnold, A. Hänel, A. Nestler, and A. Brosius, 'New Approaches for the Determination of Specific Values for Process Models in Machining Using Artificial Neural Networks', *Procedia Manuf.*, vol. 11, pp. 1463–1470, 2017.

[52] A. M. Zain, H. Haron, and S. Sharif, 'Prediction of surface roughness in the end milling machining using Artificial Neural Network', *Expert Syst. Appl.*, vol. 37, no. 2, pp. 1755–1768, 2010.

[53] G. Kant and K. Singh, 'Predictive Modelling and Optimization of Machining Parameters to Minimize Surface Roughness using Artificial Neural Network Coupled with Genetic Algorithm', *Procedia CIRP*, vol. 31, pp. 453–458, 2015.

[54] N. Ghosh *et al.*, 'Estimation of tool wear during CNC milling using neural network-based sensor fusion', *Mech. Syst. Signal Process.*, vol. 21, no. 1, pp. 466–479, 2007.

[55] B. Das, S. Roy, R. N. Rai, and S. C. Saha, 'Study on machinability of in situ Al–4.5%Cu–TiC metal matrix composite-surface finish, cutting force prediction using ANN', *CIRP J. Manuf. Sci. Technol.*, vol. 12, pp. 67–78, 2016.

[56] C. Drouillet, J. Karandikar, C. Nath, A. Journeaux, M. El, and T. Kurfess, 'Tool life predictions in milling using spindle power with the neural network technique', *J. Manuf. Process.*, vol. 22, pp. 161–168, 2016.

[57] P. Krishnakumar, K. Rameshkumar, and K. I. Ramachandran, 'Tool wear condition prediction using vibration signals in high speed machining (HSM) of Titanium (Ti-6Al-

4V) alloy', *Procedia Comput. Sci.*, vol. 50, pp. 270–275, 2015.

[58] T. Benkedjouh, K. Medjaher, N. Zerhouni, and S. Rechak, 'Health assessment and life prediction of cutting tools based on support vector regression', *J. Intell. Manuf.*, vol. 26, no. 2, pp. 213–223, 2015.

[59] I. Svalina, G. Šimunović, T. Šarić, and R. Lujić, 'Evolutionary neuro-fuzzy system for surface roughness evaluation', *Appl. Soft Comput. J.*, vol. 52, pp. 593–604, 2017.

[60] H. Öktem, T. Erzurumlu, and H. Kurtaran, 'Application of response surface methodology in the optimization of cutting conditions for surface roughness', *J. Mater. Process. Technol.*, vol. 170, no. 1–2, pp. 11–16, 2005.

[61] C. Li, X. Chen, Y. Tang, and L. Li, 'Selection of optimum parameters in multi-pass face milling for maximum energy ef fi ciency and minimum production cost', *J. Clean. Prod.*, vol. 140, pp. 1805–1818, 2017.

[62] B. Cuka and D. W. Kim, 'Fuzzy logic based tool condition monitoring for end-milling', *Robot. Comput. Integr. Manuf.*, vol. 47, no. December 2016, pp. 22–36, 2017.

[63] T. L. (Bill) Tseng, U. Konada, and Y. (James) Kwon, 'A novel approach to predict surface roughness in machining operations using fuzzy set theory', *J. Comput. Des. Eng.*, vol. 3, no. 1, pp. 1–13, 2016.

[64] D. Wu *et al.*, 'A fog computing-based framework for process monitoring and prognosis in cyber-manufacturing', *J. Manuf. Syst.*, vol. 43, pp. 25–34, 2017.

[65] J. Zhang, B. Starly, Y. Cai, P. H. Cohen, and Y. Lee, 'Particle learning in online tool wear diagnosis and prognosis ☆', *J. Manuf. Process.*, vol. 28, pp. 457–463, 2017.

[66] W. Yang, Q. Zhou, and K. Tsui, 'Differential evolution-based feature selection and parameter optimisation for extreme learning machine in tool wear estimation', *Int. J. Prod. Res.*, vol. 7543, p. 0, 2016.

[67] C. K. Madhusudana, H. Kumar, and S. Narendranath, 'Condition monitoring of face milling tool using K-star algorithm and histogram features of vibration signal', *Eng. Sci. Technol. an Int. J.*, vol. 19, no. 3, pp. 1543–1551, 2016.

[68] C. Lu, 'Study on prediction of surface quality in machining process', *J. Mater. Process. Technol.*, vol. 205, no. 1–3, pp. 439–450, 2008.

[69] K. Singh, S. Saxena, and G. Kant, 'Optimization of Machining Parameters to Minimize Surface Roughness using Integrated ANN-GA Approach', *Procedia CIRP*, vol. 29, pp. 305–310, 2015.

[70] T. Özel and Y. Karpat, 'Predictive modeling of surface roughness and tool wear in hard turning using regression and neural networks', *Int. J. Mach. Tools Manuf.*, vol. 45, no. 4–5, pp. 467–479, 2005.

[71] C. Ahilan, S. Kumanan, N. Sivakumaran, and J. Edwin Raja Dhas, 'Modeling and prediction of machining quality in CNC turning process using intelligent hybrid decision making tools', *Appl. Soft Comput. J.*, vol. 13, no. 3, pp. 1543–1551, 2013.

[72] M. Mia and N. R. Dhar, 'Response surface and neural network based predictive models of cutting temperature in hard turning', *J. Adv. Res.*, vol. 7, no. 6, pp. 1035–1044, 2016.

[73] M. Mia and N. R. Dhar, 'Prediction of surface roughness in hard turning under high pressure coolant using Artificial Neural Network', *Meas. J. Int. Meas. Confed.*, vol. 92, pp. 464–474, 2016.

[74] B. Das, S. Roy, R. N. Rai, and S. C. Saha, 'Studies on effect of cutting parameters on surface roughness of Al-Cu-TiC MMCs: An artificial neural network approach', *Procedia Comput. Sci.*, vol. 45, no. C, pp. 745–752, 2015.

[75] G. M. A. Acayaba and P. M. de Escalona, 'Prediction of surface roughness in low speed turning of AISI316 austenitic stainless steel', *CIRP J. Manuf. Sci. Technol.*, vol. 11, pp. 62–67, 2015.

[76] M. Hanief, M. F. Wani, and M. S. Charoo, 'Modeling and prediction of cutting forces during the turning of red brass (C23000) using ANN and regression analysis', *Eng. Sci. Technol. an Int. J.*, vol. 20, no. 3, pp. 1220–1226, 2017.

[77] D. R. Salgado and F. J. Alonso, 'An approach based on current and sound signals for in-process tool wear monitoring', *Int. J. Mach. Tools Manuf.*, vol. 47, no. 14, pp. 2140–2152, 2007.

[78] D. R. Salgado, F. J. Alonso, I. Cambero, and A. Marcelo, 'In-process surface roughness prediction system using cutting vibrations in turning', *Int. J. Adv. Manuf. Technol.*, vol. 43, no. 1–2, pp. 40–51, 2009.

[79] S. Dutta, S. K. Pal, and R. Sen, 'On-machine tool prediction of flank wear from machined surface images using texture analyses and support vector regression', *Precis. Eng.*, vol. 43, pp. 34–42, 2016.

[80] M. Elangovan, V. Sugumaran, K. I. Ramachandran, and S. Ravikumar, 'Effect of SVM kernel functions on classification of

vibration signals of a single point cutting tool', *Expert Syst. Appl.*, vol. 38, no. 12, pp. 15202–15207, 2011.

[81] K. Singh and G. Kant, 'Optimization of Machining Parameters for Improving Energy Efficiency using Integrated Response Surface Methodology and Genetic Algorithm Approach', *Procedia CIRP*, vol. 61, pp. 517–522, 2017.

[82] R. Teti, 'Genetic algorithm-based optimization of cutting parameters in turning processes', *Procedia CIRP*, vol. 7, pp. 323–328, 2013.

[83] T. L. (Bill) Tseng, F. Jiang, and Y. (James) Kwon, 'Hybrid Type II fuzzy system & data mining approach for surface finish', *J. Comput. Des. Eng.*, vol. 2, no. 3, pp. 137–147, 2015.

[84] R. Dhabale, V. S. Jatti, and T. P. Singh, 'Multi-objective Optimization of Turning Process During Machining of AlMg1SiCu Using Non-dominated Sorted Genetic Algorithm', *Procedia Mater. Sci.*, vol. 6, no. Icmpc, pp. 961–966, 2014.

[85] I. Hanafi, F. M. Cabrera, F. Dimane, and J. T. Manzanares, 'Application of Particle Swarm Optimization for Optimizing the Process Parameters in Turning of PEEK CF30 Composites', *Procedia Technol.*, vol. 22, no. October 2015, pp. 195–202, 2016.

[86] A. R. Yildiz, 'Optimization of cutting parameters in multi-pass turning using artificial bee colony-based approach', *Inf. Sci. (Ny).*, vol. 220, pp. 399–407, 2013.

7 BIOGRAPHY

Prof Gert Adriaan Oosthuizen is employed as an Associate Professor at the Department of Industrial Engineering at Stellenbosch University. He is director of the Stellenbosch Technology Centre and his research areas include smart, sustainable manufacturing and social innovation.

Anli du Preez holds a B.Eng in Industrial Engineering from the University of Stellenbosch. She is currently a postgraduate student in the Department of Industrial Engineering. Her specific focus and area of research interest is in the application of machine learning in the manufacturing field.

Development and Characterisation of a Photocurable Gelatin/Alginate Bioink for 3D Bioprinting

H. Mishbak[1,2], C. Vyas[2], G. Cooper[2], P. Bartolo[2]

[1] School of Engineering, University of Thi-Qar, Thi-Qar, Iraq

[2] School of Mechanical, Aerospace and Civil Engineering, University of Manchester, Manchester, UK

Abstract

Injury to articular cartilage tissue can lead to long term pain and loss of mobility. The poor regenerative capacity of articular cartilage has led to different treatment strategies, among them, tissue engineering and bioprinting have been widely investigated. Many studies have focused on the use of a single polymeric material processed with or without cells for cartilage applications. This paper focuses on the design of a suitable photopolymerisable bioink based on alginate/gelatin for 3D bioprinting in cartilage tissue engineering. Alginate/Gelatin were chemically-modified with methacrylate anhydride groups and mixed with a photoinitiator solution to obtain a photocrosslinked hydrogels, functionalisation was confirmed through nuclear magnetic resonance. Resulting bioinks were processed using ultraviolet light (UV) radiation (8 mW/cm^2) for 8 min and were assessed for their morphological characteristics including swelling and degradation. Mechanical compression test and rheological characterisation were investigated. 3D cell culture shows that this system has overcome some cell compatibility issues associated with a lack of adhering motifs in alginate alone and 2D cell culture as the cell morphology was spherical compared to the 2D culture.

Keywords

Hybrid hydrogels, functionalisation, rheological characterisation, mechanical characterisation and cell encapsulation

1 INTRODUCTION

Hydrogels are a three-dimensional (3D) hydrated water-swollen polymer networks that can swell up to hundreds times their own weight [1]. Hydrogels have been receiving attention due to their potential use in a wide range of biomedical applications, including tissue engineering scaffolds [2], drug delivery [3], contact lenses, corneal implants and wound dressings [4], and articular cartilage repair [5][6]. Hydrogels are attractive class of biomaterial for the development or regeneration of soft tissues due to their excellent biocompatibility, oxygen and nutrient permeability, the capability to modify and tailor their mechanical properties, and mimicking the native extracellular matrix (ECM) [7]. Articular cartilage tissue engineering holds great potential for enhancing articular cartilage therapy that could allow the repair and regeneration of injured or diseased tissue. A variety of biodegradable and biocompatible hydrogels have already been used for cartilage applications [8]. Synthetic hydrogels have been investigated for the last two decades as a potential biomaterial for mimicking the ECM and can be tailored to have similar properties of ECM in terms of mechanical, morphological and biological properties [9]. However, typically synthetic hydrogels exhibit linear elasticity which is in contrast to physiological ECM that is characterised by a viscoelastic behaviour [8]. Furthermore, synthetic hydrogels have limitations regarding their biocompatibility and degradation properties compared with natural polymers[10].

Natural hydrogels are an alternative biomaterial that have inherent biocompatibility and biodegradability, and have potential for cartilage applications [11]. Among them, alginate has been widely investigated due to its relatively low cost, natural origin and ease of handling. Gelatin is another natural polymer that has been extensively explored due to its biocompatibility, biodegradability, promotion of cell adhesion and proliferation [12].

Alginate gels are currently being explored for cell encapsulation and drug delivery [13][14][15]. However, alginate has limitations such as lack of cell binding motifs, poor mechanical properties and slow gelation speed which affects gel uniformity and strength [16].

To improve the properties of alginate hydrogels and to overcome its limitations, alginate precursor polymer solution has been mixed with gelatin polymer solution to form a hybrid hydrogel system. This prominent approach in 3D hybrid hydrogel, the combination of different polymers, aims to create a hydrogel network with improved properties.

In this study, we present the synthesis and mechanical characterisation of an alginate-gelatin methacrylate hydrogel (Alg-Gel)MA and demonstrate dynamic control over the viscoelastic properties. It has been demonstrated that the hybrid hydrogel system prepared from alginate methacrylate and gelatin methacrylate has overcome the limitations of using single hydrogel system. The hybrid hydrogel system enhances the mechanical properties, improves swelling and degradation kinetics, and allows cell adhesion and appropriate morphology.

2 MATERIALS AND METHODS

2.1 Synthesis of hybrid hydrogel

Hybrid hydrogel system was synthesised in four steps:

2.1.1 Photoinitiator solution preparation

The photoinitiator solution was prepared by dissolving VA-086, 2,2'-Azobis[2-methyl-N-(2-hydroxyethyl) propionamide azo initiator (Wako Pure Chemical Industries, USA) in Dulbecco's Phosphate Buffered Saline (DPBS) (Sigma-Aldrich, UK) at 40°C for 30 min, to a final concentration of 0.5-1.5% w/v. The solution was filter-sterilised for further use.

2.1.2 Alginate methacrylate preparation AlgMA

Sodium alginate (Sigma-Aldrich, UK) was dissolved in DPBS at a concentration of 2 wt. % and then mixed with methacrylate anhydride (MA) (Sigma-Aldrich, UK) at 20% v/v of alginate solution. The pH of the solution was maintained between 7.4-8.0 during the reaction time by the addition of 5M NaOH. A reaction time of 24 h was used. After the chemical modification, the polymer solution was precipitated with cold ethanol, dried in an oven overnight at 40-50ºC and purified through dialysis for six days. The solution was frozen at -80ºC and the polymer recovered by lyophilisation.

2.1.3 Gelatin methacrylate preparation GelMA

Methacrylated gelatin was prepared by dissolving gelatin bovine skin type B (Sigma-Aldrich, UK) at a concentration of 12.5 wt. %, in DPBS at a temperature of 40ºC. After gelatin dissolution, MA (10x molar excess) was added under vigorous stirring. The pH of the solution was kept at ~8.0 during the reaction. The reaction time was 6 hours. The solution was purified through dialysis for six days, frozen at -80°C and the polymer recovered by lyophilisation.

2.1.4 Hybrid hydrogel preparation (Alg-Gel)MA

Hybrid hydrogel system was synthesised by dissolving 2 wt. % AlgMA in the photoinitiator solution prepared in step 1. Dissolving 10 wt. % GelMA in the photoinitiator solution prepared in step 1. Mixing both solutions at 50:50 v/v %. Crosslinked discs (cell-free constructs) were produced by pipetting the (Alg-Gel)MA solution into custom-made cylindrical Teflon moulds (diameter: 8 mm; height 4 mm) (Fig. 1). Photopolymerisation was conducted using a 365 nm UV light (Dymax 2000-EC, Dymax, Germany) with irradiation at 8 mW/cm^2 for 8 min.

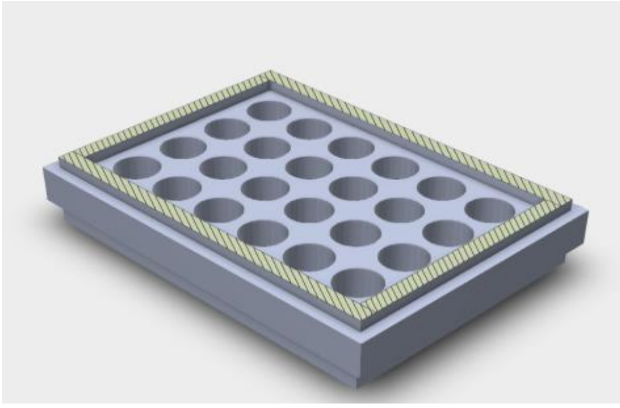


Figure 1- Custom-made cylindrical Teflon moulds (diameter: 8 mm; height 4 mm).

2.2 Structural conformations of the hybrid hydrogels and rheological characterisation

Nuclear magnetic resonance spectroscopy (1H NMR, B400 Bruker Avance III, USA) at 400 MHz was used to evaluate the structure of the hybrid system (Alg-Gel)MA. The hydrogels were dissolved in deuterium oxide (D_2O), transferred to NMR tubes and the spectra acquired with 128 scans.

The rheological analyses were carried out using a DHR2 rheometer (TA Instrument, USA). Rotational tests were considered. The rotational speed depends on viscosity computed by means of stress and shear rate. Amplitude oscillatory shear measurements tests were considered, the shear storage modulus, G′, loss modulus, G″ and linear-viscoelastic (LVE) behaviour were quantified.

2.3 Morphology, swelling and degradation kinetics of the hybrid hydrogels

Porosity is an important property of a hydrogel structure due to the influence on the mechanical performance of the scaffold, permeability and the supply of oxygen, nutrients and removal of waste products. A porous structure provides free space for cells to proliferate, interact with other cells, and produce ECM. The morphology of the internal structure of the hydrogels was investigated by scanning electron microscopy (SEM, S3000N, Hitachi, Japan). The hydrogels specimens were frozen for 3 days then lyophilised and coated using a platinum sputter-coating.

The swelling and degradation kinetics of the hydrogels were investigated. The swelling ratio can impact on the permeability of nutrients, gases and waste products. Hydrogel degradation should also match the rate of tissue regeneration. Hybrid hydrogel cell-free discs were frozen at -80ºC, then, lyophilised, and the dry weights (Wi) were measured. The dry hydrogel samples were immersed in Dulbecco’s Modified Eagle’s Medium (DMEM) (Sigma, UK) supplemented with 10% fetal bovine serum (FBS) (Thermofisher, UK) at pH 7 and incubated at 37°C to reach an equilibrium swelling state. The DMEM was replaced every two days. Over the course of 10 days, samples were removed from the DMEM and the swollen hydrogel sample weights (Ws) were measured. The swelling ratio (Q) was calculated using Eq (1):

$$Q=Ws/Wi \quad (1)$$

Where Q is swelling ratio, Wi is the dry weight and Ws is the swollen hydrogel sample weight. After weighing the swollen hydrogels, the samples were lyophilised and weighed (Wd). The percentage of mass loss was calculated using Eq (2):

$$\text{Mass loss }(\%)=(Wi-Wd)/Wi\times100 \quad (2)$$

(n=3)

2.4 Mechanical assessment

The mechanical properties of the hybrid system were assessed. The compression test was performed with a constant strain rate (20%) of the specimen height, compression tests using the Instron 3344 machine (Instron, UK) equipped with a 10 N load cell. Crosslinked (Alg-Gel)MA hydrogel discs were prepared as described in section 2.1 and maintained in DPBS at 37ºC. After 24 h of incubation, cell-free constructs were measured at equilibrium swelling on day 1 to determine both the diameter and thickness, and unconfined compression tests were performed on the hydrogel discs at room temperature and wet conditions (n=3 for each specimen).

2.5 Cell viability

The cytotoxicity of the hydrogels was observed through the encapsulation of primary human chondrocyte cells (Cell Applications, USA) and a live/dead assay kit (Thermofisher Scientific, UK). Cells were cultured till passage four in a complete human chondrocyte growth medium (Cell Applications, USA). Cells were harvested and encapsulated at density of 1×10^6 cells/mL in the hydrogel solution. 100 µL of the gel/cell solution was pipetted into a 48-well plate and exposed to UV radiation to cause gelation and cell culture media was added. After 3 days the cytotoxicity was observed by removing cell culture media, washing gently in DPBS adding a 2 µM calcein-AM and 4 µM ethidium homodimer-1 solution, and incubating for 25 min. The hydrogels were then observed using a confocal fluorescence microscope (TCS-SP5, Leica, Germany).

3 RESULTS AND DISCUSSION

3.1 Functionalisation confirmation and rheological characterisation

The functionalisation of the hydrogel has been confirmed through the investigation of the internal structure of the crosslinked (Alg-Gel)MA system. The reaction was performed under basic conditions, allowing the introduction of photoreactive methacrylate into the polymer backbone, as confirmed by NMR analysis. The reaction between both polymers have resulted in the appearance of new characteristic peaks of MA at 5.63 ppm and 6.09 ppm attributed to the methylene group in the vinyl bond, and a peak at 1.82 ppm assigned to the methyl group, which are not present in the non-modified polymer (Fig. 2).

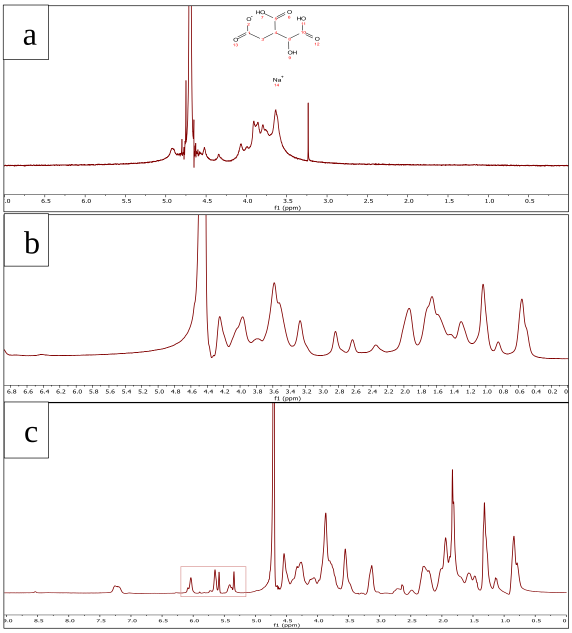


Figure 2- Representative 1H NMR spectra of a) alginate, (b) gelatin, and (c) hybrid system (Alg-Gel) methacrylate.

Rheological characterisation was conducted on the hydrogel samples to quantify viscoelastic behaviour which is directly related to the polymer molecular structure and functionalisation process (Fig. 3). Furthermore, monitoring the moduli verses strain curve, would determine the linear viscoelastic region (LVE) where the viscoelastic behaviour observed is independent of stress or strain values. Results show that (Alg-Gel)MA exhibits shear-thinning behaviour which is characterised by decreasing viscosity as a function of increased shear rate. The hydrogel discs perform as a rigid body when low deformation is applied on the polymer, G' and G" are constant, the sample structure is undisturbed. Once the moduli start to decrease, the structure is disturbed, which means the end of the LVE-region is reached. This means that the material behaves as a rigid body at low stresses but flows as a viscous fluid at high stresses.

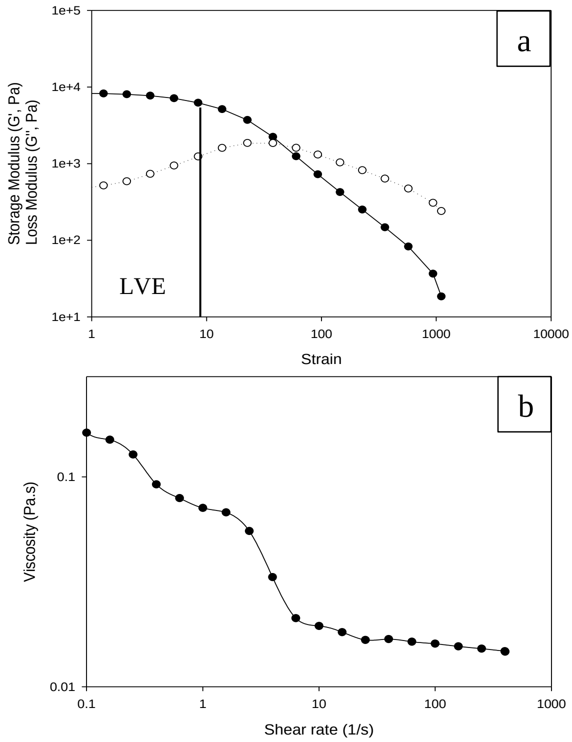


Figure 3- Rheological behaviour of hybrid hydrogel a) storage modulus (G') and loss modulus (G'') versus strain rate; b) viscosity versus shear rate.

3.2 Morphology, swelling and degradation kinetics of the hydrogel

SEM imaging provides visualisation of the hydrogel structure. Figure 4 shows SEM image of the (Alg-Gel)MA cross-linked hydrogel. The hybrid system hydrogels have a cellular structure associated with small pore size and high number of pores. This structure provides a permeable structure for nutrient, gas, and waste product diffusion and allows cell migration throughout the hydrogel.

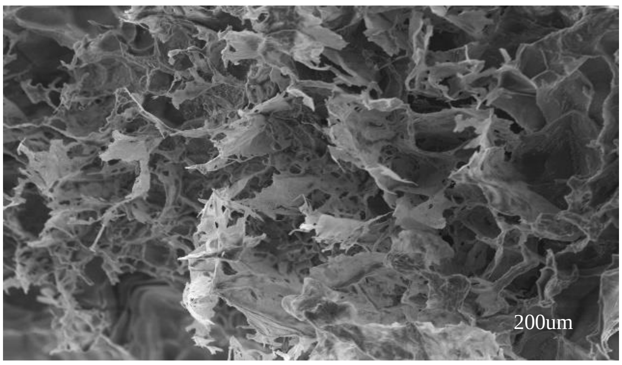


Figure 4 –SEM image of the hybrid hydrogel system which is characterised by small pore size due to high crosslinking density.

The swelling behaviour is an important property of a hydrogel (Figure 5a), when the crosslinked (Alg-Gel)MA begins to swell, the free volume within the sample is eliminated. The hydrogel sample continues to absorb DMEM until it reaches a state of equilibrium. This state is accomplished when the osmotic pressure from the swelling and the elasticity of the hydrogel network is equal. Results showed that the crosslinked hydrogel obtained from samples prepared present a low swelling ratio of 15%. These results show that the degree of crosslinking controls the swelling properties of the hydrogel and alters the mechanical properties. The hybrid hydrogel reached an equilibrium status at day 5.

Hydrogel degradation should match the rate of development and growth of the engineered tissue construct and ECM formation in order to maintain mechanical integrity during cartilage development [18]. The mass loss was approximately 40% by day 18 (Figure 5b), however, the degradation rate can be tailored by increasing the crosslinking density or increasing the polymer solution/photoinitiators concentration or by increasing the exposure time to the UV source.

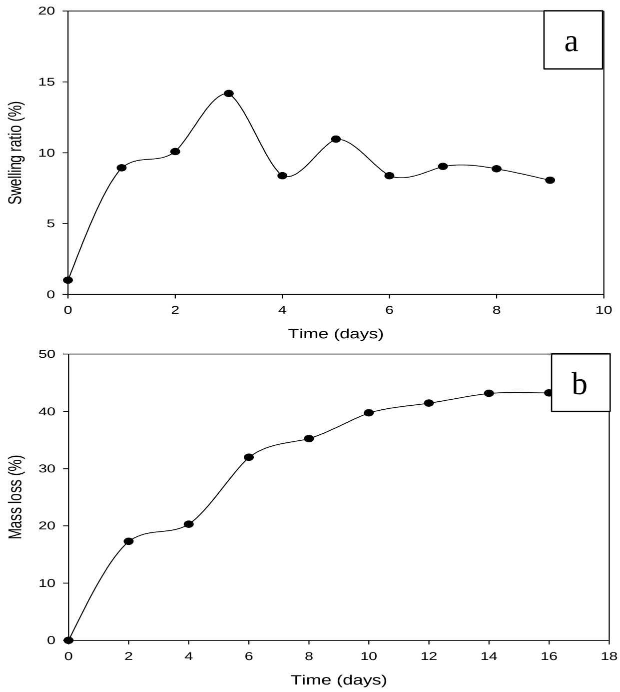


Figure 5- a) Swelling behaviour and b) degradation profile of hybrid hydrogel system

3.3 Mechanical characterisation

The Young's modulus was determined for the hybrid system (Alg-Gel)MA by means of an unconfined compression test (Fig. 6). The hydrogel discs (cell-free constructs) were measured on day 1 after equilibrium swelling. The hybrid hydrogel (Alg-Gel)MA exhibit a suitable mechanical properties with approximate Young's modulus of 35 kPa, within the range of cartilage tissue. However, the Young's modulus varied between the different groups with a range of 10-125 kPa and is affected by polymer concentration, photoinitiator concentration, and UV exposure time.

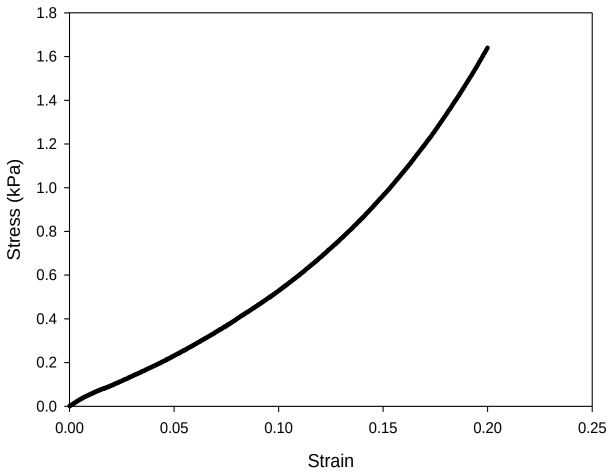


Figure 6- Strain-stress curve obtained from the crosslinked hybrid hydrogel (Alg-Gel)MA.

3.4 Cell viability

Cell viability was observed using a live/dead assay (Fig. 7). Both the hydrogel and tissue culture plastic (TCP) samples demonstrated high cell viability more than 95% with only a few dead cells observed. This demonstrates that the photoinitiator and UV exposure in the hydrogel system had limited cytotoxic effect on the cells.

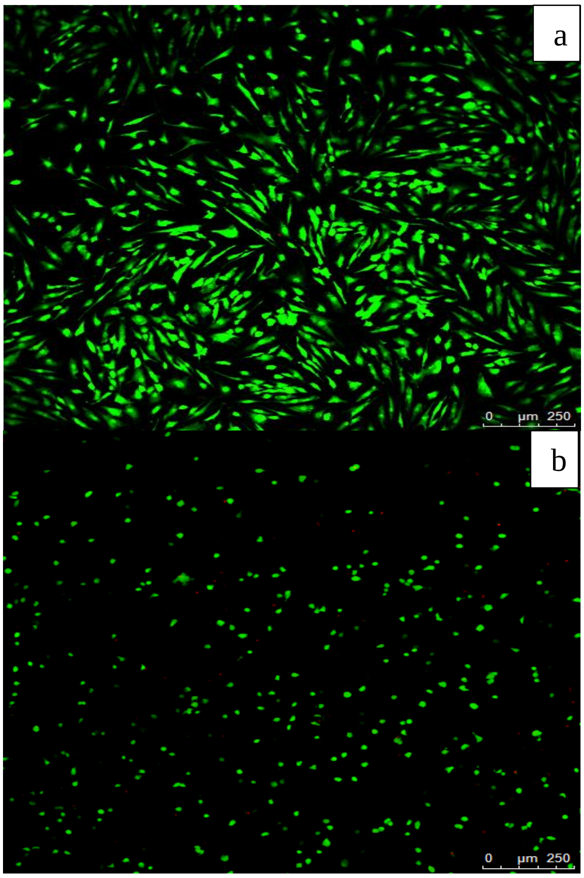


Figure 7- a) TCP and b) 3D cell culture with cells encapsulated in the hybrid hydrogels.

The chondrocytes morphology has been evaluated. Whilst the chondrocytes cultured on 2D TCP have an elongated and fibroblastic morphology, cells encapsulated in the hybrid hydrogel (3D) have a rounded morphology similar to the native morphology of chondrocytes in articular cartilage tissue. This demonstrates the potential of the hydrogel for cartilage tissue engineering applications.

4 CONCLUSIONS

The results of the study confirm that the (Alg-Gel)MA hybrid hydrogel system was functionalised with methacrylate groups and confirmed through NMR. Porosity, pore size, and permeability were assessed by SEM as the hybrid system exhibited a cellular structure. The rheological and mechanical assessment show that the design of a hybrid polymer gel system containing both natural polymers enhance the mechanical properties and allows tailorable mechanics. The rheological analysis demonstrates suitable printability due the shear thinning presented in the hydrogel system. Cell viability confirmed that UV exposure and photoinitiator had no cytotoxic effect on the cells. The network structure of the hydrogel provides cells adhesion sites and the pore structure provides the space for cells spread and migration. The interconnecting pores in the hydrogel also providing the suitable environment of the matrix for cell growth. Chondrocyte cell morphology assessments indicated a spherical shape for the crosslinked hybrid system scaffolds compared to 2D TCP. The development of the hybrid (Alg-Gel)MA hydrogel shows suitability to be used as a bioink for 3D bioprinting of articular cartilage tissue engineering. Further investigation is required to assess the printability and long-term chondrogenic maintenance of chondrocytes.

5 ACKNOWLEDGEMENT

Mr. H.H. Mishbak wishes to acknowledge the support of the Government of Iraq for supporting his PhD through a grant provided by Higher Committee for Development Education Iraq (HCED). The authors wish to acknowledge the funding provided by the Engineering and Physical Sciences Research Council (EPSRC) and Medical Research Council (MRC) Centre for Doctoral Training in Regenerative Medicine (EP/L014904/1).

6 REFERENCES

[1] RQ. Chai, Y. Jiao, and X. Yu, "Hydrogels for Biomedical Applications: Their Characteristics and the Mechanisms behind Them," *Gels*, vol. 3, no. 1. 2017.

[2] A. Cheng et al., "Advances in Porous Scaffold Design for Bone and Cartilage Tissue Engineering and Regeneration," Tissue Eng. Part B Rev., Aug. 2018.

[3] J. P. R. S. Lakes et al., "Tissue engineering," Biomaterials, vol. 8, no. 1, pp. 1–11, Dec. 2015.

[4] L. R. Feksa, E. A. Troian, C. D. Muller, F. Viegas, A. B. Machado, and V. C. Rech, "Chapter 11 - Hydrogels for biomedical applications," A. M. B. T.-N. for the E. of C. Grumezescu Tissues and Organs, Ed. William Andrew Publishing, 2018, pp. 403–438.

[5] A. R. Armiento, M. J. Stoddart, M. Alini, and D. Eglin, "Biomaterials for articular cartilage tissue engineering: Learning from biology," Acta Biomater., vol. 65, pp. 1–20, 2018.

[6] C. J. Tormos and S. V Madihally, "5 - Chitosan for cardiac tissue engineering and regeneration," J. A. Jennings and J. D. B. T.-C. B. B. V. 2 Bumgardner, Eds. Woodhead Publishing, 2017, pp. 115–143.

[7] J. Visser et al., "Reinforcement of hydrogels using three-dimensionally printed microfibres," Nat. Commun., vol. 6, p. 6933, Apr. 2015.

[8] X. Zhang and W. Z. and M. Yang*, "Application of hydrogels in cartilage tissue engineering," Current Stem Cell Research & Therapy, vol. 13, no. 7. pp. 497–516, 2018.

[9] N. Eslahi, M. Abdorahim, and A. Simchi, "Smart Polymeric Hydrogels for Cartilage Tissue Engineering: A Review on the Chemistry and Biological Functions," Biomacromolecules, vol. 17, no. 11, pp. 3441–3463, Nov. 2016.

[10] S. Yan et al., "Injectable In Situ Self-Cross-Linking Hydrogels Based on Poly(l-glutamic acid) and Alginate for Cartilage Tissue Engineering," Biomacromolecules, vol. 15, no. 12, pp. 4495–4508, Dec. 2014.

[11] P. B. Malafaya, G. A. Silva, and R. L. Reis, "Natural–origin polymers as carriers and scaffolds for biomolecules and cell delivery in tissue engineering applications," Adv. Drug Deliv. Rev., vol. 59, no. 4, pp. 207–233, 2007.

[12] P. Jaipan, A. Nguyen, and R. J. Narayan, "Gelatin-based hydrogels for biomedical applications," MRS Commun., vol. 7, no. 3, pp. 416–426, 2017.

[13] V. Guarino, R. Altobelli, F. della Sala, A. Borzacchiello, and L. Ambrosio, "Alginate Processing Routes to Fabricate Bioinspired Platforms for Tissue Engineering and Drug Delivery BT - Alginates and Their Biomedical Applications," B. H. A. Rehm and M. F. Moradali, Eds. Singapore: Springer Singapore, 2018, pp. 101–120.

[14] P. Stagnaro, I. Schizzi, R. Utzeri, E. Marsano, and M. Castellano, "Alginate-polymethacrylate hybrid hydrogels for potential osteochondral tissue regeneration," Carbohydr. Polym., vol. 185, pp. 56–62, 2018.

[15] S. T. Bendtsen, "Alginate Hydrogels for Bone Tissue Regeneration," 2017.

[16] J. Sun and H. Tan, "Alginate-based biomaterials for regenerative medicine applications," Materials (Basel)., vol. 6, no. 4, pp. 1285–1309, 2013.

[17] S. N. Pawar, "Chapter 8 - Chemical Modification of Alginate," J. Venkatesan, S. Anil, and S.-K. B. T.-S. P. Kim, Eds. Elsevier, 2017, pp. 111–155.

[18] V. Dhote, S. Skaalure, U. Akalp, J. Roberts, S. J. Bryant, and F. J. Vernerey, "On the role of hydrogel structure and degradation in controlling the transport of cell-secreted matrix molecules for engineered cartilage," J. Mech. Behav. Biomed. Mater., vol. 19, pp. 61–74, Mar. 2013.

7 BIOGRAPHY

Hussein is a PhD researcher at MACE, holds a BSc degree in Automated manufacturing systems and MSc degree in Advanced Automation systems from Baghdad university in Iraq. He works as a lecturer at the University of Dhi Qar in Iraq.

Cian is a PhD researcher in the EPSRC and MRC Centre for Doctoral Training in Regenerative Medicine and the School of Mechanical, Aerospace and Civil Engineering. He holds a BSc degree in Nanotechnology for the University of Leeds and a MSc degree in Nanotechnology and Regenerative Medicine from University College London (UCL).

Glen studied as a post graduate student in MACE on the Mechanical Engineering Design MSc and then stayed on to complete a PhD in the Design Reliable Systems funded jointly by EPSRC & Volvo Cars.

Paulo Bartolo holds a PhD degree in Polymer Physics from the University of Reading (UK, 2001), a Master of Science in Mechanical Engineering (1996) and a Licenciatura in Mechanical Engineering (1993), both from the Technical University of Lisbon (Portugal).

Influence of Cutting Speed on the Surface Properties In turning of Fe17Cr2Ni0.2C Iron Based Thermally Sprayed Coatings

H. Liborius[1], G. Paczkowski[2], A. Nestler[1], T. Grund[2], T. Mehner[2], A. Schubert[1], T. Lampke[2]

[1]Chemnitz University of Technology, Professorship Micromanufacturing Technology, Chemnitz, Germany

[2]Chemnitz University of Technology, Materials and Surface Engineering Group, Chemnitz, Germany

Abstract

To enhance the wear resistance and friction properties of lightweight materials, such as aluminium alloys, thermally sprayed coatings can be used. A field of application for thermally sprayed coatings are cylinder linings. For this case, finish machining is currently done by honing. A substitution by finish boring would lead to a reduced environmental stress and maybe reduced machining time. For basic investigations, cylindrical specimens are machined by turning with different cutting speeds (100 m/min – 400 m/min). The tools used comprise CBN cutting edges with a rake angle of 0°. Fe17Cr2Ni0.2C coatings are applied by atmospheric plasma spraying on substrates characterised by a helical dovetail microstructure. The analyses of the geometrical properties of the machined surfaces by tactile measurements, SEM, and 3D laser scanning microscopy, show a decrease of the values for Ra, Rz, Rk, and Rpk with increasing cutting speed in the range up to 300 m/min. Furthermore, the valley void volume decreases with raising cutting speed. The absolute values of the residual stresses (detected by XRD) are hardly affected by the cutting speed in the analysed range. However, the orientation of the principal stresses changes. The results enable machining of thermally sprayed coatings using tools with geometrically defined cutting edges. Detailed knowledge of the effects in turning of such coatings can contribute to a substitution of the honing process and hence to realise a material and energy-efficient production.

Keywords

Surface properties, Thermally sprayed coating, Turning

1 INTRODUCTION

The reduction of pollutant emission and fuel consumption are continuous objectives in the automotive industry. Strict legislation for passenger cars concerning exhaust emission and the scarcity of resources are boosters for this trend. The reduction of the weight of combustion engine blocks is a possibility contributing to this aim. Grey cast iron can be substituted by an aluminium alloy as base material. However, these alloys do not exhibit sufficient chemical resistance against fuels with their additives. For this, thermally sprayed coatings are used as cylinder linings. These coatings exhibit a thickness of a few hundred micrometres and a high roughness. They have to be machined to meet the geometrical and surface requirements for the movement of the pistons in the cylinders. Currently, finish machining is done by honing. The abrasive process is chosen to open the microstructure inherent pores of the coating, which act as lubricant reservoirs in service. A substitution by finish boring would lead to a reduced environmental stress due to avoiding the honing sludge and presumably decreased machining time.

For a high adhesive strength between the substrate and the coating, the substrate has to be roughened or microstructured, which influences the properties of the whole compound. Roughening can be realised by different machining processes, such as grit blasting, water jet cutting or laser beam machining. Another possibility is microstructuring by machining processes with geometrically defined cutting edges to generate dovetail microstructures. Bobzin et al. [1] show that these specific microstructures lead to a higher adhesive tensile strength than grit blasted surfaces. This adhesive tensile strength is higher than the 30 MPa required for thermally sprayed coatings used as cylinder lining [2]. For an interlocking between the substrate and the coating, a minimum dovetail microstructure height of about 70 µm is necessary [3]. An increase leads to a higher adhesive tensile strength. Liborius et al. [4] show an increase of the adhesive tensile strength with raising number of structure elements per centimetre. Additionally, the influence of specific substrate microstructures on the morphology of the coating is shown. A higher number of structure elements per centimetre results in a higher oxide proportion and hardness of the coating.

For finish machining of thermally sprayed coatings by tools with defined cutting edges, only few results are published. Cutting processes are mainly used for premachining, where accuracy of the macrogeometry and tool wear are the most important factors. In finish boring experiments of Ding et al. [5], the influence of the cutting material on the tool wear was analysed for machining of iron based thermally sprayed coatings. After a cutting length of 390 m with a cutting speed of 300 m/min, a feed of 0.1 mm and the depth of cut of 0.25 mm, the flank wear was measured. For cermet tools, the flank wear land width was about 2.1 mm and for indexable inserts with diamond tips as well as the coated cemented carbide tools about 1 mm. Additionally, two CBN tools were tested. The CBN with about 50 % boron nitride and a ceramic binder showed a flank wear land width comparable to the diamond tipped tools. A significant reduction of wear was achieved by the use of CBN tools with about 90 % boron nitride and a metallic binder. The measured flank wear land width was about 0.1 mm. Additional experiments using this cutting material for finish boring in the range of cutting speed between 80 m/min and 400 m/min, showed an increased tool wear for machining with lubrication compared to dry cutting. In similar experiments of Ding et al. [6] with a feed of 0.19 mm and a depth of cut of 0.25 mm using the mentioned tools with 90 % boron nitride and a metallic binder, a linear increase of the flank wear land width was measured with a raising cutting length. For the three analysed cutting speeds (85 m/min, 200 m/min and 400 m/min), an increase of the tool wear with increasing cutting speed was shown, whereby the difference in flank wear land width between 85 m/min and 200 m/min was very small. Machining with a cutting speed of 400 m/min resulted in a significantly higher flank wear land width. The fine boring with CBN tools (90 % boron nitride) led to a decrease of the average surface roughness values of 50 % compared to machining with the cemented carbide tools. The differences were explained by the reduced tool wear in the case of using CBN. Analyses of the surface by SEM showed feed marks as well as pores and fractures. The investigations of López et al. [7] concerning turning of Ni5Al thermally sprayed coatings (hardness about 80 HRB, porosity less than 0.5 %) also show that CBN should be preferred as cutting material. The coatings applied by atmospheric plasma spray equipment (Sulzer-Metco) were machined by different cutting speeds (30 m/min – 260 m/min), feeds (0.05 mm – 0.20 mm) and depths of cut (0.05 mm, 0.12 mm). Generally, the surface roughness values after machining were higher than the kinematic roughness. The ratio between the measured surface roughness depth Rz and the kinematic roughness decreased with an increasing cutting speed up to 200 m/min. A further increase of the cutting speed resulted in constant or slightly increased values for the roughness.

Abrasive processes, such as honing open a defined proportion of the microstructure inherent pores acting as lubricant reservoir. Honing experiments for different thermally sprayed coatings [8] showed a slight decrease of the surface roughness values with increasing cutting speed. Additionally, pore opening was analysed. Some pores at the surface were closed after machining as a result of the formation of burr in the abrasive process.

Due to the results, microstructuring of the substrate is important for generating an adequate adhesive strength between the substrate and the coating. Dovetail microstructures ensure this, because of the interlocking between the substrate and the coating material. Regarding the state of the art, there are no investigations concerning the influence of the cutting speed in turning on the surface properties of iron based thermally sprayed coatings. The mechanisms in the cutting process, e. g. for opening of pores have to be investigated. The influence of the cutting speed on the properties of the surface layer is also not analysed yet. With the increased knowledge, a substitution of honing by finish boring becomes possible.

2 EXPERIMENTAL SETUP AND METHODES

2.1 Substrate Microstructuring

Microstructuring by tools with geometrically defined cutting edges results in the generation of surfaces with reproducible and adjustable properties. The geometry of the dovetail microstructures used in the experiments is shown schematically in Figure 1.

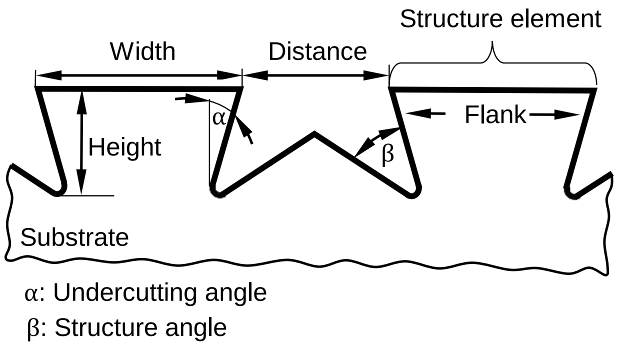


Figure 1 - Geometry of the dovetail microstructure

The microstructures are designed with a height of 70 µm, a width of 93 µm, a distance between the structure elements of 240 µm, a structure angle of 50° and an undercutting angle of 15°. The structure elements are as high as the required height for interlocking, reported by Hoffmeister et al. [3]. The height of 70 µm enables machining of the coating without damaging the substrate. The undercutting angle of 15° provides an interlocking between the substrate and the coating, i. e. it is possible to fill the undercuts with coating material.

Cylindrical specimens consisting of the alloy EN AW-5754 with a length of 50 mm and a diameter of 50 mm were machined by turning. The geometry of the sample is presented in Figure 2. Microstructuring of the substrates as well as the machining experiments (2.3) were carried out on a SPINNER PD 32 precision lathe. The samples were clamped at the centric through-hole with a mandrel

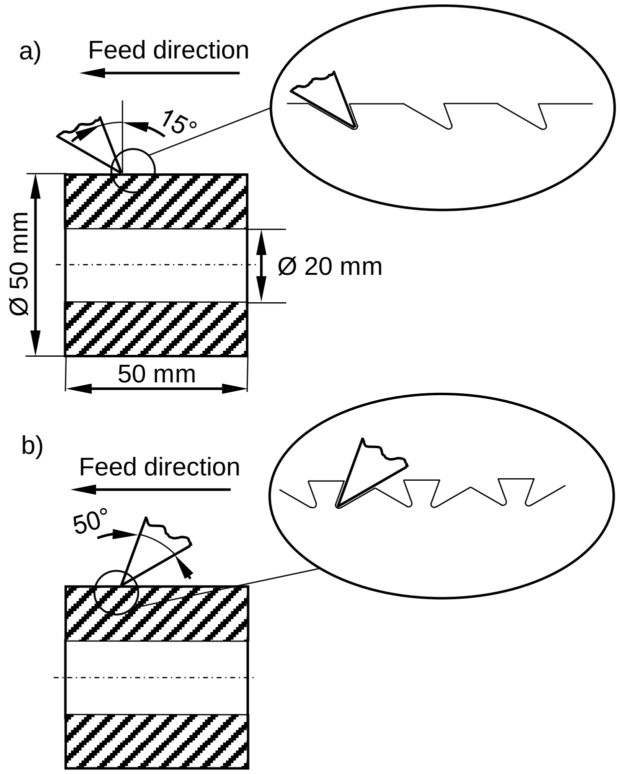


Figure 2 - Microstructuring of the surface: a) first and b) second machining step

For machining the substrates, one tool for cutting the left flank and another tool for cutting the right flank of the dovetail microstructures were used. The tools comprise CVD diamond tips with a corner radius of 10 µm and a corner angle of 50°.

2.2 Coating process

Two methods of thermal spraying were employed for the coating of the specimens. The primer coating (NiAl 80/20) was applied on the microstructured substrates by open-arc wire spraying (System: Perfect Spray – SMS Group), which features dense, low-oxide layers with adhesion by interlocking and diffusion. For the processing of the functional material (FeCrNi (GTV 30.42.2); -90 +45 µm) on the primer in the following coating step, an atmospheric plasma spraying (APS) system of the brand F6 (Company: GTV – Verschleißschutz GmbH) was employed. This one anode cathode system stands out for the high process stability and reproducibility of its coating results. The coating parameters are listed in Table 1.

Current (A)	640
Power (kW)	40
Plasma gas (l/min)	Ar: 60 / H: 7
Powder feeding gas (l/min)	Ar: 3
Line spacing of the robot (mm)	8
Coating distance (mm)	130
Coating angle	75°
Speed of the robot (mm/s)	10
Rotational speed (min^{-1})	120
Number of passages	6

Table 1 - APS parameters for FeCrNi

2.3 Machining experiments

Regarding the roughness of more than 100 µm and the different diameters between 51.1 mm and 51.3 mm of the specimens after the APS process, premachining by several cutting steps was necessary to ensure constant conditions and a low roughness for the machining experiments.

For the investigations regarding the influence of the cutting speed on the surface properties, the depth of cut and the feed were kept constant, both with 0.05 mm. In the experiments, no lubrication was used. The investigations were done for the cutting speeds of 100 m/min, 150 m/min, 200 m/min, 250 m/min, 300 m/min, 350 m/min, and 400 m/min. Indexable inserts of the type CCGW 09T304 (Sumitomo) with CBN tips were used for the turning experiments. The cutting material consists of 90 % to 95 % boron nitride with a grain size of about 1 µm, bounded with a cobalt-based binder. The cutting edges had no chamfer, resulting in a rake angle of 0°. The tool holder used led to a nominal tool cutting edge angle of 95°.

2.4 Analysing of the machined coatings

2.4.1 Geometrical surface properties

The geometrical surface properties were detected by tactile and optical measurement methods. The surface roughness of all samples machined was measured in the feed direction using a stylus instrument Mahr type LD 120. The measuring length was 4 mm. The filtering of the profile was done in accordance to ISO 11562. For a validation of the roughness values, each specimen was measured at three different positions. In addition, the samples were measured with a 3D laser scanning microscope Keyence type VK-9700. The size of the measured area was 2.0 mm (in feed direction) × 0.5 mm (in cutting direction). The objective of the optical measurements was to determine the volume of the pores and of the material pulled out of the surface. For this, the software MountainsMap was used. To measure the vallay void volume as accurately as necessary, first a form filter (polynomial 2nd order) was applied. Afterwards, the periodic kinematic roughness profile was removed from the surface (using "line orientation"). The Abbott curve of the resulting surface is characterised by a small core roughness. Following this, the tangent in this area has a small slope. Therefore, the lower material ratio (Smr2) is determined, which gives an approximated and reproducible value for the proportion of pores and pulled out material. The Smr2 is used as upper boundary for calculating the valley void volume (Vvv). Additionally, the machined surfaces were analysed by SEM using a Zeiss LEO 1455VP microscope.

2.4.2 Residual stresses

The residual stresses in the surface layer were determined after the machining of the coating by X-ray diffraction utilizing the $\sin^2 \psi$ method (diffractometer D8 Discover, Bruker AXS). The measurements were done with a cobalt anode on the {310}/{301} lattice planes of martensite using Young's modulus $E^{\{310\}}$ = 181 GPa and Poisson's ratio $\mathbf{n}^{\{310\}}$ = 0.32. An area with a diameter of about 0.7 mm was irradiated.

3 RESULTS AND DISCUSSION

3.1 Geometrical surface properties

The machined surfaces are characterised by feed marks, cracks, opened pores and areas, where material is pulled out. This material could be splats as well as oxides. A characteristic example is exhibited in Figure 3.

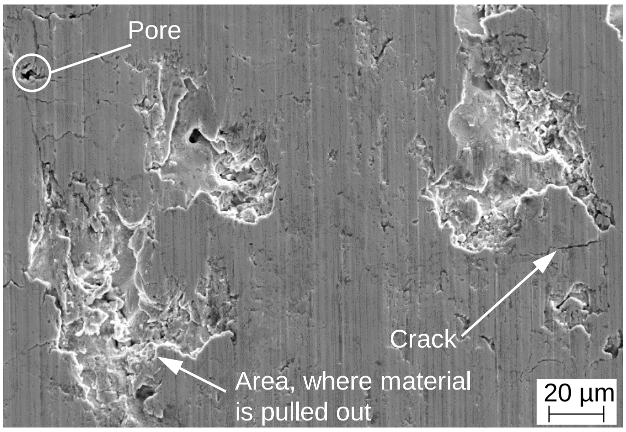


Figure 3 - Machined surface (v_c = 200 m/min) with opened pores, cracks and areas, where the material is pulled out

The differentiation of opened pores and pulled out material is not distinct. Larger connected areas beyond the depth of the kinematic roughness profile exhibit pulled out splats, maybe in addition with oxides and pores. All these irregularities significantly influence the surface roughness values. Considering this, measurement values have to be evaluated critically.

In Figure 4 the influence of the cutting speed on the surface parameters is shown. Independently of the cutting speed, the machined surfaces exhibit a kinematic roughness profile with feed marks in a distance of 0.05 mm. For the evaluation the mean values of the tactile measurements (arithmetic mean surface roughness Ra, core roughness Rk, reduced peak height Rpk, surface roughness depth Rz) are calculated out of three single measurements. The respective minimum and maximum represent the highest and lowest value out of the three measurements. Ra and Rz allow evaluating the cutting process by comparing the calculated kinematic roughness and the measured values. Depending on the cutting speed, the differences indicate changes in the material cutting mechanisms and irregularities of the surface. For the service conditions of cylinder liners consisting of grey cast iron, Flores et al. [9] suggest analysing Rpk, Rk and Rvk. Rk and Rpk influence the tribological behaviour of the surface, which is a common objective in machining of thermally sprayed coatings. In the analysis Rvk is not regarded. The parameter is significantly influenced by pores and the areas of pulled out material, which are analysed using a volume parameter.

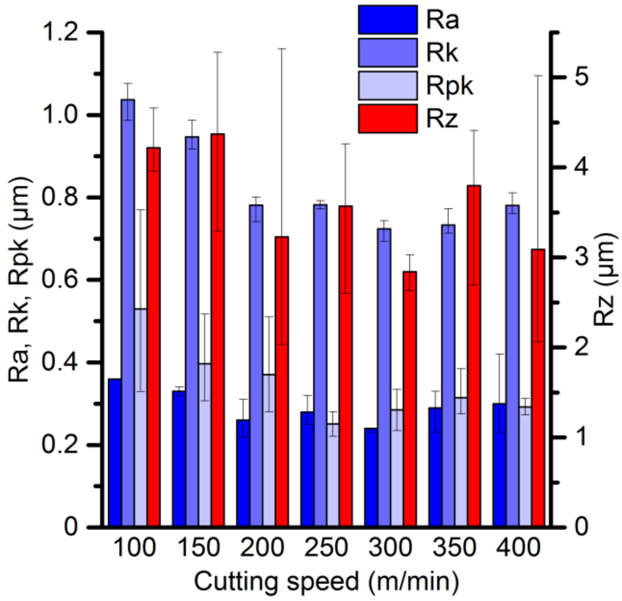


Figure 4 - Influence of the cutting speed on Ra, Rpk, Rk and Rz (f = 0.05 mm, a_p = 0.05 mm)

The surface parameters analysed show nearly the same qualitative trend. In the range of cutting speed between 100 m/min and 200 m/min, the mean values decrease. Higher cutting speeds up to 400 m/min only lead to small changes. Minima of the mean values are reached for cutting speeds of 250 m/min (Rpk) or 300 m/min (Ra, Rk, Rz). Ra exhibit the highest value (0.36 µm) in the considered range after machining with a cutting speed of 100 m/min. The lowest mean value is achieved after turning with a cutting speed of 300 m/min (0.29 µm). The highest value for Rk is 1.04 µm. From that, Rk decreases with increasing cutting speed down to

0.73 µm (300 m/min). However, in the range of cutting speed between 200 m/min and 400 m/min, the mean values are nearly constant. The measured values for Rpk decrease from 0.53 µm (100 m/min) to 0.25 µm (250 m/min) followed by an increase of this value. Rz is 4.22 µm and 4.37 µm after machining with the respective cutting speeds of 100 m/min and 150 m/min. Higher cutting speeds result in lower mean values. The minimum is reached for 300 m/min. It has to be mentioned that by machining with cutting speeds of 250 m/min and 350 m/min Rz also increases to 3.57 µm and 3.8 µm. All in all, the values for Rz are much higher than the calculated kinematic roughness of 0.78 µm.

The described relations can be explained by different processes. The cutting speed influences the shear angle, the temperature and the friction properties in the shear zone. Otherwise, the layer-wise structure of the coating consisting of particles and oxides with different sizes and shapes influence the cutting mechanisms strongly compared to other materials. Particles and oxides could be deformed, pulled out or get cut in the shear zone. This is influenced by the shear angle, which is affected by the rake angle and the cutting speed. Regarding the constant rake angle, the shear angle is mainly influenced by the cutting speed. Machining with low cutting speeds results in lower temperatures in the shear zone, which impedes the cutting of the material, due to the lower effect of strain softening compared to higher cutting speeds. Additionally, the low shear angle also impedes the cutting of the particles of the coating. This maybe results in more particles, which are deformed. With an increasing cutting speed the shear angle and the temperatures in the shear zone rise. Consequently, the cutting of the particles becomes easier. The slight increase of the roughness values by machining with the highest cutting speeds applied might be a result of the increasing tool wear, which changes the rake angle in the shear zone. Subsequently, the cutting speed dependent differences of the roughness values are a result of these mechanisms. However, large areas of cut particles alter the roughness significantly. In Figure 5, the influence of the cutting speed on the valley void volume (Vvv) of the surface is shown. This volume represents the sum of the volume of opened inherent pores and pulled out material. The large areas of pulled out material affect the Vvv significantly. The diagram shows a decrease of the valley void volume with increasing cutting speed in the range of cutting speed up to 300 m/min. After machining with a cutting speed of 100 m/min the Vvv is 0.185 ml/m^2. The lowest valley void volume is reached after machining with cutting speeds of 300 m/min and 400 m/min with 0.084 ml/m^2 and 0.083 ml/m^2, respectively. The mean value of the specimen machined with a cutting speed of 350 m/min shows an increased value (0.154 ml/m^2), which maybe can be explained by unstable cutting condition such as higher tool wear or dissenting coating properties for the machining of this specimen. The trend correlates with the measured roughness values. The higher mean values at lower cuttings speeds occur due the larger areas of pulled out material. Because of the low temperature and shear angle in the shear zone, larger connected parts of the coating are pulled out. The qualitative differences are shown in Figure 6.

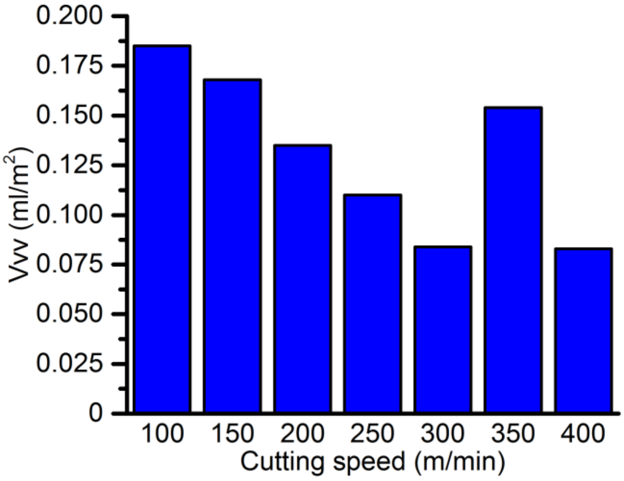


Figure 5 - Influence of the cutting speed on Vvv (f = 0.05 mm, a_p = 0.05 mm)

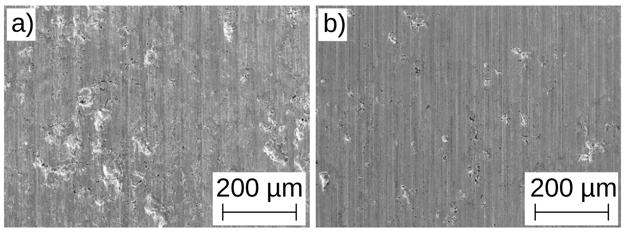


Figure 6 - Surfaces detected by SEM after machining with a cutting speed of a) 100 m/min and b) 300 m/min

Regarding the results for machining of thermally sprayed coatings, cutting speeds between 200 m/min and 400 m/min are beneficial. There are less pulled out particles at the surface.

3.2 Properties of the surface layer

The residual stress state of machined thermally spayed coatings results from a complex interaction of various mechanisms. To be able to identify the orientation of the measured residual stresses, it is insufficient to analyse the size and the orientation of the force components because this would disregard the influence additional factors such as the temperature. The microstructure of the thermally sprayed coatings, characterised by a small grain size, is a result of the plastic deformation in the flow zone (individual particle boundaries) due to the cutting forces, resulting in a micro-grain formation inside the surface area of the workpiece. The conducted XRD analyses confirm the lack of a crystallographic texture.

The residual stresses were also determined at the non-machined surfaces. The stress analysis was conducted on the austenitic Ni/Fe and the martensitic Ni/Cr/Fe solid-solution phases. It is remarkable that the iron alloy is almost free of residual stresses in the surface area after the coating process. Independently of the cutting speed, an increase in compressive residual stresses was visible after machining. The amount of residual stresses of the coating increases from 6 MPa (as sprayed) to 120 MPa compressive residual stress on average after the machining process. Only the angular orientation of the principal stresses varies significantly with increasing cutting speed. The orientation of the angles is shown in Figure 7 and the measured angles of σ_1 to the x-axis depending on the cutting speed are exhibited in Table 2.

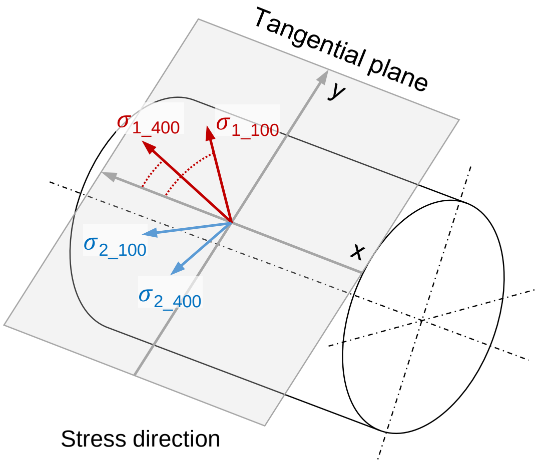


Figure 7 - Directions of the principal stresses σ_1 and σ_2 for different cutting speeds

Cutting speed (m/min)	Angle σ_1 to x-axis
100	47°
150	42°
200	40°
250	37°
300	31°
350	26°
400	16°

Table 2 - Angle σ_1 to the x-axis depending on the cutting speed

The angle decreases with rising cutting speed. One possible explanation lies in a greater thermal relaxation of the machined layer with increasing cutting speed. The influence of plastic deformation should also be regarded. This deformation results in higher degree crystallite deformation.

4 SUMMARY AND CONCLUSIONS

The results of the experiments show that the specific dovetail microstructure of the substrate generates an adequate coating adhesion and combined form fit between substrate and thermally sprayed coating for the subsequent turning process. Independently of the cutting speed, the specimens machined with constant feed of 0.05 mm and depth of cut of 0.05 mm exhibited a kinematic roughness surface profile, but however, characterised by pores, cracks and pulled out material. The cutting speed influences the surface parameters Ra, Rz, Rk, Rpk. In the range of cutting speed between 100 m/min and 300 m/min (250 m/min for Rpk), the mean values decrease with increasing cutting speed. Although, the surface parameters show only small changes in the range of cutting speed between 200 m/min and 400 m/min. The Vvv decrease with increasing cutting speed. This is also a result of the smaller amount of pulled out particles with raising cutting speed. The orientations of the first principal stresses successively align to the cutting direction with increasing cutting speed, whereas their absolute value is hardly affected.

Further investigations using electron backscatter diffraction are necessary to analyse the influence of the cutting speed on the grain size and orientation as well as the grain size distribution. Additional experiments analysing the effects of the feed, the depth of cut and the rake angle are essential for an enhanced understanding of the mechanisms in cutting of thermally sprayed coatings. Furthermore, tribological tests show the influence of different surface properties on friction and wear.

The relations between the cutting speed and the surface properties detected enhance the understanding of machining of thermally sprayed coatings. This represents a first step for the substitution of honing by processes with geometrically defined cutting edges for finish machining of thermally sprayed coatings.

5 ACKNOWLEDGEMENT

The authors acknowledge the DFG (German Research Foundation) for supporting this work carried out within the project “Functional surface design by complementarily matched thermal spray and cutting processes” (SCHU 1484/19-1, LA 1274/38-1).

6 REFERENCES

[1] Bobzin, K. et al., 2008, Thermal spraying of cylinder bores with the Plasma Transferred Wire Arc process, Surface & Coatings Technology, 202: 4438-4443

[2] Barbezat, G., Schmid, J., 2001, Plasmabeschichtung von Zylinderkurbelgehäusen und ihre Bearbeitung durch Honen, MTZ Motortechnische Zeitschrift, 62: 314-320

[3] Hoffmeister, H. W., Schell, C., 2008, Mechanical roughing of cylinder bores in light metal crankcases, Production Engineering, 2: 365-370

[4] Liborius, H. et al., 2018, Influence of dovetail microstructures on adhesive tensile strength and layer morphology of thermally sprayed metal coatings, Procedia CIRP, 71: 299-304

[5] Ding, K., Sasahara, H., 2012, Study on the Machining of Iron-Based Thermal Spray Coating for Sleeveless Engine Cylinder, Advanced Materials Research, 472-475: 991-996

[6] Ding, K. et al., 2010, Investigation on the Cutting Process of Plasma Sprayed Iron Base Alloys, Key Engineering Materials, 447-448: 821-825

[7] López de Lacalle, L. N. et al., 2001, Turning of thick Thermal Spray Coatings, Journal of Thermal Spray Technology, 10: 249-254

[8] Hoffmeister, H.-W., Grosse, T., Gerdes, A., 2012, Investigation of the influence of different process setting parameters on the surface formation at honing of thermally sprayed layers, Procedia CIRP, 1: 371-376

[9] Flores, G., Abeln, T., Klink, U., 2007, Funktionsgerechte Endbearbeitung von Zylinderbohrungen aus Gusseisen, MTZ, 03: 181-185

7 BIOGRAPHY

Hendrik Liborius holds M. Sc. from Chemnitz University of Technology. He is working on his PhD at the Professorship Micromanufacturing Technology. His research area is the microstructuring of substrates for the application of thermally sprayed coatings and the mechanisms in machining of these coatings.

Gerd Paczkowski works at the Institute of Materials Science and Engineering (IWW) of Chemnitz University of Technology. His scientific areas include the process development of thermal spray systems, computational fluid dynamics and the development of thermal spray feedstock materials.

Andreas Nestler obtained his PhD degree in Mechanical Engineering from Chemnitz University of Technology. He is leader of the Teaching and Research Group Machining and Force-controlled Processes. His main research lies in machining of aluminium matrix composites.

Thomas Grund is Chief Executive Senior Engineer at the Materials and Surface Engineering Group at Chemnitz University of Technology. He works on the fields of soldering and brazing technology as well as thermal spraying.

Thomas Mehner obtained his PhD degree in Mechanical Engineering at Chemnitz University of Technology. His main fields of research include X-ray diffraction as well as the relations between corrosion and microstructure.

Andreas Schubert obtained his PhD degree in Mechanical Engineering from Technische Universität Dresden. In 2003 he was appointed Professor for Micromanufacturing Technology at Chemnitz University of Technology. Furthermore, he is head of the Competence Center Micromanufacturing and Surface Technologies - KoMOT at Fraunhofer IWU in Chemnitz.

Thomas Lampke is the current holder of the chair of the Materials and Surface Engineering Group of Chemnitz University of Technology. Together with his staff members, he does research in the fields of chemical coating and electroplating, thermal spraying, electrolytic and plasma-electrolytic anodising, interface engineering in material joints and composites, corrosions and wear protection as well as materials modelling and simulation.

Finishing of Gear Wheels Using Abrasive Brushes

E. Uhlmann[1], J. Thalau[1]
[1] Institute for Machine Tools and Factory Management, Technische Universität Berlin, Germany

Abstract
An improved surface finish of tooth flanks can enhance the lifetime, performance and acoustic behaviour of gearboxes. High requirements regarding the gear geometry pose a challenge for surface finishing applications. The high geometrical accuracy which can be achieved by grinding processes is not to be negatively impacted by a finishing application. In this context, the finishing of gears using abrasive brushes presents a promising alternative to common finishing applications. The possibility of integrating the finishing process into the grinding machine could lead to the substitution of a further process step outside the grinding machine in the manufacturing chain of gears. Against this background, different brushing tools for the finishing of gears were tested and the technological feasibility of a machine-integrated finishing application using abrasive brushes was investigated. The tools were tested under varying process parameters to obtain relative material ratios of 50 % ≤ Rmr(5 %; -0.4 µm) ≤ 70 % on the surface of the tooth flanks while preserving the micro grooves resulting from the preceding profile grinding process. Furthermore, the geometrical deviations from the target profile ought not to be negatively influenced by the brushing process. A cylindrical brush with filaments profiled according to the tooth gap geometry was identified as the most promising type of brushing tool. By adapting the specifications of the filaments it was possible to specifically influence the quality of the surface finish, as well as the structure of brushing tracks on the finished tooth flanks. Furthermore, with the appropriate choice of process parameters it was possible to superpose the grinding grooves running parallel to the feed direction with chaotically arranged brushing marks.

Keywords
manufacturing; gear wheels; finishing; abrasive brushes

1 INTRODUCTION

An increasing renunciation of fossil and nuclear energy sources, causing a switch to renewable energies, leads to scale effects in the production of wind energy farms on the one hand and to increasing requirements regarding their performance, efficiency and durability. Growing rotor diameters, especially of off-shore plants, lead to decreasing rotational speeds and increasing torques while maintenance, repair and exchange of damaged parts become more and more complex and expensive [1, 2]. These developments affect in particular the gear boxes of large wind energy plants. Thus, the quality of gear wheels becomes essential for an effective change from fossil to renewable energy sources.

After the soft machining and heat treatment, large-module hobs are usually hard machined by discontinuous profile grinding. To reduce the risk for pitting on tooth flanks and increase the tooth bearing strength, gear wheels undergo more and more often a finishing application after grinding to improve the quality of the tooth flank surface. Alongside a longer lifetime, finer surfaces and higher material ratios of the profile Rmr on the tooth flanks also lead to an increasing efficiency of the gear box [3, 4, 5, 6, 7]. Furthermore, specific surface structures on the tooth flanks can reduce the noise emissions in operation [8, 9].

To improve the surface quality of ground tooth flanks, vibratory finishing is a widespread application. In most vibratory finishing applications active chemicals are being added to the compound to create an oxide layer on the surface of the gear wheel in order to facilitate the material removal and accelerate the finishing process [10]. Due to the extensive equipment necessary and the use of chemicals, many manufactures are looking for alternative finishing processes [11, 12, 13]. The use of elastic bonded grinding wheels allows finishing the tooth flanks to comparably surface qualities while no additional equipment is necessary since the process can be performed on conventional grinding machines. While the elastic bond allows high surface finishes, the dynamic behaviour of the tools often leads to challenges regarding the geometrical accuracy of the finished workpiece. The inhomogeneous expansion of the profiled grinding wheel requires complex geometrical compensations which lead to a specific and time-consuming process layout for every new workpiece [14]. Due to their flexible filaments, abrasive brushes adapt to the workpiece geometry and offer a promising alternative for finishing processes of gear wheels. They can, just like elastic bonded grinding wheels be used on conventional grinding machines and have already been used to finish functional surfaces of hardened steel and ceramic parts [15, 16, 17]. In this context, different brushing tools were developed to assess the technological potential of abrasive bushes for finishing high quality gear wheels.

2 ABRASIVE BRUSHES

Abrasive brushes consist of extruded polymer filaments embedded in a resin hub, figure 1. Within these filaments are abrasives, e.g. Silicium Carbide (SiC), that usually have a volume fraction of around 30 %. Depending on the machined material other abrasives like corundum (Al_2O_3) or Diamond can be used [18]. The hub diameter d_h can be designed to fit the applicable machine system. The rigidity of the abrasive filaments can be influenced by their length l_f, the size of the abrasive grains and the filament diameter d_f.

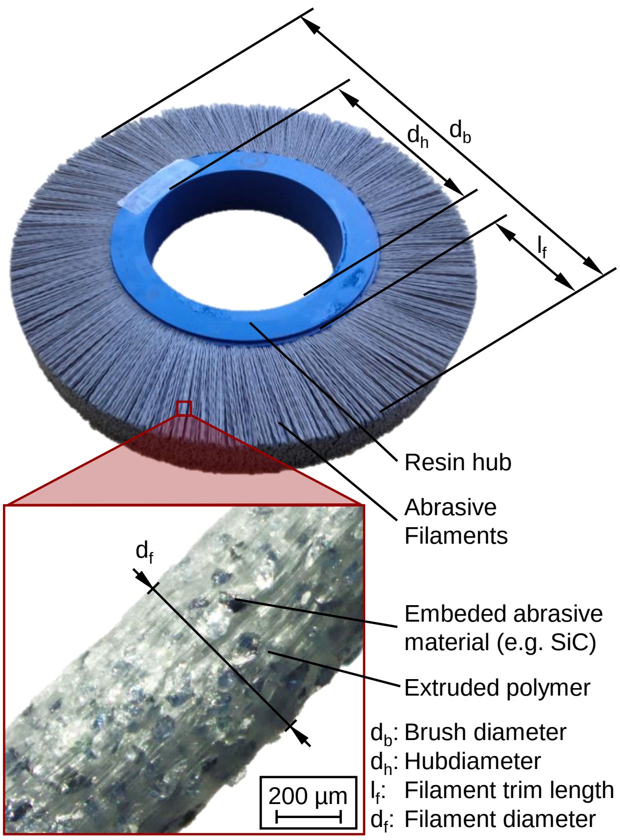


Figure 1 - Abrasive filament brush design

The abrasive filaments require a preload against the workpiece resulting in a bending force exerted by the filaments on the workpiece surface. Depending on the rotational speed v_b and the radial infeed Δx the contact pressure can be adjusted to produce the desired surface finish [19]. The tool depicted in figure 1 has a cylindrical filament stock in which all filaments have the same length. When finishing gear wheels, the shape of the flanks leads to inconsistent preloads of the brushing tool along the profile. Therefore, a profiled brushing tool was developed to match the shape of the tooth gap. To investigate the influence of the tool design on the finishing process, cylindrical as wells as profiled brushes are being examined in experimental investigations. Furthermore, different specifications of filaments were used in experimental investigations to identify their influence on the resulting surface finish as well as on the geometrical deviations of the workpiece. Table 1 provides an overview of the examined tool specifications for this work.

Tool no.		1	2	3	4	5
Abrasive	-	#120	#240	#120	#240	#320
Filament diameter d_f	mm	1.10	0.75	1.10	0.75	0.60
Stock shape	-	A: Cylindrical		B: Profiled		
Filament length l_f	mm	60				
Brush diameter d_b	mm	290				

Table 1 - Specifications of examined brushing tools

3 EXPERIMENTAL SETUP

3.1 Machine setup

The brushing experiments are carried out on the gear profile grinding machine ZE 800 of the KAPP NILES GMBH & CO. KG, Coburg, Germany. To measure the machining forces the grinding machine is equipped with the dynamometer Z21384 supplied by the KISTLER INTRUMENTE AG, Winterthur, Switzerland. The dynamometer is mounted beneath the workpiece clamping. To evaluate the machining forces in radial, axial as well as in tangential direction, figure 2, the angular position of the workpiece as well as the helix angle must be considered.

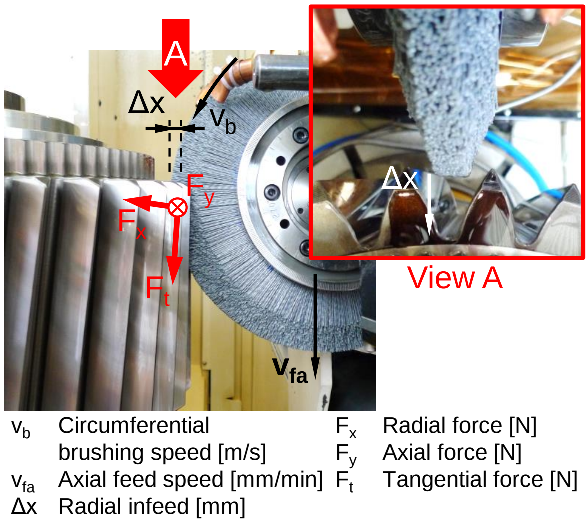


Figure 2 - Experimental setup

A software tool was developed to analyse the machining forces for each individual stroke of the bushing tool through the workpiece in consideration of the workpiece position and geometry.

3.2 Workpieces

Due to the high requirements regarding the shape accuracy as well as the surface finish of gear boxes in wind energy plants, the planetary gears used therein were chosen as test workpieces for this work. These gears have a normal module of $m_n = 10$ mm, $z = 39$ teeth, a helix angle of $\beta = 7.5°$ and consist of the material 18CrNiMo7-6, which is hardened to 61 ± 1.5 HRC.

Prior to the brushing experiments the gears were ground by discontinuous profile grinding using a conventional corundum grinding wheel with a grid size of #80. After grinding, the surface roughness was measured at three points on the tooth flank along the width of the gear wheel. By grinding an average relative material ratio of Rmr(5 %; -0.4 µm) = 27.0 % was achieved at the tip of the teeth while the average relative material ratio at the root of the teeth was a little higher with Rmr(5 %; -0.4 µm) = 36.8 %. Similar results can be observed when considering the arithmetic mean value of the profile ordinate values with Ra = 0.41 µm at the tip of the teeth and Ra = 0.33 µm at the root respectively. These values as well as the workpiece itself represent a typical industrial application.

3.3 Roughness and shape measurement

The roughness of the machined tooth flanks was measured orthogonal to the feed direction at three points along the width of the gear wheel, always at the top, the middle and the root of the tooth using the tactile roughness measuring instrument MarSurf M 300 C by MAHR GMBH, Göttingen, Germany. These measurements were executed on both sides of the tooth gap resulting in a total of 18 measurements for each machined gap. Since it is the goal to only remove the roughness peaks on the tooth flanks by the brushing process while preserving the grinding grooves to store lubrication fluid, the machined surface is characterized by two superposed processes. To quantify the influence of the brushing process on the ground tooth flanks the abbot-firestone-curve was used to identify the material ratio in a depth of 0.4 µm from the top of the roughness profile.

Furthermore, the geometry of the gear wheels was measured prior to the brushing process as well as afterwards using the measuring centre ZMC 550 by CARL ZEISS AG, Oberkochen, Germany. Based on these measurements the profile slope deviation $f_{H\alpha}$ as well as the profile form deviation $f_{f\alpha}$ were evaluated according to ISO 1328-1 [20]. These values were then used to elaborate whether the different brushing tools and processes cause geometrical alterations of the tooth flanks.

3.4 Design of experiments

To assess the performance of the different brushing tools as well as to estimate the potential for a machine-integrated gear finishing process using abrasive brushes, the tools were tested under varying process intensities. To adjust the process intensity, following process parameters were varied: the radial infeed Δx, the circumferential brushing speed v_b, the feed speed v_f and the number of strokes through a tooth gap $n_{Strokes}$. While a variation of the axial feed speed v_{fa} and the number of strokes through the tooth gap $n_{Strokes}$ solely influences the quantity of filament impacts on the flank surface, an alteration of the radial infeed Δx as well as the circumferential brushing speed v_b leads to a change of the impact energy of the single filaments. This is due to the larger filament deformation at higher infeeds or the higher kinematic energy at greater circumferential speeds respectively. To access the performance of the brushing process three different process intensities were tested for each brushing tool, table 2.

Intensity		Min.	Med.	Max.
Radial infeed Δx	mm	1	3	5
Circumferential brushing speed v_b	m/s	10	15	20
Axial feed speed v_{fa}	mm/min	400	300	200
Strokes $n_{Strokes}$	-	2	2	4

Table 2 - Experimental process parameters

4 RESULTS AND DISCUSSION

4.1 Tool shape

A significant challenge for many finishing applications is to improve the surface finish by removing roughness peaks while not removing too much material to change the geometry of the functional surface. To assess whether the different tool designs and process parameters lead to an inhomogeneous work result or deviations to the tooth flank profile, the change of the profile slope deviation $\Delta f_{H\alpha}$ through the brushing process was determined. Figure 3 illustrates how the different brushing tools and processes influence the surface finish on the one hand and the geometrical deviation on the other. It can be observed, that both types of tools can, depending on the process parameters and the filament specifications, improve the surface quality significantly and relative material ratios of Rmr(5 %; -0.4 µm) > 95 % (Ra < 0,08 µm) are achievable. Nevertheless, when using the cylindrical brushing tool, it becomes obvious that for higher surface finishes the geometrical deviation of the tooth flank increases. Due to the negative change of the profile slope deviation $\Delta f_{H\alpha}$, it can be assumed that higher process intensities lead to a tilting of the tooth flanks towards the material side of the flank.

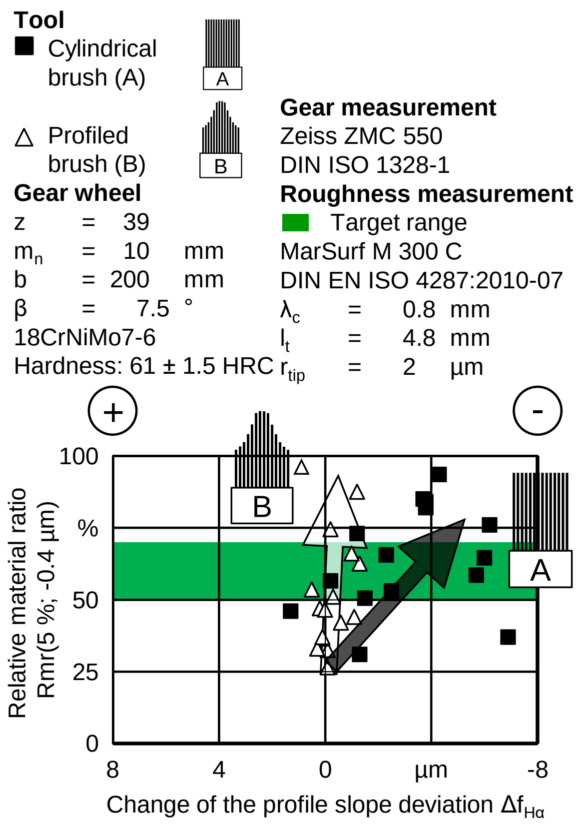


Figure 3 - Relative material ratio and shape deviation of brushed tooth flanks

This is plausible considering the significantly higher preload of the brushing tool at the tip of the tooth as well as the stronger compression of the brush which result in a higher pressure on the workpiece surface exerted by the brushing tool. In comparison, almost no geometrical alterations of the flank profile can be detected using profiled brushing tools. Even for very high relative material ratios of Rmr(5 %; -0.4 µm) > 90 % no significant change of the profile slope deviation $\Delta f_{H\alpha}$ was observed. Accordingly, no geometrical changes were observed for brushed surfaces within the target range of 50 % ≤ Rmr(5 %; -0.4 µm) ≤ 70 %.

4.2 Filament specifications

So far, only the tool shape was considered to assess the process behaviour of a machine-integrated finishing application using abrasive brushes. As indicated in the beginning, the filament specifications as well as the process intensity have a substantial impact on the work result. This is confirmed by the many different surface finishes generated ranging between 26 % < Rmr(5 %; -0.4 µm) < 98 % depending on the filament specifications and the process intensity, figure 3. Since no significant geometrical alterations occurred using the profiled bush, the following examinations will be based on this tool design. The process behaviour as well as the surface finish generated with different process intensities and filament specifications is shown in Figure 4.

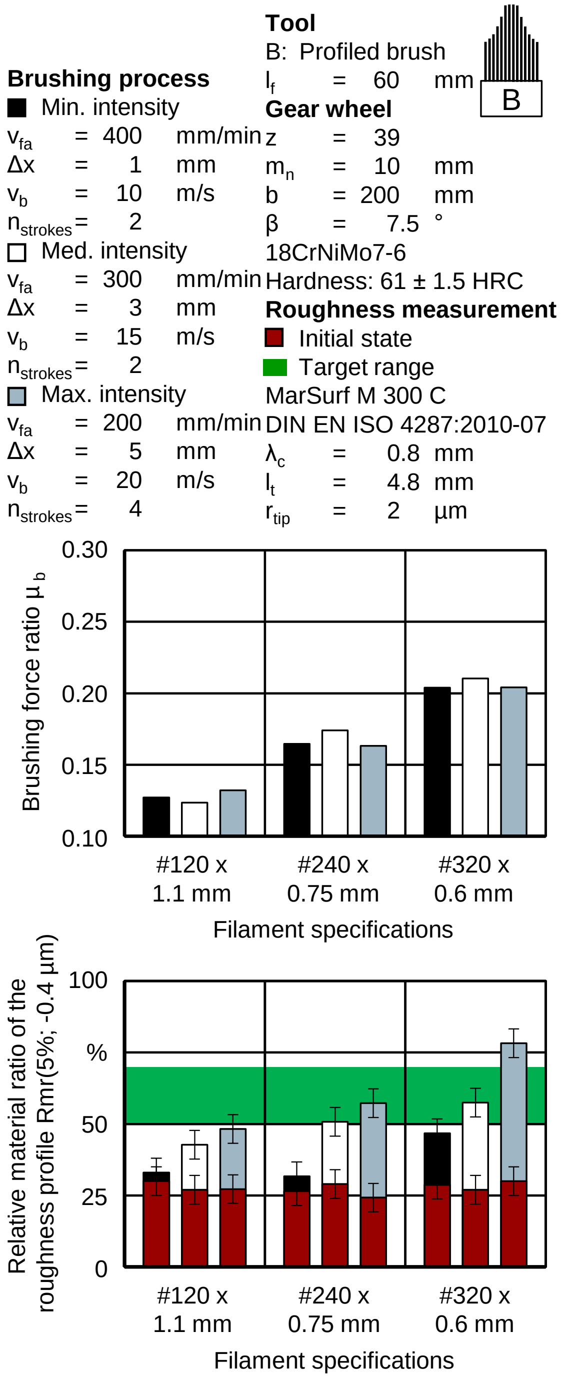


Figure 4 - Influence of the tool specifications of the process behaviour and the surface finish

To evaluate the process behaviour, the brushing force ratio μ_b was calculated from the tangential force F_t and the radial force F_x according to equation 1.

$$\mu_b = \frac{F_t}{F_x} \quad (1)$$

A high brushing force ratio μ_b indicates an efficient machining process, while little values suggest a less efficient material removal. For finer filaments with smaller grid sizes and diameters, higher brushing force ratios μ_b were detected independently of the process parameters, figure 4. This is plausible considering that the smaller grid sizes lead to a significant increase of possible cutting edges and chip formation processes as long as the volume ratio of abrasives within the filaments remains unchanged. Also the smaller diameter of the filaments leads to an increase of filament surface area, also resulting in a higher quantity of active grains. This assumption is supported by the higher surface qualities achieved by finer filaments. Although the different process intensities have almost no influence on the brushing force ratio μ_b, higher intensities considerably impact the surface finish of the brushed tooth flanks. While an average relative material ratio of Rmr(5 %; -0.4 µm) = 41.33 % was reached with a filament diameter of d_f = 1.1 mm and a grid size of #120, the use of filaments with a diameter of d_f = 0.6 mm and a grid size of #320 led to an increase of nearly 50 % with an average relative material ratio of Rmr(5 %; -0.4 µm) = 60.83 %.

With this in mind, similar surface qualities can be reached with all filament specifications by adapting the process parameters. The surfaces generated by the low intensity using the finest filaments (d_f = 0.6 mm, #320), the medium intensity using medium-fine filaments (d_f = 0.75 mm, #240) and the high intensity using coarser filaments (d_f = 1.1 mm, #120) all show relative material ratios of Rmr(5 %; -0.4 µm) ≈ 50 %. Even though the relative material ratios Rmr(5 %; -0.4 µm) are similar, microscopic images of the brushed tooth flanks show significantly different textures on these surfaces, figure 5. All surfaces still show grinding grooves in the feed direction that are superposed by chaotically arranged brushing marks. The width and intensity of the brushing marks depend on the grid size of the abrasives within the filaments. Also, coarser abrasives seem to remove more of the grinding grooves than the finer filament stock.

4.3 Finishing of an entire gear wheel

The previous investigations show that it is possible to use abrasive brushes to significantly improve the surface quality of ground tooth flanks. It was also shown that high surface qualities can be attained without influencing the geometrical accuracy achieved by the previous grinding process. Nevertheless, these results base on single gaps being brushed for each tool and process combination. To estimate the industrial usability of a finishing application using abrasive bushes, entire workpieces need to be finished in order to assess the process stability and wear behaviour of the tools. Based on the previous results, a promising combination of tool specifications and process parameters was selected to finish an entire gear wheel in consideration of the target range for the relative material ratio of 50 % ≤ Rmr(5 %; -0.4 µm) ≤ 70 %. With regard to the more effective material removal process using finer filaments, a brushing tool with a filament diameter of d_f = 0.6 mm and an abrasive grid size of #320 was chosen. With this tool a feed speed of v_{fa} = 200 mm/min, a circumferential brushing speed of v_b = 20 m/s, a radial infeed of Δx = 3 mm and $n_{strokes}$ = 2 for each tooth gap was used. The resulting surface finish throughout an entire gear wheel finish is shown in figure 6.

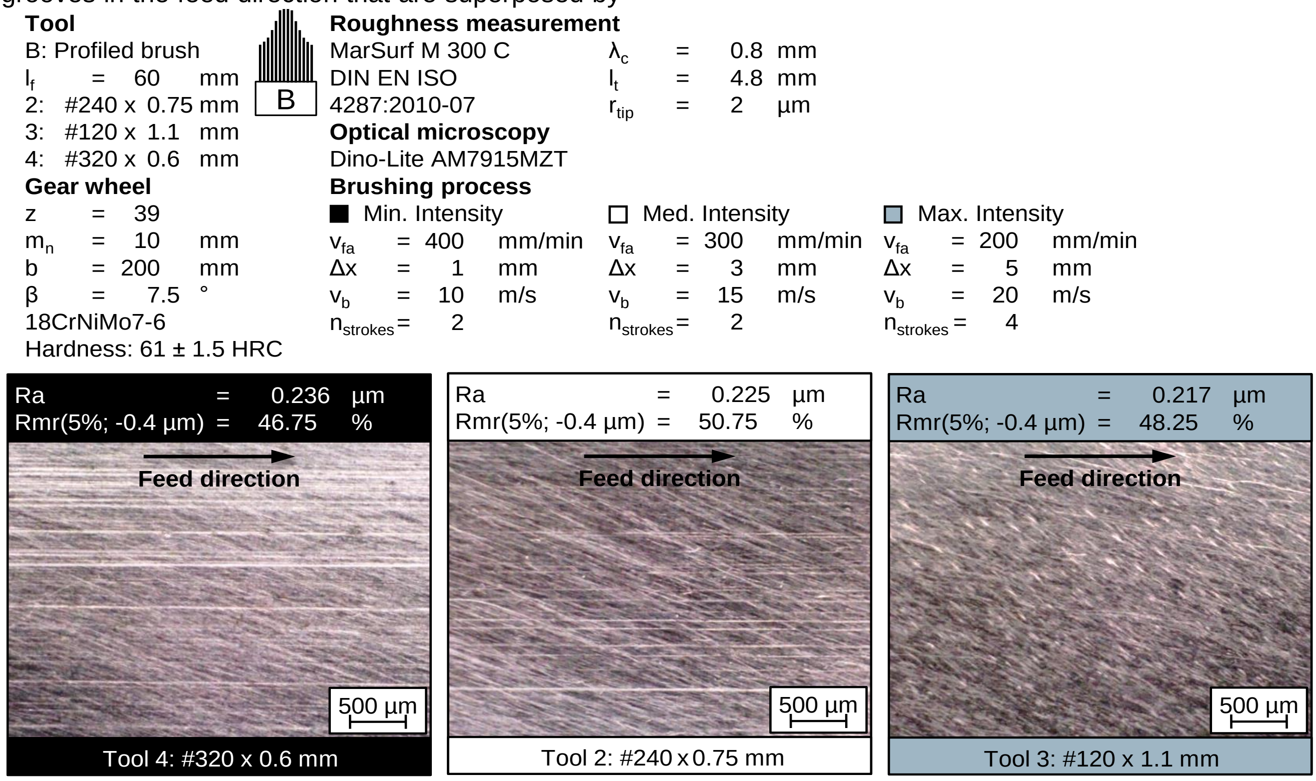


Figure 5 - Influence of the tool specifications of the process behaviour and the surface finish

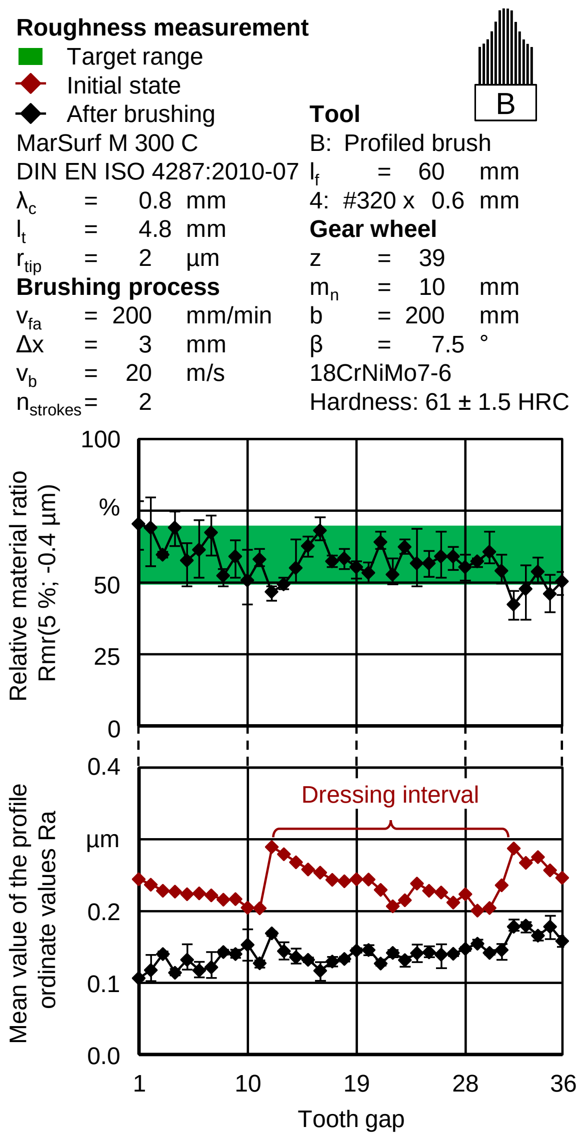


Figure 6 - Finishing of a complete gear wheel

It can be overserved that the target range can be achieved for most tooth gaps. The largest variations of the surface finish occur around the 12th and the 31st gap of the gear wheel. These variations result from the dressing interval of the preceding grinding process. The dressing causes a significantly sharper and coarser grinding wheel topography, which results in higher roughness values and lower relative material ratios Rmr respectively. As the brushing process was designed to preserve most of the grinding grooves the surface finish of the preceding grinding process heavily impacts the final surface quality of the tooth flanks after the brushing process. Nevertheless, the surface quality itself as well as the variation of the surface finish throughout the entire gear wheel was significantly improved by the brushing process.

5 SUMMARY AND CONCLUSION

An enhanced surface quality on the tooth flanks of gear wheels can improve the performance, lifetime and efficiency as well as reduce the operational noise-emissions. The most widely used gear-finishing applications require either extensive additional equipment or a permanent process optimization.

Abrasive brushes show great potential to be used in a machine-integrated finishing application. Depending on the filament specifications and the process parameters, it is possible to generate specific surface qualities, while preserving the grinding grooves resulting from the preceding grinding process. The resulting surface texture lacks roughness peaks, usually damaging the counter wheel under tribological load, yet the grinding groves still remain serving as reservoir for lubrication, and thus assumingly improving the tribological behaviour of the gear wheel.

While the presented work shows the technical potential of abrasive brushes for a gear-finishing application, there are still outstanding issues before the process can be used in an industrial context. In further research, more detailed investigations regarding the process stability as well as the wear behaviour are planned.

6 ACKNOWLEDGEMENT

The authors would like to acknowledge the support and funding for this work by the Forschungsvereinigung Antriebstechnik (FVA) and the contributing companies.

7 REFERENCES

[1] N.N., 2013, Leistungssteigerung der Windkraftanlagen - Größenwachstum von Windenergieanlagen. Bundesverband für Windenergie, www.wind-energie.de.

[2] N.N., 2010: Produzierendes Gewerbe - Produktion im Produzierenden Gewerbe. Statistisches Bundesamt, Fachserie 4, Reihe 3.1.

[3] Kapp, M., 2010, Hartfeinbearbeitete Antriebskomponenten zur Steigerung der Energieeffizienz in Antrieben. Proceedings des XIII. Internationalen Produktionstechnischen Praktikum.

[4] Niskanen, P.; Manesh, A, 2003, Reducing wear with superfinish technology. The AMPTIAC Quarterly, Volume 7, Number 1.

[5] Schwienbacher, S., 2008, Einfluss von Schleifbrand auf die Flankentragfähigkeit einsatzgehärteter. Zahnräder; Dissertation TU München.

[6] Vucetic, D.; Sander, B., 2008, Große Moduln produktiver schleifen. Verfahrensbedingte Prozessgrenzen beim Profilschleifen aushebeln. WB Werkstatt und Betrieb, Vol. 141, Issue 5, 58 - 61.

[7] N.N., 1991, Verzahnungen in Windgetrieben - Optimierung durch Superfinishing. Windkraft Journal & Natürlich Energien - unabhängige

Fachzeitschrift. Seevetal, Verl. Natürliche Energie.

[8] Hohle, A. C., 2002, Auswirkungen von Rauheit, Oberflächenstruktur und Fertigungsabweichung auf das Lauf- und Geräuschverhalten hartfeinbearbeiteter hochüberdeckender Zylinderräder, Rheinisch-Westfälische Technische Hochschule Aachen, Dissertation.

[9] Otmar, A.; Tröbst, S., 2005, Untersuchungen von Zahnradgeräuschen an schweren Nutzfahrzeugdieselmotoren mittels MAN-Geräuschindex. Motor- und Aggregat-Akustik 2, Vol. 51, 181 - 189.

[10] Rösler Oberflächentechnik GmbH, 2008, Gleitschliffanlage für Superfinishing - Die Spitzen glätten. Journal für Oberflächentechnik, 48 ‐ 49.

[11] N.N., 1974, Bundes-Immissionsschutzgesetz (BImschG): Gesetz zum Schutz vor schädlichen Umwelteinwirkungen durch Luftverunreinigungen, Geräusche, Erschütterungen und ähnliche Vorgänge.

[12] N.N., 2011, Additional sound emission provisions. Economic Commission for Europe, Regulation No. 51.

[13] N.N., 2007, Richtlinie 2007/34/EG der Kommission. Amtsblatt der Europäischen Union, 155/49.

[14] Wagner, A., 2016, Feinschleifen von Verzahnungen mit elastischen Schleifscheiben. Forschungsberichte aus der Stiftung Institut für Werkstofftechnik, Vol. 70, Aachen: Shaker Verlag, Universität Bremen, Dissertation.

[15] Rosemann, B., 2013, Fertigungsverfahren Bürstspanen - Potenziale eines hochflexiblen Bearbeitungsverfahren. In: Jahrbuch Schleifen, Honen, Läppen und Polieren, Essen und Ruhr, 376 – 385.

[16] Uhlmann, E. H.; Schmiedel, C., 2013, Markt- und Trendstudie – Nachbearbeitungsverfahren 2012 - Schwerpunkt Strömungsschleifen. Märkte kennen - Trends nutzen.

[17] Uhlmann, E.; Bäcker, C.; Lypovka, P., 2013, Kinematische Modulation beim Zahnflanken-Profilschleifen von Zahnstangen. wt Werkstattstechnik online 103, 875 ‐ 881.

[18] Rentschler, J.; Muckenfuß, G., 2013, Neue Anwendungsmöglichkeiten durch hochtemperaturbeständige Schleiffilamente in der Oberflächenbearbeitung. In: Jahrbuch Schleifen, Honen, Läppen und Polieren, Essen und Ruhr, 387 – 403.

[19] Landenberger, D.; Steinhilper, R.; Rosemann, B., 2007, Verbesserung der Oberflächengüte durch Bürstspanen. VDI-Z, 67 – 69.

[20] DIN ISO 1328-1, Teil 1, 2018, Cylindrical gears - ISO system of flank tolerance classification - Part 1: Definitions and allowable values of deviations relevant to flanks of gear teeth. Berlin, Beuth.

8 BIOGRAPHY

Prof. Dr. h. c. Dr.-Ing. Eckart Uhlmann studied mechanical engineering at the Technische Universität Berlin. From 1986 to 1994 he was research engineer and chief engineer at the Institute for Machine Tools and Factory Management (IWF). In 1997 he became Professor at the Chair of Machine Tools and Manufacturing Technology. Also, he is the director of the Fraunhofer Institute for Production Systems and Design Technology (IPK) in Berlin.

Janis Thalau, M.Sc. studied mechanical engineering at the University of Bremen, the Cape Peninsula University of Technology in Cape Town and the Technische Universität Berlin. He currently works as head of grinding technology at the Institute for Machine Tools and Factory Management (IWF). His doctoral dissertation focusses on the sustainable manufacturing of gear wheels.

Manufacturing of Functional Aluminum Components by Using Sheet Bulk Metal Forming Processes

R. Schulte[1], M. Merklein[1]
[1]Institute of Manufacturing Technology, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

Abstract
Scientific investigations as well as industrial applications have shown that sheet-bulk metal forming is an expedient and efficient approach to manufacture functional components out of steel materials. However, not only high-strength steel but also light metals, such as aluminum alloys, become more important in various fields of application. Especially the automotive industry faces serious ecological and economic challenges. To overcome those challenges, enhanced material efficiency combined with increased load bearing capacity is crucial. Under these conditions, sheet-bulk metal forming processes are applied to aluminum sheets. First of all conventional aluminum sheets are used to manufacture functional components by a combined single-stage deep drawing and upsetting process. In a next step a Tailored Blank will be manufactured by an orbital forming process prior to the forming of the functional components and then applied to improve the material flow control and enhance the die filling. The objective of this paper is to investigate the transferability of the findings for a mild deep drawing steel to an aluminum alloy to provide the scientific groundwork for the application of aluminum alloys in the process presented. Therefore, the process is analyzed in a numerical simulation to investigate the material flow and to identify process limits. Experimental tests are presented to validate the numerical results and to gain a deeper process understanding regarding the microstructure and the strain hardening. Based on these results, innovative approaches are derived to enhance limits of conventional processes using conventional semi-finished products.

Keywords
Sheet-bulk metal forming, process combination, aluminum

1 INTRODUCTION

Lightweight design is an important factor to meet challenges regarding ecologic and economic circumstances. Based on the fields of application, the substitution of materials as well as their combination can lead to a significant reduction of weight.

An increase of efficiency and the avoidance of wasting essential resources is a key factor not only in general, but especially in sheet metal forming [1]. In this context, innovative processes or process combinations were established to overcome limitations of conventional processes. A promising approach to efficiently manufacture functional components is sheet-bulk metal forming (SBMF). Within this process class, bulk forming operations are applied to sheet metal, often in combination with conventional sheet forming operations to enhance the design flexibility. This technology enables the forming of parts with functional elements having a higher thickness than the initial sheet thickness while increasing the strength due to strain hardening by the intended 3-dimensional stress and strain states. [2] Therefore, the main research topic is the control of the material flow [3]. Besides the process, the material plays a key role to meet the requirements for various components. A widespread approach is the substitution of steel by light metals [4]. Aluminum is yet not only used for car bodyworks but is also used to manufacture functional components [5]. However, the reduction of weight is often accompanied by a reduction of strength. Therefore, the specific application has to be taken into account when applying various materials. For a gearwheel in the gearbox steel seems to be suitable, whereas for belt drives the strength of aluminium is sufficient. According to the load, hybrid solutions can also be promising, as it enables a combination of different materials with locally adapted characteristics [6].

2 FORMING CONCEPT AND PART GEOMETRY

In this section the forming concept, the geometry of the functional component and the materials applied will be introduced. Afterwards, the numerical modelling will be presented.

2.1 Forming concept

The process combination used to manufacture functional components out of sheet metal combines a deep drawing and an upsetting operation which proceeds in a single stage on a triple-acting hydraulic press Lasco TZP 400.

Initially, the blank is clamped between drawing punch and upsetting punch by a hydraulic cylinder with the clamping force F_c. The basic cup shape is drawn as the drawing die moves downwards applying the drawing force F_d. Within the same press stroke the upsetting punch displaces the drawing punch until the part is upset between upsetting punch and upsetting plate with the upsetting force F_u. The reduction of the cup height in the upsetting process forces the material to flow radially into the gear cavity. The process kinematic is displayed in Figure 1.

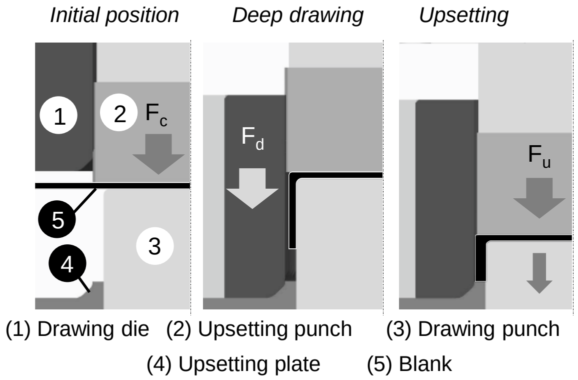


Figure 1 - Kinematic of the combined deep drawing and upsetting process

2.2 Part geometry and materials applied

The functional component manufactured in the process combination is a cup with external gearing as presented in Figure 2a). The gearing has an external diameter of 82.72 mm and consists of 80 teeth with a flank angle of 90° each.

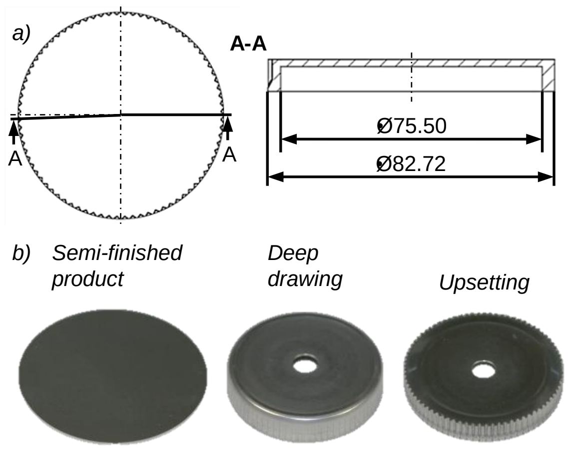


Figure 2 - a) Part geometry and b) Steps from semi-finished product to final component

The inner diameter amounts to 75.50 mm and the height depends on the upsetting force applied. The target height then determines the target volume for tooth and gearing. The target volume is set by the volume of the die cavity and amounts to 108.22 mm^3 for a single tooth and consequently to 8657.6 mm^3 for the whole gearing. The die filling is calculated by relating the volume measured to the target volume. For this purpose the experimental part is digitized with a 3D scanner and then analysed.

As semi-finished product a blank with an initial sheet thickness of $t_0 = 2$ mm is used. The blank has an outer diameter of $d_0 = 100$ mm and a positioning hole with a diameter of 13.2 mm in the center. Figure 2b) shows the semi-finished product, the deep drawn cup and the final shape after upsetting. Previous investigations focused on the forming of different steel materials. In this investigation, the results for the application of the mild deep drawing steel DC04 will be expanded by results for AA5182. The yield strength of the aluminium alloy amounts to 135.10 MPa and is 25% lower than the yield strength of DC04. The micro hardness of 71.6 HV0.05 is 42.5% lower. A comparison of the flow curves and the mechanical properties is presented in Figure 3.

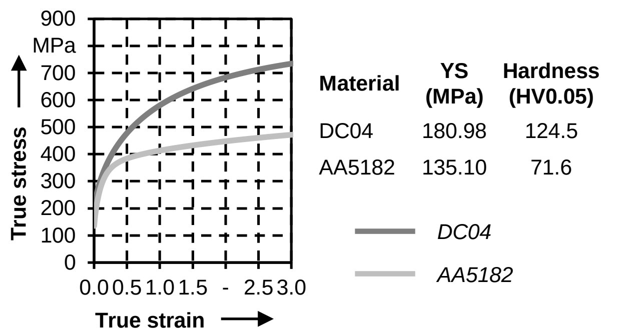


Figure 3 - Mechanical properties of the materials

The data for AA5182 was used to create a material card for the numerical investigation of the forming processes.

2.3 Numerical modelling

For the numerical investigation the implicit FE-code simufact.forming 12 is used. The tools are modelled as rigid bodies according to the tool concept presented above. Due to the cyclical and symmetrical arrangement of the functional elements, a 4.5°-section of the part geometry is used to model the process. To prevent contact problems and to enhance numerical stability a tool segment of 6.5° is used. In contrast to the experimental set-up the numerical process is divided into the two process steps to decrease the calculation time. Furthermore, this enables a specific adaption of the meshing for deep drawing and upsetting. As the deep drawing is a typical sheet metal forming operation, the Sheetmesher is used to generate the mesh for the subsequent forming operation with 6 elements in thickness direction and an edge length of 0.35 mm. For the upsetting process Hexmesher is used with an edge length of 0.2 mm and additional refinement boxes considering the 3-dimensional stress and strain states. Both meshing algorithms create hexagonal elements. The model was validated by an alignment of the numerical results with experimental data. Therefore, a quantitative comparison of the geometric dimensions as well as a qualitative comparison of the strain hardening induced by the experimental process and the effective plastic strain calculated in the numerical model was made with high correlation.

3 COMPARISON OF FORMING RESULTS FOR STEEL AND ALUMINUM ALLOYS IN A SBMF PROCESS COMBINATION

The aluminum alloy AA5182 is applied to manufacture functional components in the combined deep drawing and upsetting process. Results for both process steps are presented and compared to the results for DC04 to analyse their transferability.

3.1 Deep drawing

Due to the lower strength the aluminium alloy also shows a lower force level during deep drawing. Though, the force-stroke curve of AA5182 shows the characteristic progression as displayed in Figure 4. In the beginning of the process, the drawing force increases as the outer radius of the drawing die is in contact with the workpiece. With ongoing stroke the contact zone shifts from the outer to the inner radius of the die and the force steeply increases up to the maximum drawing force of 38.15 kN for AA5182 and 55.09 kN for DC04. After the force decreases with ongoing stroke a further increase can be detected for both materials when the internal gearing of the drawing die gets in contact with the workpiece and causes ironing.

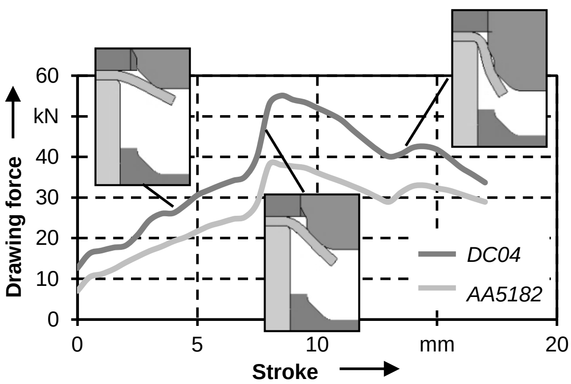


Figure 4 - Force-stroke curves for AA5182 and DC04 during deep drawing

After deep drawing, the analyses of the geometric part properties show similar results for aluminium and steel as well. Figure 5a) presents the resulting sheet thickness after deep drawing calculated by the forming simulation and measured after the experimental test in the area of the drawing punch radius in Figure 5b). The numerical as well as the experimental investigation show the thinning of the material in the area of the drawing punch radius as well as the thickening due to tangential compression at the bottom of the cup wall. The minimum sheet thickness amounts to 1.77 mm for AA5182 and 1.78 mm for DC04. The cup height is 14.75 mm and 14.84 mm, respectively.

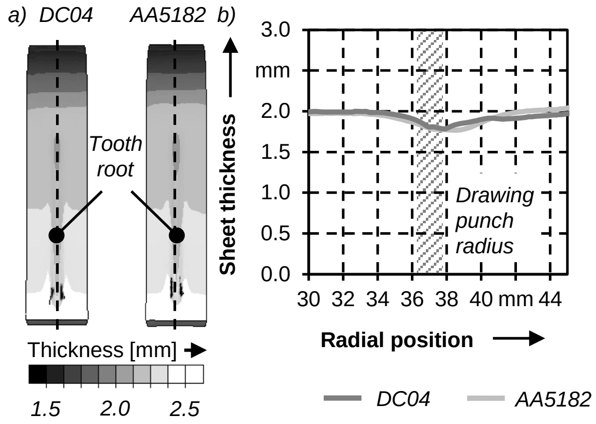


Figure 5 - Geometric properties of the deep drawn cups

The mechanical properties of the drawn cups are evaluated by the resulting strain hardening. The micro hardness of the parts is presented in Figure 6. After deep drawing, there is no significant difference detectable for the hardness in cup wall and bottom due to the slight plasticizing of the material.

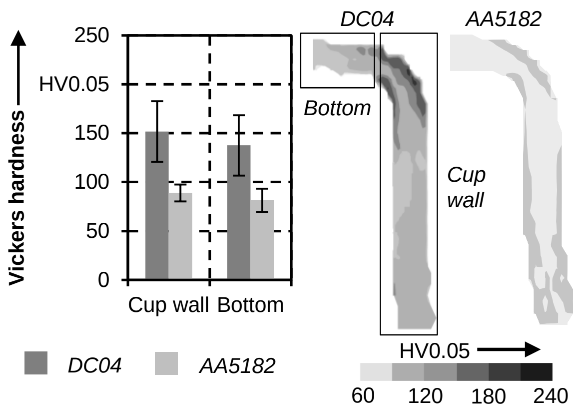


Figure 6 - Mechanical properties of the deep drawn cups

For AA5182, the mean hardness in the cup wall is 88.89 HV0.05, which represents an increase of the initial hardness of over 26%. The mean hardness in the bottom of the cup is 81.44 HV0.05 and 14% higher than the initial hardness. Although there is no significant difference between cup wall and bottom, the mean values correspond to the trend determined for DC04. The mean values for the mild deep drawing steel amount to 151.71 HV0.05 in the wall and 137.62 HV0.05 for the bottom. Disregarding the different level of strength, both of the materials show similar results after deep drawing. The area of the drawing punch radius shows strain hardening due to bending stress at the inner and outer side and the hardness in the lower area of the cup wall increases due to tangential compression.

3.2 Upsetting

After the cup is drawn the upsetting process begins to reduce the height of the cup and to form the functional elements. Therefore, the drawn cup is the input variable for the upsetting process.

The force-stroke curves for the upsetting of the steel and aluminum cups are shown in Figure 7a). Again, the force required for AA5182 is lower than for DC04 due to the lower strength of the material. Furthermore, the gradient is lower due to the minor strain hardening exponent. Both cups are upset to a target height of 11 mm. The process proceeds stroke-controlled. For the aluminum alloy, the maximum upsetting force is $F_u = 168$ kN and is thereby 49% lower than for the mild deep drawing steel. The height for the cups is 10.98 mm for AA5182 and for DC04 as well. The die filling, which represents the material volume per tooth in correlation to the volume in the die, is 77.46% for AA5182 and 78.27% for DC04, as presented in Figure 7b). These values show that the geometric results for the two materials are transferable as the differences of the mean values are in between the standard deviations of ±0.39 and ±0.82.

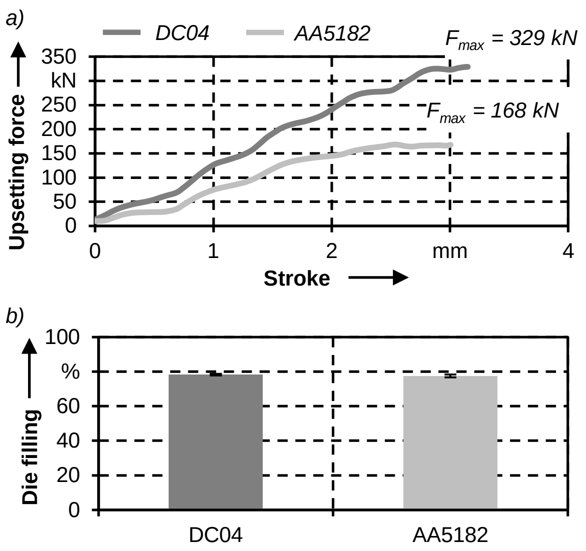


Figure 7 - a) Force-stroke curves for the upsetting of DC04 and AA5182 and b) Resulting die filling for a cup height of 11 mm

The results for the evaluation of the mechanical properties of the final components are displayed in Figure 8. The qualitative comparison shows similar results for steel and aluminum. The material buckles under the upsetting force causing inhomogeneous strain hardening in the cup wall. Quantitative differences occur in accordance with the different level of initial material strength. For AA5182, the hardness in the cup wall increases by more than 43% up to 102.50 HV0.05. The hardness in the bottom amounts to 82.47 HV0.05. Thus, the mean value of the hardness in the bottom wall is about 20% lower than in the wall. The standard deviation is ±12.3 and ±9.11, respectively. For the mild deep drawing steel the mean values for the hardness are 195.58 HV0.05 ±39.12 in the cup wall and 155.73 HV0.05 ±53.56 in the bottom of the cup. Due to more severe strain hardening, the steel shows a higher standard deviation in the wall and in the bottom of the cup as well.

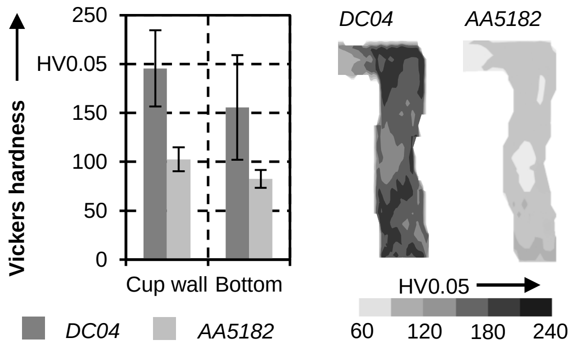


Figure 8 - Mechanical properties of final components

The results for the application of steel in this SBMF process are transferable to the application of AA5182. Nevertheless, the field of application decides whether an aluminium component is suitable or not as the reduction of weight is accompanied by a loss of strength and wear resistance. Therefore, the approach of using Tailored Blanks was identified as expedient [7]. In this investigation, another approach will be presented and analysed.

4 APPLICATION OF HYBRID SEMI-FINISHED PRODUCTS

To meet the challenge of reducing weight on the one hand without losing strength on the other hand, hybrid semi-finished products are applied in the process combination.

4.1 Hybrid semi-finished products

Conventional semi-finished products can be steel or aluminum blanks with an initial sheet thickness of $t_0 = 2$ mm as presented above. Figure 9a) shows the design of the semi-finished products and Figure 9b) presents the arrangement of the materials.

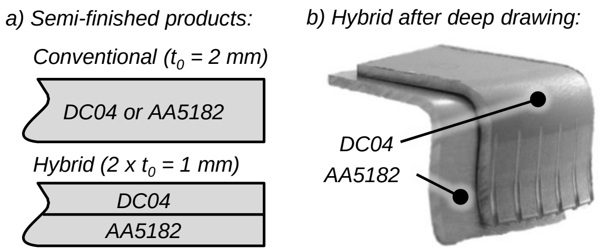


Figure 9 - Design of hybrid semi-finished products and parts

Subsequently, conventional semi-finished products will be replaced by hybrid blanks consisting of one steel and one aluminum blank with an initial sheet thickness of t_0 = 1 mm each. As aluminum alloy AA5182 is used, as steel the mild deep drawing steel DC04 was chosen, as presented in the investigations above. The steel blank with a significant higher strength is positioned on the upper side at the beginning of the forming operations so that the gear teeth are formed out of steel while the base frame is aluminum. The hybrid semi-finished products are subsequently applied to the combined deep drawing and upsetting process. Due to the consistent outer dimensions no adaptions to the processes have to be made.

4.2 Manufacturing of functional components out of hybrid semi-finished products

During the deep drawing process the drawing force required for the hybrid semi-finished product is significantly lower than the force required to draw the cup out of the mild deep drawing steel. For the two blanks with lower sheet thickness the height of the drawn cup is h = 15.3 mm and thereby about 0.5 mm higher than for the cups drawn with conventional semi-finished products.

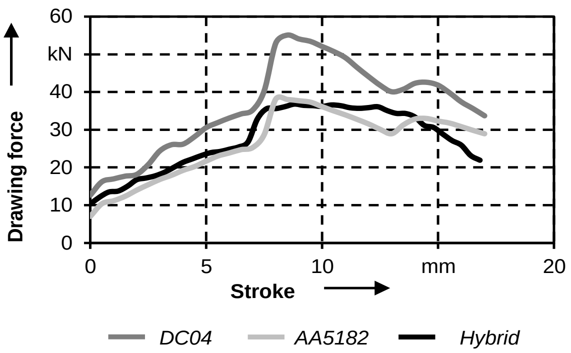


Figure 10 - Force-stroke curves of conventional and hybrid semi-finished products in deep drawing

In contrast to the results presented above, the upsetting process proceeds force-controlled to ensure a sufficient form closure of the blanks due to buckling of the cup wall. Therefore, the maximum upsetting force is set to F_u = 600 kN. The resulting cup height is h = 8.2 mm. Top and bottom of the experimental part are presented in Figure 11 with AA5182 at the inside and DC04 at the outside of the component. The cross-section of the etched sample shows the form-locked connection of the two different materials. Furthermore, the qualitative hardness distribution is displayed in Figure 11 showing the significant increase of the hardness from the inner to the outer side. At the inner side the mean hardness is 136.99 HV0.05 ±16.59 and at the outer side it is 211.35 HV0.05 ±38.75. That amounts to a mean hardness of 175.07 HV0.05 ±49.88 in the cup wall. In the bottom of the cup the mean hardness is 123.63 HV0.05 ±24.90, whereas the values for AA5182 and DC04 are 116.33 HV0.05 ±22.57 and 133.81 HV0.05 ±24.78, respectively.

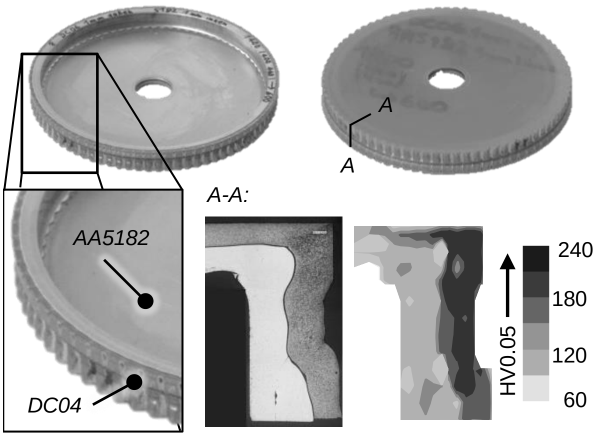


Figure 11 - Geometrical and mechanical properties of hybrid functional components

4.3 Evaluation of weight saving potential for different semi-finished product strategies

The different semi-finished product strategies presented result in substantially varying part weights. For an ideal blank with a constant sheet thickness of t_0 = 2mm the calculated weight is 121.16 g for steel and 41.83 g for aluminium assuming a density of 7.85 g/cm^3 and 2.71 g/cm^3 respectively. For the hybrid semi-finished product the theoretical weight amounts to 81.49 g. After the experiments the parts were cleaned and weighed on a precision scale. The comparison of measured part weights is shown in Figure 12.

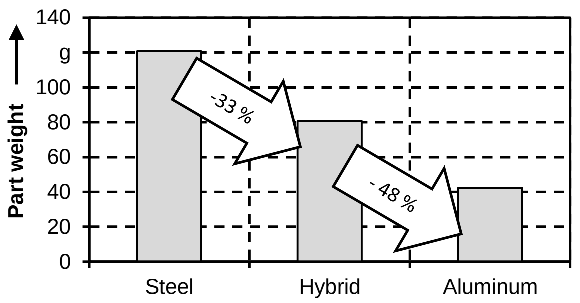


Figure 12 - Weight reduced by hybrid components and light metals compared to the application of steel

For the application of a conventional steel blank the part weight is 120.73 g ±0.77, for an aluminium blank it is 42.39 g ±0.07 and for the hybrid semi-finished product it is 80.82 g ±0.08. If the field of application allows changing the semi-finished product strategy, the weight can be reduced by 33% using hybrid semi-finished products and by further 48% for aluminium blanks compared to steel.

5 CONCLUSIONS

Current challenges regarding forming technology require innovative processes and process combinations as well as adapted semi-finished process strategies. The investigation presented analysed the transferability of gained results for steel onto aluminium and introduced the application of hybrid semi-finished products to SBMF. Of course, mild deep drawing steel offers high strength combined with high formability and is therefore appropriate for many applications. A major disadvantage though is the relatively high density. The application of blanks made out of the aluminium alloy AA5182 showed the transferability of the results obtained for the application of the mild deep drawing steel DC04. Geometrical as well as mechanical properties of the parts investigated show a high qualitative consistency after deep drawing as well as after upsetting of the components. These findings are relevant if the parts manufactured in a forming process can be formed out of aluminium with regard to their field of application. For parts which require higher strength the application of hybrid semi-finished products is a promising approach. Instead of a conventional steel or aluminium blank with an initial sheet thickness of t_0 = 2 mm, a steel and an aluminium blank with a sheet thickness of t_0 = 1 mm each were positioned onto each other and then formed. Thus, the strength of a steel gearing can be combined with the reduced weight of an aluminium base frame within the forming process. Thereby, a form-locked connection is especially valuable for materials with restricted joinability.

The investigations shows the need for further research regarding the application and combination of other materials as well as the identification of process limits for those semi-finished products in SBMF processes, such as the application of precipitation hardenable aluminium alloys in combination with local laser heat treatment. Furthermore, for the manufacturing of Tailored Blanks, which enabled an improvement of geometrical and mechanical part properties for conventional blanks, the transferability of obtained results onto hybrid semi-finished products should be analysed.

6 ACKNOWLEDGEMENTS

This work was supported by the German Research Foundation (DFG) within the scope of the Transregional Collaborative Research Centre on (CRC/TR 73) in the subproject A1.

7 REFERENCES

[1] Cooper, D. R.; Rossie, K. E.; Gutowski, T. G., 2017, The energy requirements and environmental impacts of sheet metal forming: An analysis of five forming processes. Journal of Materials Processing, Volume 244, pp. 116-135.

[2] Merklein, M.; Allwood, J. M.; Behrens, B.-A.; Brosius, A.; Hagenah, H.; Kuzman, K.; Mori, K.; Tekkaya, A.E.; Weckenmann, A., 2012, Bulk Forming of Sheet Metal. Annals of the CIRP, Volume 61/2, pp. 725-745.

[3] Mori, K.; Abe, Y.; Osakada, K.; Hiramatsu, S.; 2011, Plate forging of tailored blanks having local thickening for deep drawing of square cups. Journal of Materials Processing Technology, Volume 211(9), pp. 1569-1574.

[4] Henriksson, F.; Johansen, K., 2016, On Material Substitution in Automotive BIWs-From Steel to Aluminum Body Sides. Procedia CIRP, Volume 50, pp. 683-688

[5] Sieczkarek, P.; Kwiatkowski, L.; Tekkaya, A. E.; Krebs, E.; Kersting, P.; Tillmann, W.; Herper, J., 2013, Innovative tools to improve incremental bulk forming processes. Key Engineering Materials, Volume 554-557, pp. 1490-1497.

[6] Behrens, B.-A.; Chugreev, A.; Matthias, T., 2018, Characterisation of the joining zone of serially arranged hybrid semi-finished components. AIP Conference Proceedings, Volume 1960, article number 040002

[7] Schulte, R.; Hildenbrand, P.; Vogel, M.; Lechner, M.; Merklein, M., 2017, Analysis of fundamental dependencies between manufacturing and processing Tailored Blanks in sheet-bulk metal forming processes. Procedia Engineering, Volume 207, pp. 305-310.

8 BIOGRAPHY

Robert Schulte has obtained his Master in Industrial Engineering from the Friedrich-Alexander University in 2015 and is researcher at the Chair of Manufacturing Technology in the field of sheet-bulk metal forming.

Marion Merklein is ordinaria of the Chair of Manufacturing Technology. She is member of different national and international research associations (e.g. CIRP) and speaker of the Collaborative Research Centre TR73.

Composite and Metallic Forming Manufacturability Criteria for Incorporation into Material Selection Methodology

J. Serfontein[1, 2], D. Dimitrov[2], W. Gerber[1]

[1]Research and Development Department, Aerosud Aviation, Centurion, Pretoria, South Africa

[2] Department of Industrial Engineering, University of Stellenbosch, South Africa

Abstract

Aircraft panel components, manufactured by a forming process, are either comprised of metallic or continuous fibre reinforced thermoplastic (CFRTP) composite material. These materials are traditionally chosen based on the intended functional requirements. Some manual adjustment to the tool is usually needed to achieve the desired shape during the prototype phase. The focus of this study is to utilise forming simulation techniques to improve the traditional process to conceptualise and validate press tools and sheet blanks. In addition, to develop an evaluation model for the material selection process. With the study still ongoing, the literature review and the metallic forming tool and blank conceptualisation workflow conducted in ESI Group Pam-stamp will be presented. The literature reviewed for the study was selected based on three main areas of interest, namely, material selection databases, metallic sheet forming and CFRTP sheet forming. The metallic component will be manufactured using a fluid cell press forming process. The forming tool and blank conceptualisation workflow are comprised of five functions: Computer-aided design (CAD), Material characterisation, forming simulation optimisation, manufacturability through simulation and concept validation. Four of the steps have been completed thus far with the final concept validation to be completed.

Keywords

CFRTP sheet forming, Fluid cell forming, Material selection, Sheet metal forming, Thermoforming

1 INTRODUCTION

Over many decades aircraft manufacturers have significantly increased the amount of composite-based materials used in aircraft structures compared to the use of metallic aluminium and titanium alloy materials [1]. Two significant composite materials currently in use are Glass or Carbon Fibre Reinforced Plastics (FRP/GFRP/CFRP) and Glass or Carbon Fibre Reinforced Thermoplastic (FRTP/GFRTP/CRFTP). A specific difference in an FRP is the thermoset plastic matrix which cures in an exothermic reaction and cannot be reheated and melted as opposed to an FRTP which is a thermoplastic matrix which can harden and repeatedly softened by reheating [1, 2].

The focus of this study is to develop the means to effectively choose the correct material option, based on the component geometry and structural requirements in the most cost-effective way when manufactured by a stamp or fluid cell forming process. Stamp forming is described as a punch-press operation [3]. The forming process is also termed matched mould forming and further detailed as thermoforming [4]. The process is used to manufacture the GFRTP interior lining panel (ILP) using a metallic male press tool and a female silicon tool. Aerosud Aviation is a core supplier to major aircraft manufacturing companies such as Airbus and Boeing and is the design authority on the Airbus A400M. A typical panel is depicted in Figure 1. ILPs are interior components designed to cover and protect substructure beneath it and must sustain various loading conditions.

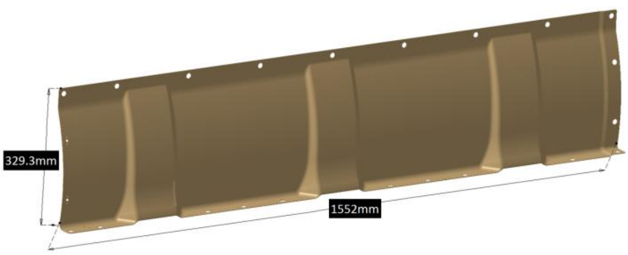


Figure 1 - Schematic of a geometrically complex ILP

The ILP is geometrically complex as multiple curves make up the component, all of which make it difficult for manufacture by forming process. Extensive "trial-and-error" forming tests are required with many GFRTP sheets used to reach an acceptable component finally. The "C" used in CFRTP is also commonly used for continuous and denotes that the thermoplastic composite is comprised of long uninterrupted fibres and not short, discrete fibres [4]. Short fibre composites do not carry high loads compared to long fibre composites due to the matrix carrying the loading as it is the primary load path within the laminate. Fibres are the primary load path in continuous fibre composite laminates; the matrix binds the fibres [4].

Forming an ILP from a single sheet of GFRTP posed extremely challenging. Through the use of thermoforming simulation tools and added blank holding techniques, a conforming single piece ILP that met the curvature around the peaks were as realised, see Figure 2.

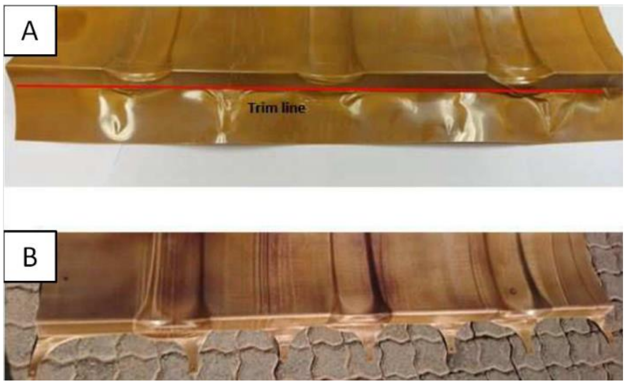


Figure 2 - A - Oversized ILP, B - Optimized ILP [6]

The metallic sheet variant chosen for the study is 2024 series aluminium alloy, 2mm thickness clad sheet. This alloy has excellent fatigue resistance and is the typical choice for aircraft sheet panels [9]. In T3 temper condition, the material has limited formability. However, the material is heated solution treatable which increases formability. The material characterisation of 2mm sheet in the W temper condition which is required to attain accurate forming simulation results was conducted during this study. The design strategy and concepts using a forming simulation were conducted through a case study component detailed in section 2. The final results from the study will be used to conceptualise the forming tools to realise a metallic ILP component.

Material selection is an essential factor in any design. In order to effectively choose the appropriate material, a thorough understanding of the material's property characteristics, component geometry, structural requirements and manufacturability must be considered. The study will focus on establishing an understanding of the formability of thermoplastic and metallic components through simulation and realised component. The understanding of the two manufacturing processes will lay the foundation for developing a material evaluation model.

The material selection method will incorporate two simulation validation studies of the thermoforming and metallic forming process. A study in the simulated formability of the composite ILP has been conducted using Aniform virtual forming software [5, 6]. Further simulations on the composite model will be conducted using the ESI group software packages [7] and the ESI group software packages and Dynaform [8] for the metallic variant. A comparison study of the simulation results will be conducted to establish which is most appropriate in the design process of formed composite and metallic aircraft components. Simulation studies in the design process have shown to be most effective in initial prototyping, hence integrating the right tool into the design process, will improve material selection methods. Costing and strength-to-weight ratio models will also form part of the selection methodology.

1.1 Traditional Material Selection Methods

Material selection studies reviewed the use of simulation techniques in the initial stages of product design. Ismail et al. [10] described several of the material screening methods, material comparison and choosing methods available. The purpose of the study was to review the research work done in the field of material selection. The link between material selection and the inherent manufacturing process most suited was portrayed by many selection systems. The use of computer simulation in the selection of material is described as a means to compare selected material against each other without trial-and-error studies. The selection process that incorporates a computer simulation by Ismail et al. is defined as optimisation methods.

Aguiar et al. [11] conducted an extensive study analysing over 87 digital tools that provided material selection methodology in product design. The research study concludes with no reference to manufacturability of a component. Manufacturing processes are stated as one of the most common criteria for material selection. This entails selecting the material best suited for a manufacturing process, but it by no means gauges the manufacturability of a component.

Both the papers by Ismail et al. [10] and Aguiar et al. [11] refer to the work by Michael F. Ashby, Materials Selection in Mechanical Design [12]. The approach followed emphasised design and material as opposed to material "science" [12]. Ashby's material selection makes use of material and process charts but is limited to the initial screening of materials [10].

Ljungberg [13] conducted a study on material selection and design for structural polymers. Ljungberg describes the concept of Design for Manufacturing (DFM) which integrates the material selection process. A proposed material may be provided as a possible solution, but without practical testing, the most suitable material cannot be chosen. Ljungberg [13] proposed a material be selected first then produce the production tool to avoid costly changes where the material is then changed. However, through the development of manufacturing simulation tools, the above concern becomes a hypothetical view as the tooling validation and verification of the concept is well understood before test prototypes are introduced.

Chakraborty and Chatterjee [14] studied material selection using preferred ranking methods and advises on four material selection methods which are based on multi-criteria decision-making (MCDM). The material finally selected to manufacture a gear clearly shows the applicability, usefulness and accuracy of the four chosen methods. However, only material properties are made use and no manufacturability characteristics.

The current means of material selection is adequate. The evolution of material selection systems with the effective coupling of simulation tools will integrate into design frameworks and is currently underway.

1.2 Composite Forming Process

Forming of some aircraft components is better suited towards thermoplastic sheets for the specific function they exhibit. Sala et al. [15] defined the forming process as a thermoforming process where a polymeric sheet is clamped and heated until softening. The sheet is then placed between a punch and die from which the desired component is formed. Thermoplastics can be reheated as opposed to thermosets which cannot be reheated after curing. The thermoforming process is well studied and documented but the research conducted is based on simple geometry. This study is conducted to establish a material evaluation model. However, the forming of complex geometric components in a thermoplastic substrate must be clearly understood. The use of simulated design and manufacture significantly increases efficiency and reduces the cost of setting up a manufacturing process [16].

Thermoplastic defects that can occur during the manufacturing process is discussed below. Sadighi et al. [19] concluded that $[0°]_4$ lay-ups are more vulnerable to wrinkling whereas $[0°/\pm45°/90°]_T$ lay-ups exhibit buckling. Wang et al. [17] stated that double-curvature shaped components are difficult to manufacture and can lead to defects. Composite component defects are wrinkling, porosities and fibre fracture to state a few. The comprehensive understanding of predicting defects correctly and further eliminating them is the critical aspect of the simulation studies which will aid the component design and material selection.

Drape forming is defined as "a method of forming thermoplastic sheet in which the sheet is clamped into a movable frame, heated, and draped over high points of a male mould" [20]. In thermoforming drapability or formability is the ability to shape a medium before hardening. The effective manufacture of a composite component is defined by the drapability/formability of the component. The study by Boisse et al. [21] draws explicitly on the link between intra-ply shear stiffness and its link to wrinkles and shear strain post the locking angle limit.

Wrinkling is the most common defect to occur in the forming of thermoplastics [22]. A study conducted by Rozant et al. [22] investigated the drapability of dry textile fabrics for stampable thermoplastic pre-forms. The study clearly defines the cause of wrinkling in the component. Figure 3 depicts the occurrence of wrinkling as the yarns move during shearing. The warp and weft yarns of the weave are angled at 90° but when shearing is induced during the forming process, the angle is reduced, and the distance between the yarns get closer together. This angel ($\theta_{Locking}$ or θ_L) is the "locking angle and is the smallest angle reached between two adjacent yarns before the onset of buckling" [22].

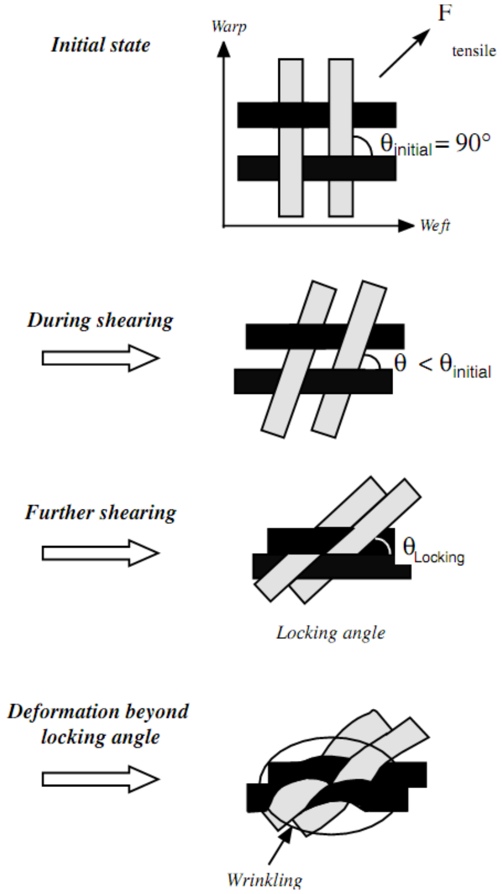


Figure 3 - The occurrence of wrinkling during the shearing of weaves [22]

Figure 3 defines the wrinkling defect, where the movement of the yarns during the forming process causes the yarns to lock. Rozant [22] goes further to describe this movement as the intra-ply shear deformation or in-plane shearing. Harrison et al. [16] conducted a study to characterise the shearing behaviour of viscous woven textile composites using two tests. The picture frame and the bias extension tests are standardised tests that can be used to characterise the material properties for thermoforming. The picture frame test is simple to perform, and results are repeatable. However, the boundary conditions imposed fail to induce the required kinematics where the edges are loosely clamped. On the other end, where the edges are tightly clamped and fibres not aligned correctly, false results are attained.

The uniaxial bias extension test "involves the clamping of a rectangular piece of woven material such that the warp and weft directions of the tows are oriented initially at 45° to the direction of the applied tensile force" Harrison et al. [16]. The test is simple to perform and provides reasonably repeatable results but in particular, provides the best method for determining the locking angle. The test induces a combination of pure shear and intra-ply slip. Hence it is a means to investigate both phenomenon Harrison et al. [16]. The study by Harrison et al. [16] concluded that both test methods are given acceptable results, but mainly the bias extension test is a reliable method for determining a material's locking angle. The biggest drawback is the need to perform an extensive image analysis [16]. The study by Boisse et al. [21] corroborates with the data presented by Harrison et al. [16], however Boisse et al. [21] conducted a forming simulation making use of a dome shape which shows the correlation between double curvature and the need for the material to shear in order to get good formability of a component.

Due to the inherent difficulties of aligning the fibres with the picture frame testing, the study conducted by Chen et al. [18] established the uniaxial bias extension test as the baseline for determining intra-ply and inter-ply shearing properties of GFRTP/CFRTP for thermoforming. The study denotes the effects of temperature and strain rate explicitly. In the case of lower temperatures in comparison to the "forming temperature", the inter-ply shear forces increase significantly as the resin viscosity increases [18]. By holding the test temperature constant and altering the speed of the test, the strain rates between the two tests were established. The study found that a higher strain rate decreases the viscosity of the resin with the lower viscosity in the resin, the inter-ply shear force increases [18]. The Chen et al. [18] study characterised the material properties and conducted simulation studies to correlate the mechanisms. The mathematical model used a thermos-visco-elastic model.

Developing material properties is needed to understand the complexity of thermoforming mechanics effectively. An additional parameter to the two defined above is bending properties. De Bilbao et al. [23] conducted a study on the bending behaviour of reinforcements. "The bending behaviour is specific because the components are structural components and out-of-plane properties cannot be directly deduced from in-plane properties" [23]. The study validated a new test method which is a cantilever type test which determines the bending moment and curvature properties of the material [23]. The study conducted by Wang et al. [17] states the bending stiffness is essential in determining the size of wrinkles.

The investigation by Mashau, M [6], in particular, the research conducted to improve the forming process to yield fewer defect components shows the effectiveness of simulation techniques. The defects discussed previously must be omitted where at all possible during the forming process to manufacture acceptable components. Through the use of simulation, the manufacturing process can be defined and investigated in greater detail. Tool design and rework during process validation will become more efficient with less costly "trial-and-error" methods will be utilised.

1.3 Metallic Forming Process

Metallic forming processes have been used extensively in the aircraft and automotive industry and have also been well studied and understood academically. Solfronk and Sobotka [24] specifically indicated sheet metal forming as one of the most spread technologies for metallic part production and how it enables part production with different shapes as well as size variation. One of the specific drivers is the quality. Advantages of sheet metal forming are the process exhibits good quality surface finish, high accuracy of defined sizes and quite high stiffness at minimal part's weight [24]. From experience, metallic component surface finish in comparison to the composite thermoplastic formed components can be superior as the later yields poor surface quality due to the requirement to use silicon tooling in the thermoforming process. In the case of male matched tooling, a far greater surface finish can be achieved.

With the maturity of forming finite element code to conduct advanced forming simulations, the formability prediction and optimisation algorithms substantially reduce the time and material wasted during the "trial-and-error" period which can be expensive experimental studies [26]. The study conducted by Solfronk and Sobotka [24] concluded that results compared between the simulation results of Pam-stamp [8] and the measurements from the formed components are agreeable. The study conducted further denotes the software enables the user to optimise tooling surfaces, proper parameter selection, the dimensional stability of stamping, compliance of thickness tolerances, the appearance of areas with minimal deformation, and critical zones that will be susceptible to wrinkling and fracture defects. The most significant advantage of the use of simulation is the cost benefits. The study by Yershov et al. [26] highlighted specifically "the ability to significantly conserve material by reducing wastage, shortening the time required for component finishing and lower labour costs". A limitation is the development of localised necking due to strain reaching a critical limit and eventually failing [27].

Similar defects are experienced in composite thermoforming is seen in metallic forming, in particular, wrinkling and fracture. Other defects include tearing, spring back and other geometric and surface defects as denoted by Kadkhodayan and Moayyedian [28]. Wrinkles in thermoformed components are not recoverable and must be dealt with in the pre-forming design phase. Wrinkling can be treated as a recoverable defect, contrary to fracture. “Wrinkling occurs when the compressive stresses go beyond a certain value during forming” [28]. Prevention of defects is essential to ensure dimensional accuracy and acceptable aesthetics [29]. In the case of wrinkling, the clamping and holding tools can “iron” out the wrinkles but only if used correctly. Wrinkling is seen as acceptable if it lies outside the trimmed zone, out-of-plane.

Defects can be prevented with the correct procedure and knowledge of how to use the simulation tools correctly, in particular, the use of Forming Limit Diagrams (FLD). Kumar et al. [25] stated “FLD is used during the design stage of any new sheet metal component for tooling shape & optimising variables. In the sheet metal industry and studies, it is widely used and considered as one of the important tools to determine the formability of sheet metals. Every sheet metal has its forming limit diagram which determines its formability, strain limit and forming regions”. Two things can be taken from this statement: one is that metal forming has a traditional means to determine the formability of a component. The second is that extensive testing is required to establish the Forming Limit Curve (FLC) for the FLD and the established curve is only applicable to a particular material and sheet thickness. A standard test method has been established to determine FLD’s: ASTM E 2218-2 [30] and ISO 12004-2.

The sheet metal forming process selected for this study is fluid cell forming and is a variant of the sheet hydroforming process. The press ram is oil and a flexible rubber diaphragm, which replaces the solid punch. Under the oil pressure, the rubber diaphragm conforms to the shape of the tool and holds the blank over the die cavity [31]. Thus a single rigid die tool is required and can be more cost-effective compared to a typical stamp forming tool and die sets.

The chosen sheet material cannot be pressed in the “as delivered” condition, which is denoted as the T3 temper condition. T3 temper is solution heat treated, cold worked and naturally aged to a suitable stable condition as received from the supplier. Solution treatment is achieved by heating to a temperature of 495°C and holding long enough to allow constituents to enter into solid solution and cooling rapidly enough to hold the constituents in solution [32]. From the heat treatment process specification ABP 3-1119 [33], the AA2024 clad sheet is processed to W temper condition by heating in a furnace for 23m30s±30s at 495±5°C, quenching in water which is no warmer than 40°C for 3m±1m within a delay after exiting the furnace of 12s.

After heat treatment and quenching, the sheet is in W temper condition, which is an unstable temper which naturally ages at room temperature after solution heat treatment. The material naturally ages from W temper condition to T42 temper condition after being cold worked by press forming [32]. The formability of the sheet significantly improves in W temper condition but the time to press after heat solution heat treatment is limited.

2 METALLIC SHEET FORMING CASE STUDY

The component used for the study is an aircraft skin panel that is manufactured by press forming process from 2024-T42 clad sheet material. The material is delivered in T3 temper condition from the supplier. The sheet is heated solution treated to W temper condition after which it is pressed formed and left to naturally ages to the T42 temper condition. Figure 4 depicts the wingtip panel component.

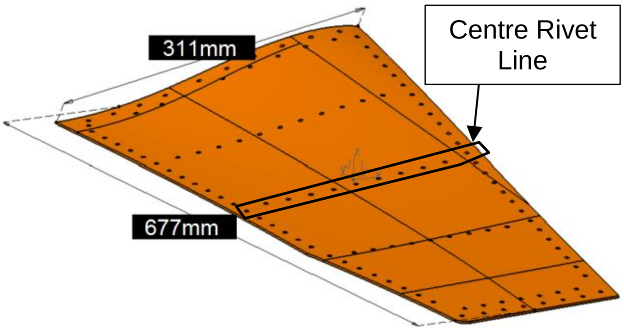


Figure 4 - Schematic of an aircraft skin panel

The component does not meet dimensional requirements using the male matched tooling process and must be manually adjusted to final form. The case study will investigate the manufacture of the component using the fluid cell press. Also, the ESI Group Pam-stamp forming simulation software is used to conceptualise the press tool. Even though the Pam-stamp material database is well populated, there are limited aircraft grade aluminium alloy properties in published literature. Material characterisation of the 2024-W temper condition was conducted to ensure accurate results are attained from the forming simulations.

2.1 Forming Process Conceptualisation Cycle

The forming process conceptualisation cycle is depicted in Figure 5. The process cycle is a means to effectively take the conceptual design of the component blank and tools from a CAD model to final realised component with blanks and tools that can reliably manufacture the component. The key to this process is the inclusion of forming simulation which aides in optimising and verifies the blank and tooling concepts without a costly trial and error process used in traditional design methods. Material characterisation forms part of this cycle but only if the required properties for conducting a simulation are not available. Each phase of the cycle is briefly described:

- CAD

  The CAD function is the creation of 3D models of the sheet blank and tooling. The models are also prepared where necessary and saved into Initial Graphics Exchange Specification (IGS or IGES) file format.
- MATERIAL CHARACTERISATION

  If the material database does not contain the desired material, characterisation will need to be conducted.
- FORMING SIMULATION

  Feasibility and optimisation studies using the blank and tool concepts.
- MANUFACTURABILITY

  Concept acceptance based on the outputs from the forming simulation phase.
- CONCEPT VALIDATION

  Final tooling and blank concepts will be manufactured. Press trials and trimming of the component followed by measurement will qualify the components for serial production use.

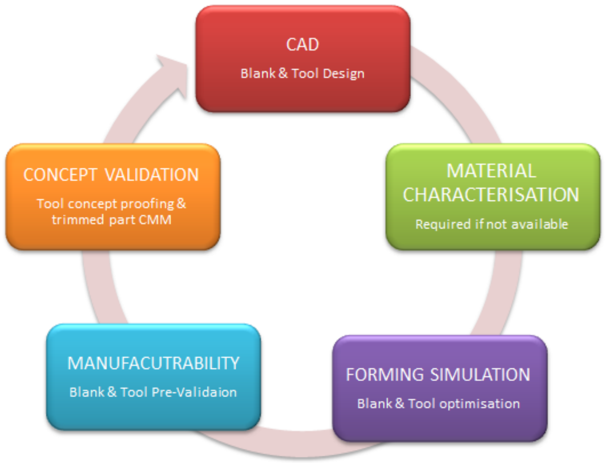


Figure 5 - Forming Process Conceptualisation Cycle

Completing the final conceptualisation cycle will yield an optimal solution. Multiple cycles may be required where complex components are investigated.

2.2 Material Characterisation

Forming simulation tools have been adopted into the automotive industry to a far higher degree than the aerospace industry judging by some materials available in the ESI Group Pam-stamp material property database. The 2024 aluminium alloy clad sheet in 2mm thickness in the W temper condition is not available in the database, and material characterisation testing had to be conducted. In collaboration with ESI Group and the Technical University of Liberec in the Czech Republic, the material characterisation was conducted to attain the following:

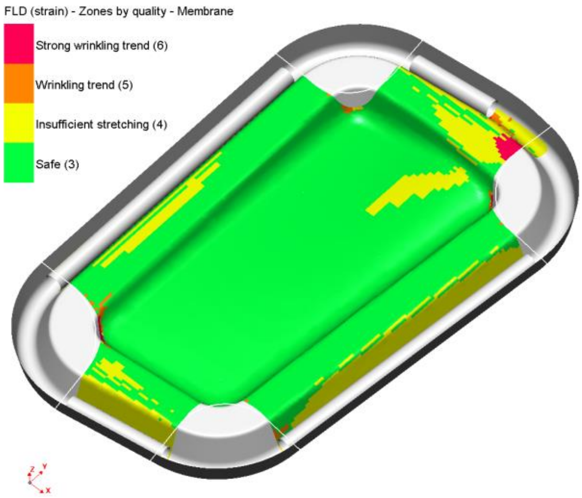
- A forming limit curve of 2mm 2024 aluminium alloy clad sheet that has been heat solution treated to –W temper condition.
- Material stress-strain curve based on the Vegter-Lite plasticity law
- Isotropic hardening curves based on the Krupkowsky law
- Kinematic hardening curves based on the Yoshida law.
- Variability in Young's modulus which changes as the material gets work hardened considerably during the forming process.
- Variable friction curves dependant on pressure and velocity.

2.3 Die and Blank Design Conceptualisation

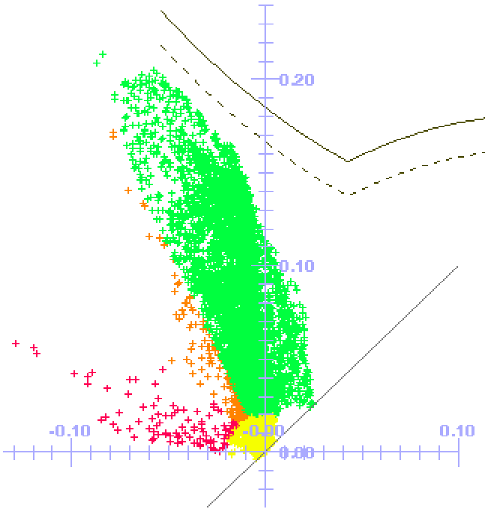
In total 15 simulation concepts were conducted before a final concept was chosen. As the forming process used for this study is the fluid cell, hydroforming process, a bathtub shaped die was opted for as it is a standard die design for this process [32]. Initial concepts were aimed at near net shaped blanks which yielded poor results. An overlapping blank sheet concept was attempted throughout the study. The shape ensures the edges of the blank hang over the sides of the die. This results in the blank's edges becoming constrained due to the friction forces between the blank and die. The blank continues to form into the die cavity, and due to the added forces at its edges, in-plane stretching in the blank is successfully achieved.

In-plane stretching is essential to ensure the blank will have reduced spring back. Insufficient stretching results in the blank returning to its initial flat state. The extent of in-plane stretching that can be achieved is assessed by mean of a forming limit diagram (FLD). In metallic forming, an FLD is used to determine the degree of formability a tooling concept can achieve. Also, it will give a clear indication where fracture and wrinkling can occur. Design changes with further simulation studies can be conducted before a final tooling concept is manufactured.

Figure 6 (a) depicts the FLD zone by the quality of the final state of the forming concept, and it clearly shows that a sufficient amount of in-plane stretching has been achieved with this concept. Figure 6 (b) is the FLD diagram for the aluminium alloy 2024-W 2mm thick clad sheet. It depicts the blank is not susceptible to fracture.



(a)



(b)

Figure 6 - A - FLD zones by quality, B - FLD diagram of the second stage concept

The result is achieved by a two-stage forming concept. The first stage creates a pre-form where the overhang flanges of the blank drape over the edge of the die. The preform is then fitted, and roller wedges are used to restrain the blank so that sufficient in-plane stretching will occur. The first and second stage with roller wedges is depicted in Figure 7 and Figure 8 respectively.

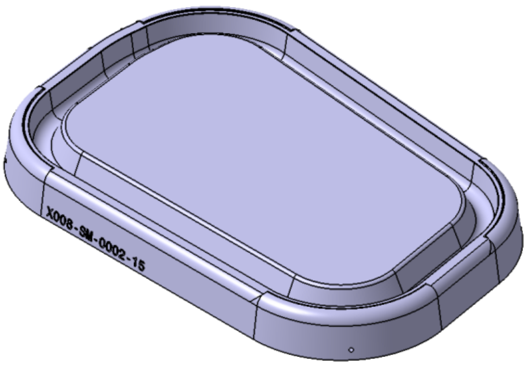


Figure 7 - First stage pre-forming die

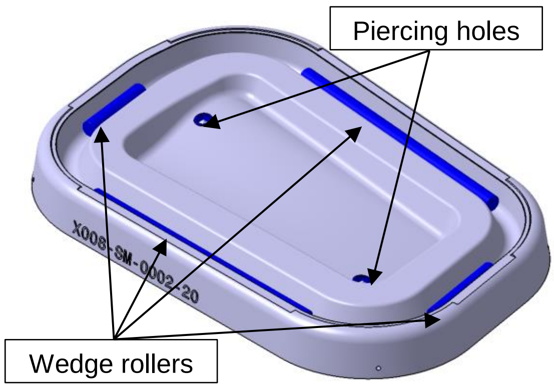


Figure 8 - Second stage forming die

After forming, the component will be trimmed and then inspected. Location tags are incorporated to aid the locating of the component for trimming. Piercing holes in the die are used to create the position of the holes in the location tags, as seen in Figure 8. The location tags and final distortion of the component are depicted in Figure 9. The component deviates from final from by 7.035mm. This amount is deemed acceptable as the component will be flattened during the assembly process as the edges are easier to hold in place during riveting. The chosen concept was chosen based on the final distortion and overall shape. The final validation is the next phase of the case study to be completed.

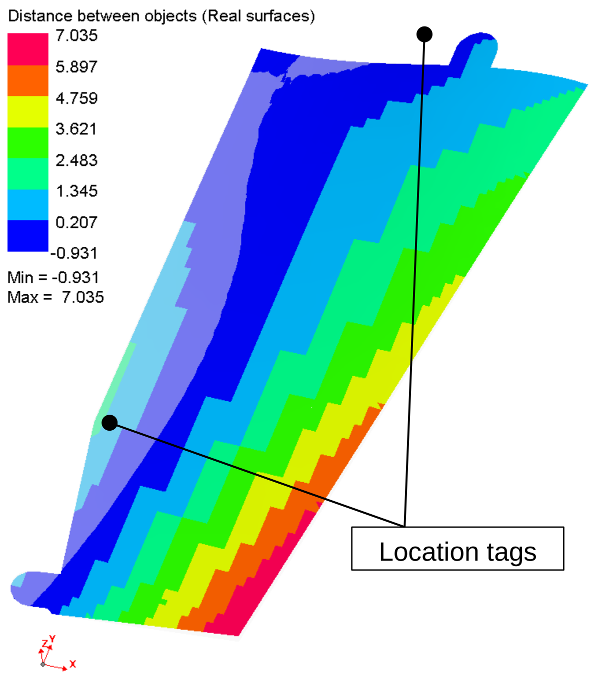


Figure 9 - Component distortion contour plot

3 CONCLUSIONS

The manufacture of aircraft panels using a forming process is an established and extensively researched topic. This study shows the degree of manufacturability in both composite and metallic materials is associated with a component's geometric complexity.

The literature reviewed for this study shows the material selection phase of component design lacks the effective use of simulation studies. This is not to be confused with simulation studies related to the validation of a concept. In concept validation, the component's material, forming die sets and the manufacturing process have been defined. Validation simulation studies give confidence in the tool, and blank design concepts before tool proofing trials are conducted.

Material selection simulations are feasibility-oriented studies where multiple concepts and material types are investigated, and the outputs are then compared against one another. The most appropriate material to manufacture the component, based on the conceptualised blank and tooling, is then used for final concept design where a validation study is then conducted.

Traditional material selection methods and processes are extensively publicised. With the incorporation of material selection simulations and a workflow to assess and compare the outputs, the concept validation process will become less dependent on "trial-and-error" methods.

The outputs from the metallic Pam-stamp case study gave insight into the component's manufacturability. A forming limit diagram is a useful tool as it gives a visual output of areas where defects can arise. It also gives areas where successful in-plane stretching can be achieved. To successfully form the ILP, sufficient in-plane stretching was required to reduce the component's spring back and loss of form back to a nearly flat sheet. Further work must be conducted on the composite process and finally defining the holistic process for the evaluation model for material selection of formed aircraft components.

4 ACKNOWLEDGMENTS

I would like to gratefully acknowledge the support of the staff of the Stellenbosch University, Department of Industrial Engineering and in particular, Prof Dimiter Dimitrov for his guidance and wisdom. This project would not have been possible without the financial and manufacturing support from Aerosud Aviation, and in particular, the research and development team. The financial support provided by the Department of Trade and Industry, in particular, the Technology and Human Resources for Industry Programme (THRIP) is greatly acknowledged.

5 REFERENCES

[1] Nui, M. C. 2000. *Composite Airframe Structures Practical Design information and data*. Hong Kong: Hong Kong Conmilit Press Limited.

[2] Jones, R. M. 1999. *Mechanics of composite materials*. Philadelphia, United States of America: Taylor and Francis Incorporated.

[3] Budynas−Nisbett. 2008. *Shigley's Mechanical Engineering Design*, 8th Edition. United States of America: The McGraw−Hill Companies.

[4] Peters, S. T. 1998. *Handbook of Composites*. New York: Chapman & Hall.

[5] Aniform. 2016. *Aniform Virtual Forming*. Available: http://www.aniform.com/. [2016, May 01].

[6] Mashau, M. An Investigation into the Manufacturing of Complex, Three Dimensional Components using Continuous Fibre Reinforced Thermoplastic Composites. Available: https://hdl.handle.net/10539/24191 [2017. January]

[7] ESI. 2016. ESI – *Virtual prototyping software*. Available: https://www.esi-group.com/. [2016, May 01].

[8] ETA. 2016. *Dynaform*. Available: http://www.eta.com/inventium/dynaform. [2016, May 01].

[9] Nui, M. C. 1995. *Airframe structural design*, 8th Edition. Hong Kong: Hong Kong Conmilit Press Limited.

[10] Ismail, M. Y., Jahan, A., Mustapha, A., Sapuan, S. M. 2009. Material screening and choosing methods - A review. Materials and Design, 31 (2010) 696-705.

[11] Aguiar, C., Ramalhete, P. S., Senos, A. M. R. 2009. Digital tools for material selection in product design. *Materials and Design*, 31 (2010) 2275–2287.

[12] Ashby, M. 2011. Materials Selection in Mechanical Design, 4th Edition. Published by Elsevier Ltd.

[13] Ljungberg, L. Y. 2002. Materials selection and design for structural polymers. *Materials and Design*, 24 (2003) 383–390.

[14] Chakraborty, C., Chatterjee, P. 2011. Material selection using preferential ranking methods. *Materials and Design*, 35 (2012) 384–393.

[15] Sala, G., Di Landro, L., Cassago, D. 2002. A numerical and experimental approach to optimise sheet stamping technologies: polymers thermoforming. *Materials and Design*, 23 (2002) 21-39.

[16] Harrison. P., Clifford. M., Long. A. 2004. Shear characterisation of viscous woven textile composites: a comparison between picture frame and bias extension experiments. *Composites Science and Technology*, 64 (2004) 1453–1465.

[17] Wang. P., Hamila. N., Boisse. P. 2013. Thermoforming simulation of multilayer composites with continuous fibres and

thermoplastic matrix. *Composites: Part B*, 52 (2013) 127–136.

[18] Chen. Q., Boisse. P., Parl. H., Saouab. A., Breard. J. 2011. Intra/inter-ply shear behaviors of continuous fiber reinforced thermoplastic composites in thermoforming processes. *Composite Structures*, 93 (2011) 1692–1703.

[19] Sadighi. M., Rabizadeh. E., Kermansaravi. F. 2008. Effects of laminate sequencing on thermoforming of thermoplastic matrix composites. Journal of materials processing technology, 201 (2008) 725–730.

[20] McGraw-Hill Dictionary of Engineering. 2002. (2nd Edition). New York: Two Penn Plaza.

[21] Boisse. P., Zouari. B., Daniel. J. 2006. Importance of in-plane shear rigidity in finite element analyses of woven fabric composite preforming. *Composites: Part A*, 37 (2006) 2201–2212.

[22] Rozant. O., Bourban. P., Mánson. J. 2000. Drapability of dry textile fabrics for stampable thermoplastic preforms. *Composites: Part A*, 31 (2000) 1167–1177.

[23] De Bilbao. E., Soulat. D., Hivet. G., Gasser. A. 2009. Experimental study of bending behaviour of reinforcements. *Experimental Mechanics*, 50 (2010) 333–351.

[24] Solfronk. P., Sobotka. J. 2011. Utilization of Forming Tool with Variable Blankholder Force for Drawing of Al Alloys. *Physics Procedia*, 22 (2011) 233 – 238.

[25] Kumar. S., Amjith. T., Anjaneyulu. C. 2016. Forming Limit Diagram Generation of Aluminum Alloy AA2014 Using Nakazima *Test Simulation Tool. Procedia Technology*, 24 (2016) 386 – 393.

[26] Yershov. A., Kotov. V., Loginov. Y. 2012. Optimization Of The Initial Form Of A Semifinished Product In Pam-Stamp 2g. *Metallurgist, Vol. 56, Nos. 3–4*, 2012 231-235.

[27] Fyllingen. Ø., Hopperstad. O., Lademo. O., Langseth. M. 2009. Estimation of forming limit diagrams by the use of the finite element method and Monte Carlo simulation. *Computers and Structures*, 87 (2009) 128–139.

[28] Kadkhodayan. M., Moayyedian. F. 2011. Analytical elastic–plastic study on flange wrinkling in deep drawing process. *Scientia Iranica B*, (2011) 18 (2), 250–260.

[29] Elangovan. K., Sathiya Narayanan. C., Narayanasamy. R. 2010. Modelling of forming limit diagram of perforated commercial pure aluminium sheets using artificial neural network. *Computational Materials Science*, 47 (2010) 1072–1078.

[30] Bong H., Barlat. F., Lee. M., Ahn. D. 2012. The forming limit diagram of ferritic stainless steel sheets: Experiments and modelling. *International Journal of Mechanical Sciences*, 64 (2012) 1–10.

[31] Altan, T., Tekkaya, A. 2012. *Sheet Metal Forming Processes and Applications*. ASM International.

[32] MMPDS-07. 2012. *Metallic Materials Properties Development and Standardization (MMPDS)*. Scientific Report. U.S. Department of Transportation, Federal Aviation Administration.

[33] APB 3-1119. 2005. *Heat treatment of aluminium*. Process specification Airbus UK.

6 BIOGRAPHY

Dimiter Dimitrov obtained his PhD degree in Technical Sciences (Manufacturing Engineering) from the Technical University of Dresden. In 1999 he was appointed Professor in Advanced Manufacturing at the University of Stellenbosch, South Africa.

Mr Gerber has 19 years' experience in the aviation industry spread over a wide spectrum of activities including strategy, operations and project management, as well as governance and business development. He specialises in deploying new technology programmes and industrialisation. Today he leads the research and development division inside Aerosud Aviation. Within this portfolio he manages several "world first" technology programmes with which Aerosud has earned international recognition for innovative solutions coupled with accelerated delivery.

A Modular Framework for Cognitive Assistance Systems in Manual Assembly

L. Merkel[1], J. Weth[2], R. Sochor[1], C. Berger[1], S. Braunreuther[1,3], G. Reinhart[1,2]
[1]Fraunhofer Research Institution for Casting, Composite and Processing Technology IGCV, Augsburg, Germany
[2]Technical University of Munich, Garching, Germany
[3]Hochschule Augsburg University of Applied Sciences, Augsburg, Germany

Abstract
Growing numbers of variants raise complexity in manufacturing systems and require higher flexibility. Especially in assembly systems, full automation is not always economical in high-variant production scenarios. Manual assembly plays an important role, since human flexibility is a key benefit when quick changes in the product portfolio have to be performed. Digital technologies create potentials for data analytics to increase transparency and productivity. In order to increase flexibility, productivity and transparency in manual assembly systems, cognitive assistance systems can be introduced. They assist assembly workers by providing, processing and collecting information to support manual tasks such as picking, joining or documentation. Currently, there is a lot of research on new technologies to improve human-machine-interaction and to provide targeted training for employees. However, employee acceptance as well as the profitability of new technologies are key factors that need to be investigated when cognitive assistance systems are introduced. Although assembly systems can benefit from these technologies, several companies are not yet making use of cognitive assistance systems. This paper presents a modular framework for cognitive assistance systems in manual assembly. A cognitive assistance system is modeled and described by its components, technologies and capabilities. This enables the design of an application-specific assistance system using a component and technology database. Since there are several ways to exchange relevant information between the assistance system, assembly workers, workplace and IT-systems, a modular framework is introduced. This framework defines six types of information flow and its needed capabilities. This framework can be used to enable a needs-based introduction process of cognitive assistance systems. An application of the framework is shown at the Learning Factory for Cyber-Physical Production Systems in Augsburg, Germany.

Keywords
Manual assembly, cognitive assistance systems, flexibility

1 INTRODUCTION

Trends such as mass customization and personalization of products increase complexity in manufacturing systems [1]. Therefore manufacturing systems need to be flexible, changeable and reconfigurable [2]. Especially in assembly scenarios with small lot sizes and short product lifecycles, full automation is often not economical because of its high reconfiguration effort. Instead, repetitive tasks should be performed by automated systems and often changing tasks by manual workers because of high human flexibility and human's capability to learn new tasks in a short period of time [3]. In manual assembly systems, physical and cognitive assistance system can be used to support manual tasks [1, 3]. Cognitive assistance systems support the workers by perceiving tasks and making decisions [1]. Studies show that rising complexity increases errors in manual assembly and cognitive assistance can reduce the number of errors [4]. In addition to their impact on quality, assistance systems can also increase productivity in manual assembly [5]. This is achieved by directly providing all relevant information, which decreases time used for searching information in documents or by asking coworkers. In addition to current technologies such as worker information systems [6], several new technologies, e.g. augmented reality [7] are currently being developed in research. This increases the variety of software and hardware technologies that can be used to support manual assembly workers. These growing amounts of options to support manual work in assembly makes it necessary to develop a systematic approach to introduce cognitive assistance systems in manual assembly systems. This approach should match the requirements of manual assembly workers with the assistance capabilities of available technologies.

2 LITERATURE REVIEW

2.1 Cognitive assistance systems and current trends

Several terms are used to describe different approaches of applying digital support to manual assembly systems. The term 'Worker information systems' (WIS) [6] is often used for systems that provide the worker with information such as working instructions. However, a large benefit of digital support lies in bidirectional information flow that creates the possibility to document information about the assembly process. In addition to 'cognitive assistance systems', the term 'digital assistance systems' [8] is also often used for similar purposes. Current research on cognitive assistance systems includes the automatic generation of information from CAD databases [9] and qualification-based assistance [10]. Furthermore, new technologies for human-machine interaction such as projections [11] or motion tracking [12] are currently being developed for assistance systems in manual assembly.

2.2 Complexity evaluation and planning of cognitive assistance systems

In order to plan a needs-based cognitive assistance system, it is necessary to analyze the complexity of the supported assembly tasks. Approaches to evaluate the assembly complexity exist that use detailed geometrical information of the assembly parts [13]. These concepts can be applied during the design phase of a product to decrease complexity by evaluating different design concepts. In order to plan a cognitive assistance system, not only geometrical information, but also further complexity drivers have to be considered. Zeltzer et al. [14] have identified 11 complexity drivers that characterize the objective and perceived complexity at mixed model assembly workstations. Claeys et al. [15] are evaluating the need for cognitive support by extrinsic factors, such as operational and product complexity as well as intrinsic factors of the operator, such as his or her experience. Hold et al. [8] are structuring assembly tasks using MTM to evaluate the complexity caused by product, workstation and operator. Zäh et al. [16] are measuring complexity of manual assembly operations by using a temporal, a cognitive and a knowledge-based factor.

The aggregation of different complexity drivers using one of the presented models is necessary to identify the workstations with the highest need for support in order to decide where to introduce a cognitive assistance system first. However, the aggregated evaluation of different complexity drivers cannot be used to design the manner of support of a cognitive assistance system since different complexity drivers result in different needs and therefore require different support technologies. A scenario with often replaced workers and standardized products might have the same overall complexity index value than a scenario with 'lot size one' assembly and experienced workers, but the form of support is different in both cases. It is therefore necessary to evaluate different complexity drivers individually and to derive their particular needed support from the assistance system in order to introduce a needs-based assistance system.

3 MODULAR FRAMEWORK FOR COGNITIVE ASSISTANCE SYSTEMS

3.1 Cognitive assistance system structure model

In order to model a cognitive assistance system, it can be described by its components, technologies and capabilities [17, 18]. Figure 1 shows a UML diagram of the developed cognitive assistance system model.

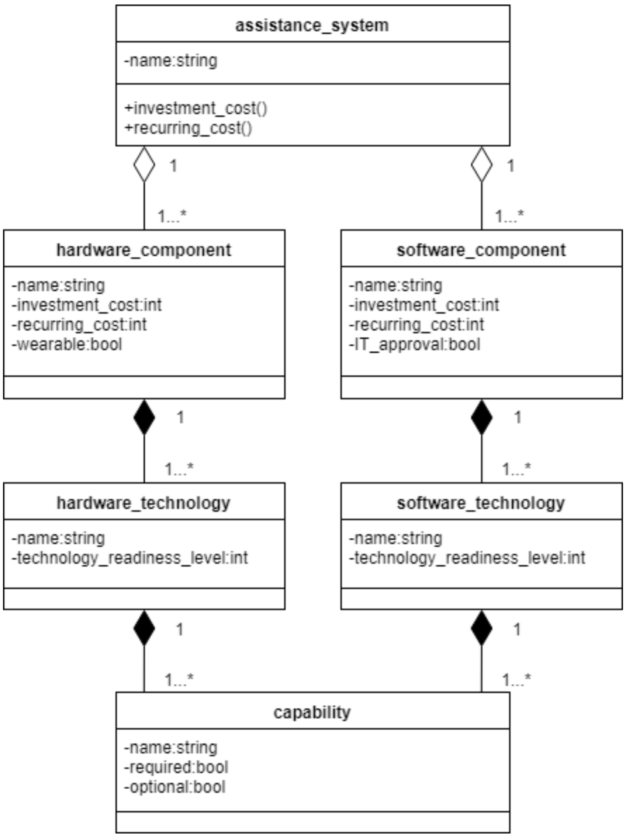


Figure 1 - UML diagram of the cognitive assistance system structure model

The elements of an assistance system are modeled as components that can be purchased and are either hardware or software components. The entirety of components creates the assistance system. A component is described by its technologies with their own technology readiness level. Furthermore, a technology offers capabilities for the assistance system. Capabilities and technologies can only exist if there is a technology or component that provides them. However, components can also exist without being applied in an assistance system. As an example, a regular computer display component consists of display technology and therefore has the capability of providing information, whereas a touchscreen also consists of input technology to receive information from the worker. The priority of a capability can be classified by its *required* and *optional* boolean value. Depending on the requirements, a technology and component database can then be used to create use case-specific solutions for a cognitive assistance system.

3.2 Information flow model and cognitive assistance functions

Components and technologies of the cognitive assistance system structure model are created by hardware and software developers and can also be used for different purposes in addition to cognitive assistance systems. Often new technologies arise from the aerospace, military or gaming industries. The capabilities in this model are introduced especially for cognitive assistance systems in assembly to characterize the flow of information going in and out of a cognitive assistance system.

Figure 2 shows the developed information flow framework for cognitive assistance systems that displays the three objects that a cognitive assistance system exchanges information with [17].

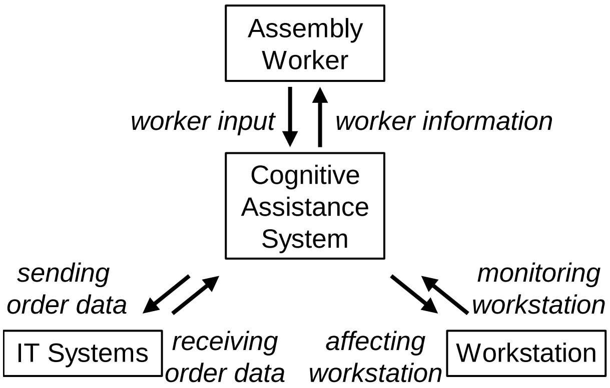


Figure 2 - Information flow framework for cognitive assistance systems

The cognitive assistance system exchanges information with the assembly worker by providing information or receiving input from the worker. Information is also exchanged with higher-level IT systems that manage order-specific information. Furthermore, a cognitive assistance system also exchanges information consisting of product, needed materials as well as tools required for the assembly with the workstation.

Since the information flow with those three objects is bidirectional, six groups of information flow capabilities can be defined.

- *Worker information*: Providing the assembly worker with information
- *Worker input*: Receiving information by manual worker input
- *Sending order data:* Sending information to higher-level IT systems
- *Receiving order data:* Receiving order-specific information from higher-level IT systems
- *Affecting workstation:* Sending information to intelligent tools or products
- *Monitoring workstation:* Monitoring the product or intelligent tools

In order to make use of an assistance system, it is always necessary to have input as well as output information flow. In order to make a needs-based selection, the term 'assistance function' is introduced. Assistance function is a set of input and output information flows that supports the worker in a specific way and is similar to a *user story* in software development. As an example, the assistance function 'providing the assembly worker with variant-specific information' consists of the information flows of *worker information* and *receiving order data*.

3.3 Approach for the definition of capabilities

In order to connect the cognitive assistance system structure model and the information flow framework, a more detailed description of the capabilities is necessary. Capabilities are used to describe the functionality of technologies and are a subgroup of the six information flow groups. Figure 3 shows the approach that was used to define these capabilities.

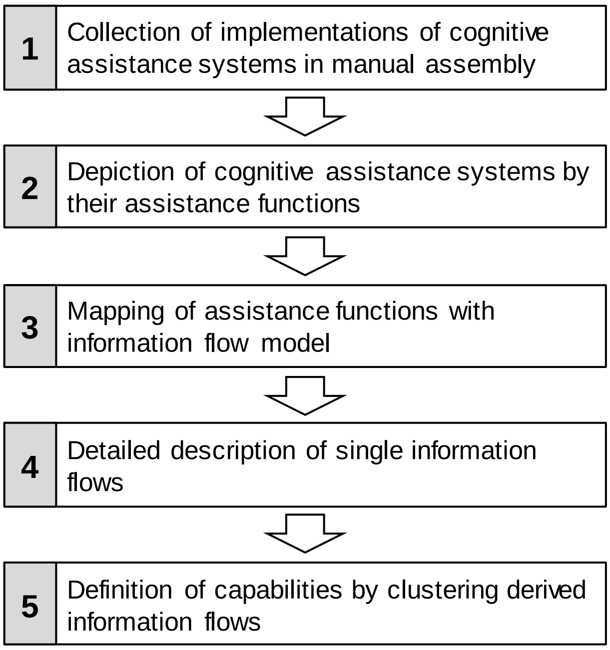


Figure 3 - Approach for the definition of capabilities of cognitive assistance systems

In the first step, 21 different implementions of assistance systems were collected from literature and site visits. The objective of this collection was to select a diverse set of cognitive assistance systems in order to create a higher significance of the results. Subsequently, steps 2 to 4 were performed for each implementation individually. In step 2, every assistance system was depicted by its cognitive assistance functions. The analyzed implementations consisted of between two and six different cognitive assistance functions. In step 3, every assistance function was then mapped with the information flow model to assign which of the six types of information in- and outputs were used by the assistance system. In step 4, every information flow was described in detail. In step 5, the information flows of all of the implementions were grouped by the six information flow types and identical information flows were clustered. This approach resulted in 18 different capabilities in the six information flow groups:

Worker Information:

- *Provision of order-specific information:* Providing the assembly worker with information such as variant-specific parts (1)
- *Assembly process description*: Providing the assembly worker with a general description of the assembly task (2)
- *Alert*: Providing the assembly worker with information about an assembly error that was made or recent changes of a product (3)

Worker input:

- *Binary documentation*: Input of simple information, e.g. to acknowledge specific quality features (4)
- *Textual documentation:* Input of complex information, such as serial numbers or reports (5)
- *Worker identification:* Manual identification of the worker, e.g. by using login information or a smartcard (6)
- *Stress recognition*: Using sensor information to evaluate the current stress level of an employee (7)

Sending order data:

- *Order information request:* Sending information, such as order numbers, to higher-level IT systems (8)
- *Order-specific documentation:* Sending order documentation to higher-level IT systems (9)
- *Checking production status:* Sending information about the current status of the assembly workstation in order to supervise the current status of the assembly system (10)

Receiving order data:

- *Receiving order-specific information:* Receiving of variant- or order-specific information that is needed to complete the assembly process correctly (11)
- *Receiving process parameters:* Receiving information to be transmitted to intelligent tools, e.g. maximum torque (12)

Affecting workstation

- *Writing on intelligent products:* Writing order-specific information on data carriers implemented on product or workpiece carrier, e.g. by sending information via RFID (13)
- *Providing process parameters:* Transmission of information to intelligent tools, e.g. maximum torque (14)

Monitoring workstation:

- *Product identification:* Unique identification of the product at the workstation, e.g. by scanning a barcode (15)
- *Detection of joining task:* Monitoring of the product to detect the completion of correctly or wrongly performed joining tasks (16)
- *Detection of picking tasks:* Monitoring of materials at the workstation that are picked by the assembly worker for a specific order (17)
- *Monitoring process parameters:* Transmitting of information from intelligent tools, e.g. torque curve (18)

4 APPLICATION OF THE FRAMEWORK

4.1 Introduction process for cognitive assistance systems

The cognitive assistance system structure model, the information flow framework and the cognitive assistance system capabilities form the modular framework for cognitive assistance system, which is used for the introduction process for cognitive assistance systems shown in Figure 4.

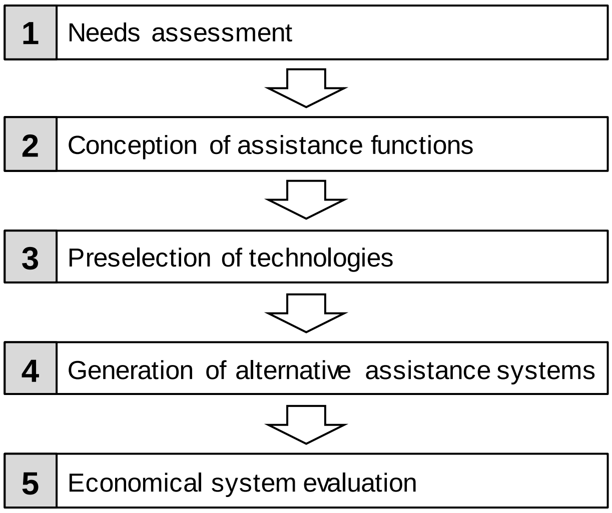


Figure 4 - Introduction process for cognitive assistance systems

In the first step of this process, the need for support is assessed by evaluating various complexity drivers at the assembly workstation. In the second step, the information flow framework is used to create a needs-based definition of assistance functions that shall be introduced by the assistance system. In this phase, the priority of an assistance system is defined as *required* or *optional*. In the third phase, a preselection of technologies and components is made by testing different technologies with the actual assembly workers to include them in the introduction process and to avoid the implementation of technologies that are not accepted by the workers because of little process reliability. In the fourth phase, the combinations of the optional assistance functions result in multiple alternative assistance systems with different assistance functions and therefore different capabilities. In this phase, an algorithm should be used to create a set of components with minimal costs. Costs can be minimized here since the functionality is defined by the assistance functions and the acceptance is ensured by the preselection of technologies in the prior phase. Finally, in the fifth phase, an economical evaluation is performed using quantitative and qualitative factors in order to make a decision for one of the alternative assistance systems that shall be introduced.

4.2 Use case

The framework and the introduction process have been applied at a high-variant gearbox assembly scenario in the learning factory [19] for cyber-physical production systems. The needs assessment and conception phase have resulted in three required and two optional assistance functions (AF) as shown in Table 1:

AF	Description	Type	Capa-bilities
A	Provision of order-specific information	Required	(1), (17)
B	Confirmation of defined quality features	Required	(4), (9)
C	Identification of product at start of assembly	Required	(8), (15)
D	Alert in case of picking errors	Optional	(3), (17)
E	Qualification level based instructions	Optional	(2), (6)

Table 1 - Alternative cognitive assistance functions

Since two assistance functions are optional, four combinations are possible that result in the generation of alternative assistance systems as shown in Table 2.

Solu-tion ID	Assistance functions	Selected components	Investment in EUR
1	A,B,C	#3, #14	13,550
2	A,B,C,D	#3, #14, #16	15,050
3	A,B,C,E	#3, #10	13,650
4	A,B,C,D,E	#3, #10, #16	15,150

Table 2 - Alternative cognitive assistance systems in the application of the framework

The component database consisted of 16 different components from which four different components (#3, #10, #14, #16) were found among the solutions as shown in the third column of Table 2. The results show that the additional cost for the assistance function E which introduces qualification level based instructions is negligible in this case. Instead, a detailed analysis of the benefits of assistance function D, which creates an alert when a picking error is performed, should be performed in order to make a selection between solution 3 and 4.

5 SUMMARY

In order to increase productivity and reduce assembly errors with cognitive assistance systems in manual assembly, several new technologies are currently being developed in research. When introducing an assistance system, it is necessary to select technologies and components according to the requirements of the assembly workers and involve them in the introduction process to ensure acceptance. A model was therefore developed that describes assistance systems by its components, technologies and capabilities. An information flow model was introduced to describe assistance functions when the concept for an assistance system is created. There are 18 capabilities that have been identified to prepare a detailed description of assistance functions and connect them with the technology and component database. The application of this method showed that alternative assistance systems can be generated using a component database and that this approach helps to identify the focus of an economic evaluation by analyzing the additional costs of the optional assistance functions

6 ACKNOWLEDGEMENTS

This publication is part of the “SynDiQuAss” project for excellence in assembly in the context of Industry 4.0 funded by the German Ministry of Education and Research (BMBF) and the European Social Fund (ESF) (grant no. 02L15A281).

7 REFERENCES

[1] Reinhart, G., 2017, Handbuch Industrie 4.0, Carl Hanser, München.

[2] ElMaraghy, H., Schuh, G., ElMaraghy, W., Piller, F., Schönsleben, P., Tsend, M., Bernard, A., 2013, Product variety management, CIRP Annals, 62:629-652.

[3] Spath, D., Ganschar, O., Gerlach, S., Hämmerle, M., Krause, T., Schlund, S., 2013, Produktionsarbeit der Zukunft - Industrie 4.0, Fraunhofer Verlag, Stuttgart.

[4] Fast-Berglund, Å., Fässberg, T., Hellman, F., Davidsson, A., Stahre J., 2013, Relations between complexity, quality and cognitive automation in mixed-model assembly, Journal of Manufacturing Systems, 32:449-455.

[5] Hinrichsen S., Bendzioch S., 2019, How Digital Assistance Systems Improve Work Productivity in Assembly, Advances in Intelligent Systems and Computing, AHFE 2018, 332-342.

[6] Lušić, M., Fischer, C., Bönig, J., Hornfeck, R., Franke, J., 2016, Worker information systems: state of the art and guideline for selection under

consideration of company specific boundary conditions, Procedia CIRP, 41:1113-1118.

[7] Syberfeldt, A., Danielsson, O., Holm, M., Wang, L., 2016, Dynamic operator instructions based on augmented reality and rule-based expert systems, Procedia CIRP 41:346-351.

[8] Hold, P., Erold, S., Reisinger, G., Sihn, W., 2017, Planning and Evaluation of Digital Assistance Systems, Procedia Manufacturing 9:143-150.

[9] Reisinger, G., Komenda, T., Hold, P., Sihn, W., 2018, A Concept towards Automated Data-Driven Reconfiguration of Digital Assistance Systems, Procedia Manufacturing 23:99-104.

[10] Müller, R., Vette-Steinkamp, M., Hörauf, L., Speicher C., Bashir, A., 2018, Worker centered cognitive assistance for dynamically created repairing jobs in rework area, Procedia CIRP 72:141-146.

[11] Rodriguez, L., Quint, F., Gorecky, D., Romero, D., Siller, H.R., 2015, Developing a Mixed Reality Assistance System based on Projection Mapping Technology for Manual Operations at Assembly Workstations, Procedia Computer Science, 75:327-333.

[12] Müller, B.C., Nguyen, T.D., Dang, Q., Duc, B., Seliger, G., Krüger, J., Kohl, H., 2016, Motion tracking applied in assembly for worker training in different locations, Procedia CIRP, 48:460-465.

[13] Samy, S. N., ElMaraghy, H., 2010, A model for measuring products assembly complexity, International Journal of Computer Integrated Manufacturing, 23/11:1015-1027.

[14] Zeltzer, L., Limère, V., Van Landeghem, H., Aghezzaf, E., Stahre, J., 2013, Measuring complexity in mixed-model assembly workstations, International Journal of Production Research, 51/15:4630-4643.

[15] Claeys, A., Hoedt, S., Soete, N., Van Landeghem, H., Cottyn, J., 2015, Framework for Evaluating Cognitive Support in Mixed Model Assembly Systems, IFAC-PapersOnLine, 48/3:924-929.

[16] Zaeh, M.F., Wiesbeck, M., Stork, S., Schuböh, A., 2009, A multi-dimensional measure for determining the complexity of manual assembly operations, Prod. Eng. Res. Devel., 3:489-496.

[17] Merkel, L., Berger, C., Schultz, C., Braunreuther, S., Reinhart, G., 2017, Application-Specific Design of Assistance Systems for Manual Work in Production, Proceedings of the 2017 IEEE IEEM, 1189-1193.

[18] Merkel, L., Starz, J., Schultz, C., Braunreuther, S., Reinhart G., 2017, Digitale Assistenzsysteme in der Produktion, wt Werkstattstechnik online, 107/3:124-128.

[19] Merkel, L.; Atug, J., Merhar, L., Schultz, C., Braunreuther, S., Reinhart, G., 2017, Teaching Smart Production: An insight into the Learning Factory for Cyber-Physical Production Systems (LVP), Procedia Manufacturing, 9:269-274.

8 BIOGRAPHY

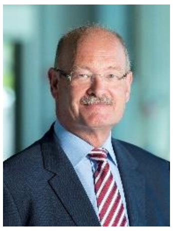
Lukas Merkel obtained his M.Sc. degree in Mechanical Engineering and Management from the Technical University of Munich. Since 2015 he is performing research on cognitive assistance systems at Fraunhofer IGCV.

Johannes Weth obtained his B.Eng. in Mechanical Engineering and is currently a graduate student at the Technical University of Munich in Mechanical Engineering and Management

Robin Sochor obtained his M.Eng. degree in Industrial Engineering. Since 2017 he is performing research on cognitive assistance systems at Fraunhofer IGCV.

Christoph Berger obtained his M.Sc. degree in electrical engineering. Since 2014 he is performing research on production planning and control at Fraunhofer IGCV.

Prof. Dr.-Ing. Stefan Braunreuther serves as professor at the Faculty of Mechanical and Process Engineering at the Augsburg University of Applied Sciences and as head of the department of production planning and control at Fraunhofer IGCV.



Prof. Dr.-Ing. Gunther Reinhart serves as director of the iwb at the Technical University of Munich and as head of the Fraunhofer IGCV. He conducts research into assembly technology, automation and robotics as well as production management and logistics.

Improving Worker Information - Proposal of a Dynamic and Individual Concept

S. Teubner[1], L. Merkel[2], G. Reinhart[1,2], F. Hagemann[3], C. Intra[3]
[1]Institute for Machine Tools and Industrial Management (*iwb*), Garching, Technical University of Munich (TUM), Germany
[2]Fraunhofer Research Institution for Casting, Composite and Processing Technology (IGCV), Augsburg, Germany
[3]MAN Truck & Bus AG, München, Germany

Abstract
Employees face a daily challenge to fulfil high quality and productivity goals, with numerous product variations to be produced. In this context, dynamic and individual worker information systems can provide support. In order to give consideration to the increasing importance of these systems, this article will first offer a definition of this type of assistance system, as well as an overview of existing approaches. Building on the key literature in the field and on expert interviews, we perform a problem-based analysis. This is to identify potential points of improvement within manual assembly that can be addressed by implementing a dynamic and individual worker information system. After clustering the potentials, we deduce requirements for the dynamic and individual information system. Based on a system-theoretical approach, we design a conceptual worker information system. Considering an information system as a black box, the requirements are formulated as system functions. To enable the system functions, we derive types of necessary system elements and relations such as models, views, and a framework of rules.

Keywords
Worker information system, worker assistance, operator instructions

1 INTRODUCTION

In today's turbulent market environment, manufacturing firms face fierce global competition. Many firms react by offering mass customized products or even order-based production. Hence, a adaptive production system, with employees playing a central role, is rendered indispensable. Employees have a high degree of mobility and flexibility for the execution of assembly work, as well as the ability to adapt to new tasks and situations. [1, 2]

At the same time, employees are challenged to meet high quality and productivity goals [3]. In this context, worker information systems can provide support and enable high-variance manual assembly [1]. Through interpretation of the information presented, the workers are enabled to generate the required knowledge for fulfilling the assembly task. Apart from the evident necessity of worker information systems, an advancement towards dynamic and individual information systems is likely to increase efficiency when informing workers about their tasks, and to reduce rework or assembly errors [4] - by for instance encompassing increased awareness of product changes, or information tailored for the individual worker.

2 DEFINITION OF DYNAMIC AND INDIVIDUAL WORKER INFORMATION SYSTEMS

In order to provide an understanding of worker information system capabilities, we will first offer a characterisation in form of a definition.

Assistance systems are utilised in various fields, such as medical technology and care, and production industries. There are thus varying aims of assistance systems, resulting in numerous different classifications of assistance systems, or use cases of assisting humans in their actions. While Sheridan focuses on the man-machine-interaction with levels of automation [5], Karafilides & Weidner develop a periodic system serving as a more holistic approach [6]. Biniok presents a concept of sociotechnical assistance-ensembles [7], taking into account the social dimension of technical developments. For the use case of manual assembly, a classification as shown in Bengler et al. [8] seems appropriate. They distinguish three types of assistance according to the human sequence of action: perception, decision, and execution.

Execution or handling assistance is usually perceived to be the "classical" type of assistance, starting from simple manipulators to complex robotic systems. A system for decision assistance could for instance consist of an intelligent tool providing preventive maintenance notes. Simple instructions as well as augmented reality systems can be classified as perception assistance since they let the worker know which task to perform. Consequently, worker information systems are a subclass of worker assistance systems. Of course all types of assistance can be combined within one system. This could be a camera-based worker-monitoring system giving hints whenever the worker does not correctly assemble. These cognitive assistance systems, combining perception and decision, are often called worker guidance systems, and significantly differ from worker information systems through the feedback-loop. Worker information systems "only" inform workers on an explicit and rule-based way through presenting task related information [9] (based on [10]). The information provided is consequently called worker information. Associated with a dynamically changing production system, dynamic information shall be understood as information depending on the product or the production processes, for example by highlighting product changes within the information presented. **Individual information** is depending on the individual worker, for instance by leveling the degree of information shown according to the qualification of the worker. [9]

3 LITERATURE REVIEW

3.1 Existing Aproaches

Within the existing literature on worker information systems, especially aspects of dynamic and individual information, have been examined. As mentioned above, information systems play an important role in production systems with a high product variety. Therefore, this review is focused on manual, customer-oriented series assembly. Furthermore, flowing assembly environments are examined, since prevailing conditions such as tact dependence and restricted time lead to additional demands regarding information presentation. Due to these circumstances, the use of (camera-based) worker guidance systems is often not suitable. Moreover, the review focuses on workplaces and assembly instructions rather than organizational, cross-worker information.

Many projects focus on knowledge-intense tasks such as maintenance or operating machines. Within APPsist [11], an information system architecture for provision of AI-based knowledge services was developed. Associated with this project, Breitkopf (né Kreggenfeld) has published a dissertation about adaptive assistance [12]. Based on an individual competence profile, he has generated an assistance process. The project Facts4Workers [13] has a similar focus as APPsist and presents a concept for the inclusion of knowledge work on the shop floor.

The next listed projects put an emphasis on interactive process guidance in manual assembly, mostly at workbenches. Within ACIPE [14], Bannat [15] and Wiesbeck [4] develop concepts for generating adaptive assembly instructions based on the monitored assembly process and recorded human state of cognition. Connected through the project motionEAP, Korn [16] and Bächler et al. [17] are also working on adaptive guidance. Korn presents a camera-based context-sensitive information system for work-benches. Bächler et al. develop an individual step-by-step instruction, also based on observed assembly failures and assembly time. Based on the analysis of worker movements, Weisner [18] (project AIM) puts more attention on the ergonomic aspects and individual assembly training. Ahnelt & Bader [19] developed a mobile assembly assistant called Plant@Hand which is able to assume the current most likely work situation, and present context-senitive step-by-step instructions. Within SmartF-IT, Kerber & Lessel [20] focus on adaptive gamification-instructions by utilising individual repetitiveness and failure-frequency.

Lang [10] has published a very fundamental dissertation on worker information systems, but did not work on the development towards dynamisation and individualisation. In contrast, Merkel et al. [21] present an approach to the application-specific design of assistance systems that can encompass dynamic and individual worker information systems.

3.2 Research Demand

Industrial need

Outputs of systems already implemented by the industry tend to be too monotonous and thus fail to attract attention [4] (e.g. the information presented is very static so that worker does not pay sufficient attention to it). These systems do usually not account for the individual information needs, arising from the specific expertise and abilities of the individual employee. The consistent presentation of processual guidance is obviously very inefficient for skilled workers who are already familiar with the repetitive phases of the process. Additionally, dynamic information depending on the production situation, such as quality issues, is rarely given. But both are crucial for increasing efficiency, respectively productivity, and for avoiding mistakes. Besides, they form a foundation of competitiveness in the production of customised goods (see introduction). Therefore, the aim of this research project is to present suitable worker information by developing a dynamic and individual worker information system.

Research gap

The literature review shows that increasing dynamism and individualisation of worker information is currently not sufficiently covered in the literature, especially the combination of both tends to be neglected. This gap particularly pertains to our focus area, that is series assembly, flowing assembly, and assembly instructions.

Ideas and concepts for individual worker guidance are available and show the importance of this research. However, they need to be transferred to worker information systems suitable for workplaces at assembly lines, which have different requirements and options than workbenches.

As the research does not offer a holistic solution for the industrial need, the design of this dynamic and individual worker information system starts with a requirement analysis presented in the next section.

4 PROBLEM-BASED ANALYSIS OF REQUIREMENTS

4.1 Approach

In order to design dynamic and individual information, a problem-based requirements analysis was conducted.

Firstly, we identified potentials for increasing efficiency and decreasing errors that can be addressed by an information system. The identification is based on four sources: expert workshops, a worker survey, work internship, associated literature.

We have conducted three expert workshops in total, one with four foreman from MAN Truck and Bus AG, another one with eight experts from MAN production planning, and a third one with two production planners from Kramer Werke GmbH. Furthermore, we have used 32 semi-structured interviews with MAN workers who are most affected by worker information systems. Additionally, the first author has worked at several assembly stations at MAN for twelve days, in order to gather additional points of improvement (also called “potentials”). We have further completed and confirmed the final collection of potentials by using relevant literature in the field of worker information systems.

The case study partner is the production plant of MAN in Munich where trucks for long-distance transport, distribution transport, or use on a construction site are produced. The production system is characterized by a manual assembly line. Kramer has a similar work environment producing (tele) wheel loaders and telescopic handlers.

Secondly, by clustering the potentials, eight requirements for the dynamic and individual worker information system could be deduced.

4.2 Results

In the following, the potentials identified and the resulting requirements are presented together.

1.) If the information presented fits all product variants, it is usually not rich in detail and not sufficiently specific for assembling without any prior additional knowledge. On the other hand, the amount of information tends to be very vast if all variants are explained in detail. Therefore, information adjusted to the specific assembly order is required. Because this kind of information is fundamentally different from static information, this requirement needs to be mentioned here even though it is already taken into account in some existing information systems.

2.) Due to a large amount of routine tasks for some high runner products, a second problem is that product or process changes are prone to be overlooked. The dynamic information system shall be able to draw attention to changes or exceptional cases.

3.) Improvements and avoidance of mistakes are rather difficult to achieve if the workers are not informed about the tasks most prone to errors. Therefore, preventive information considering past assembly quality needs to be given to draw attention to processes prone to failure.

4.) Especially foremen desire a possibility to give additional information to their workers, for example if an exhibition product is assembled. Consequently, providing an option for comments regarding products or processes is a fourth requirement.

The first four requirements are related to dynamic information. The requirements number 2 and 3 could also be specified as individual information. However, in relation to significant difficulties regarding anonymous data acquisition about who built which assembly order at which point in time, we have found that the added benefit is not sufficiently high. One could say that the information connected to requirements 2, 3 and 4 are already available to the workers. But it is often only presented in weekly meetings and not directly at the workstation connected to a specific assembly order, and therefore tends to be forgotten. The following requirements are related to individual worker information.

5.) For example for very difficult or tact time exceeding variants, an additional worker can support the main responsible worker, and the assembly task is shared. In this case, information is usually not split between the workers involved, and each role needs to search for his/her information. Therefore, the fifth requirement is to split the information between roles in multiple worker cases.

6.) For each role, the given information is usually not adjusted to the qualification level of the worker. But young and unexperienced workers have different information needs than experienced ones. This results in presenting too much or too little information for many workers. Experience levels are often assessed by foremen or team leaders and matched to a company internal qualification level. However, these assessments are in most cases not used for information systems. Consequently, a qualification specific information adjustment is required.

7.) Due to a huge range of variants, it can sometimes take several days or weeks to train new workers. The information system shall be able to support the teaching process, in addition to the previously mentioned qualification adjustment. Processes mastered can be checked, difficult ones can be marked, and an overall learning progress could be calculated.

8.) The last requirement addresses a similar problem. Workers can lose track of which assembly processes are identical on which products, and how familiar they should be with upcoming assembly orders. To help the workers, individual transparence over variants of working procedures shall be available.

5 SYSTEM DESIGN

After analysing the requirements, a dynamic and individual worker information system needs to be developed.

To design the system conceptually, a system theoretical approach can be used. It describes a system from three different perspectives - Ropohl names a hierarchical view, a functional view, and a structural view [22]. Using this abstraction level, we arrive at a generic concept of a worker information system also suitable for production systems similar to the ones of our case study partners.

From a hierarchical perspective, a worker information system is connected to super- and subsystems [22]. A supersystem can be a general employee information system that the worker information system is embedded in. Some kinds of Manufacturing Execution Systems (MES) can be understood as subsystems since they give input information. This perspective might be suitable to describe the information systems technical architecture. Since this is not the main focus of this research project, it is not detailed further.

The next two sections present a system design using the functional and the structural view. Figure 1 provides an overview.

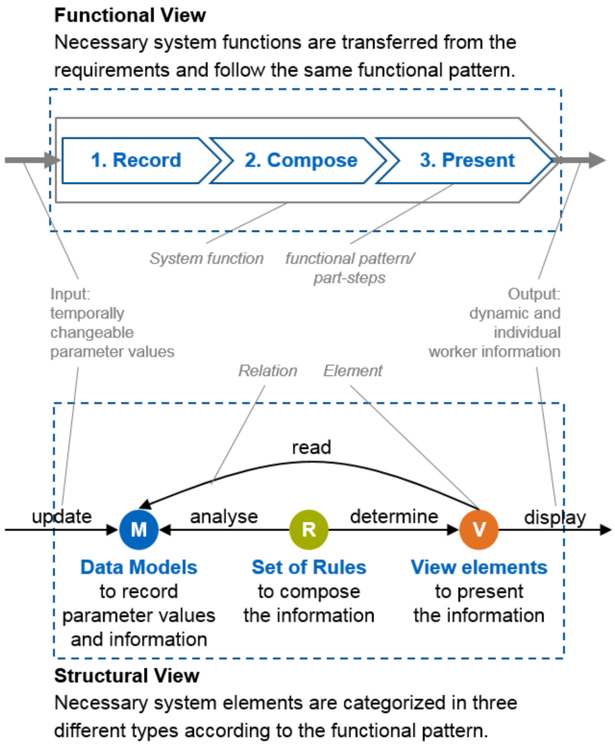


Figure 1 - Conceptual system design

5.1 System functions

Understanding the information system as a black box [22], as the functional view suggests, we transfer the requirements from section 4.2 into system functions. Following the eight requirements, eight system functions are defined.

Dynamic system functions

1. Assembly order specific information adjustment
2. Drawing attention to changes
3. Drawing attention to mistakes
4. Commenting products or processes

Individual system functions

5. Role-specific information split
6. Qualification-specific information adjustment
7. Teaching process support
8. Individual transparency of working procedures variants

All functions follow the same **functional pattern**. At first, the system input (temporally changeable parameter values) needs to be **record**ed. The second part-step is to **compose** the information depending on the parameter values. Finally, the output (dynamic and individual worker information) will be **present**ed.

5.2 System elements and relations

The structural view describes the system by its internal system elements and their relations (the network of relations is called system structure). Their interaction enables the system functions [22]. According to the functional pattern, three types of elements are considered: **data model**, **set of rules**, and **view elements**.

The **data models** record the parameter values of inputs, which the presented information shall depend on, and include the worker information itself and its changes. For example, inter alia, a product model storing product type and changes, and an instruction model including texts and pictures to describe the assembly task, would be necessary for function 2: drawing attention to changes. The **set of rules** is necessary to compose the right information. One rule could be that a product change should be displayed 10 times or for one week. **View elements** present the information. To draw attention to an aspect, a view for the instruction as well as a view element, such as a flashing box around the product change, need to be implemented.

The data models are related to and updated by the contextual situation, or more specifically by the product and process (for dynamic information) and the worker (for individual information). After every update of the data model, the set of rules analyses the parameter value changes and determines the necessary views. The views read the data models to display the composed worker information.

6 CONCLUSIONS AND OUTLOOK

The industrial need and the literature review demonstrate that a holistic concept for a dynamic and individual worker information system, suitable for customer-oriented flowing series assembly, needs to be developed.

Therefore, at first a profound problem-based requirements analysis was conducted to specify dynamization and individualization. We furthermore develop a theoretical system concept of functions and elements. Since we use a high-level abstraction process, we are able to derive a solution also suitable for firms with work environments similar to the case study partners.

Building upon the eight requirements, figures 2 and 3 provide an overview of the required system functions and necessary system elements. The figures also demonstrate that we aim at creating a modular system. Depending on the specific needs of a workstation, different functions can be selected and implemented.

The research will be continued as follows: Based on our literature reviews and studies, we are currently working on further specifying necessary data models, set of rules, and view elements. At the same time we are implementing a prototype of a dynamic and individual worker information system.

An evaluation in cooperation with the industry partner MAN will be the final step to challenge the system capabilities regarding the requirements. Utilizing this case study, our deductive approach to developing the system elements can be validated. At the same time, we can investigate whether the system helps achieve an efficient information process and reduce rework needs.

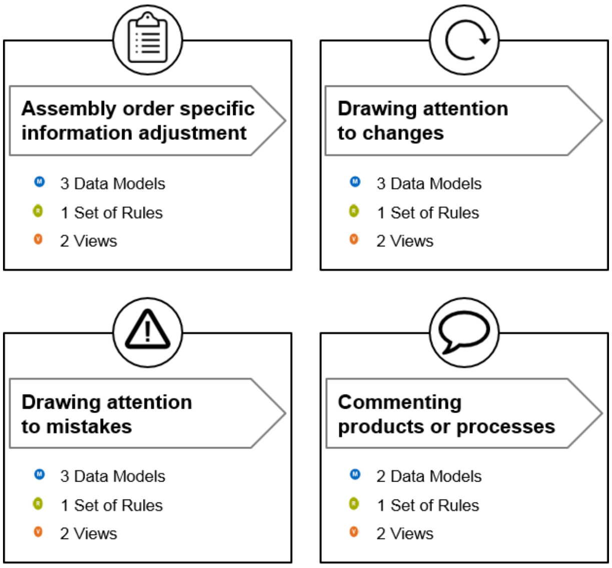


Figure 2 - Required system functions and necessary system elements for dynamic worker information

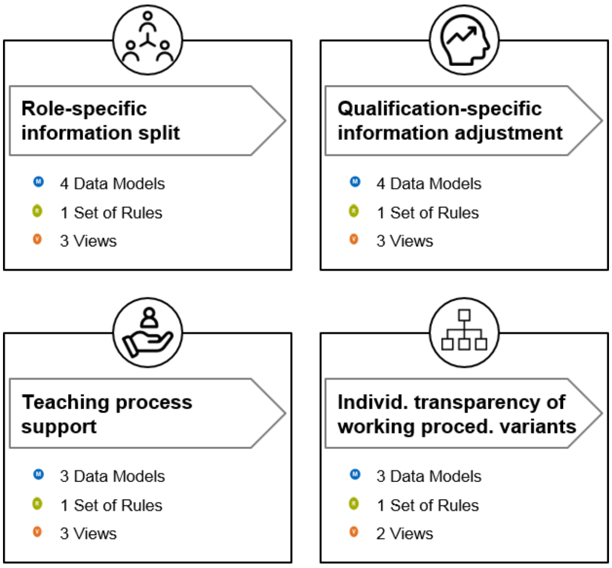


Figure 3 - Required system functions and necessary system elements for individual worker information

7 REFERENCES

[1] Feldmann, K., Lang, S., 2005. Multimediale Informationssysteme an manuellen Arbeitsplätzen: Prozesssicherheit und Effizienzsteigerung durch Bereitstellung und Rückmeldung richtiger sowie aktueller Informationen *21*, p. 25.

[2] Spath, D., Editor, 2013. *Produktionsarbeit der Zukunft – Industrie 4.0*. Fraunhofer.

[3] Aehnelt, M., Bader, S., 2014. Mobile Informationsassistenz für die Montage, in *Technische Unterstützungssysteme, die die Menschen wirklich wollen: Erste Transdisziplinäre Konferenz*, Hamburg.

[4] Wiesbeck, M., 2014. *Struktur zur Repräsentation von Montagesequenzen für die situationsorientierte Werkerführung,* 1st edn. Utz, München.

[5] Sheridan, 1992. *Telerobotics automation and human supervisory control*.

[6] Karafillidis, A., Weidner, R., 2016. Taxonomische Kriterien technischer Unterstützung: Auf dem Weg zu einem Periodensystem, in *Technische Unterstützungssysteme, die die Menschen wirklich wollen: Zweite Transdisziplinäre Konferenz*, Hamburg, p. 233.

[7] Biniok, P., 2016. Soziotechnische Assistenzenensambles: Aushandlung von Bedarf und Akzeptanz technischer Unterstützungssysteme, in *Technische Unterstützungssysteme, die die Menschen wirklich wollen: Zweite Transdisziplinäre Konferenz*, Hamburg, p. 269.

[8] Bengler, K., Lock, C., Teubner, S., Reinhart, G., 2017. Grundlegende Konzepte und Modelle, in *Handbuch Industrie 4.0*, Hanser, München, p. 54.

[9] Teubner, S., Reinhart, G., Haymerle, R., Merschbecker, U., 2016. Individuelle und dynamische Werkerinformation: Einordnung, Definition und Beschreibungsmodell, in *Technische Unterstützungssysteme, die die Menschen wirklich wollen: Zweite Transdisziplinäre Konferenz*, Hamburg, p. 349.

[10] Lang, S., 2007. *Durchgängige Mitarbeiterinformation zur Steigerung von Effizienz und Prozesssicherheit in der Produktion,* 1st edn. Meisenbach, Bamberg.

[11] Ullrich, C., Aust, M., Blach, R., Dietrich, M. *et al.*, 2015. Assistenz- und Wissensdienste für den Shopfloor, in *DeLFI Workshops 2015: Proceedings der Pre-Conference Workshops der 13. E-Learning Fachtagung Informatik (DeLFI 2015)*, p. 47.

[12] Breitkopf, N., 2018. *Adaptive Assistenz in der Produktion: Eine Methodik zur individuellen Mitarbeiterbefähigung*, Bochum.

[13] Hannola, L., Lacuera-Perez, F.J., Steinhüser, M., Kokkonen, K. *et al.*, 2017. An Evaluation Framework for Worker-Centric Solutions in Production Environments, in *24th Internation Conference on Production Research (ICPR)*.

[14] Zäh, M.F., Wiesbeck, M., Engstler, F., Friesdorf, F. *et al.*, 2007. Kognitive Assistenzsysteme in der manuellen Montage: Adaptive Montageführung mittels zustandsbasierter, umgebungsabhängiger Anweisungsgenerierung *97*, p. 644.

[15] Bannat, A., 2014. *Ein Assistenzsystem zur digitalen Werker-Unterstützung in der industriellen Produktion*.

[16] Korn, O., 2014. *Context-Aware Assistive Systems for Augmented Work. A Framework Using Gamification and Projection*.

[17] Bächler, A., Bächler, L., Autenrieth, S., Kurtz, P. *et al.*, 2015. Entwicklung von Assistenzsystemen für manuelle Industrieprozesse, in *DeLFI Workshops 2015: Proceedings der Pre-Conference Workshops der 13. E-Learning Fachtagung Informatik (DeLFI 2015)*, p. 56.

[18] Weisner, K., Knittel, M., Enderlein, H., Wischniewski, S. *et al.*, 2016. Assistenzsystem zur Individualisierung der Arbeitsgestaltung: Einsatz von Smart Devices zur kontextsensitiven Arbeitsunterstützung *111*, p. 598.

[19] Aehnelt, M., Bader, S., 2016. Providing and Adapting Information Assistance for Smart Assembly Stations, in *SAI Intelligent Systems Conference (IntelliSys)*, Springer International Publishing, Cham, p. 540.

[20] Kerber, F., Lessel, P., 2015. Adaptive und gamifizierte Werkerassistenz in der (semi-)manuellen Industrie 4.0-Montage, in *DeLFI Workshops 2015: Proceedings der Pre-Conference Workshops der 13. E-Learning Fachtagung Informatik (DeLFI 2015)*, p. 28.

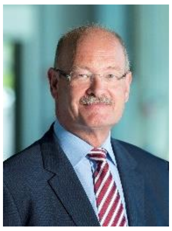
[21] Merkel, L., Berger, C., Schultz, C., Braunreuther, S. *et al.*, 2017. Application-Specific Design of Assistance Systems for Manual Work in Production, in *International Conference on Industrial Engineering and Engineering Management (IEEM)*, p. 1189.

[22] Ropohl, G., 1999. *Allgemeine Technologie: Eine Systemtheorie der Technik,* 2nd edn. Hanser, München.

8 BIOGRAPHY

Severin Teubner holds a M.Sc. degree in Mechanical Engineering from the Technical University of Munich. Currently, he is a PhD-candidate at the Institute for Machine Tools and Industrial Management. His area of research is worker information systems.

Lukas Merkel obtained his M.Sc. degree in Mechanical Engineering and Management from the Technical University of Munich. Since 2015 he is performing research on cognitive assistance systems at Fraunhofer IGCV.



Prof. Dr. Gunther Reinhart is holder of the Chair for Industrial Management and Assembly Technology of the *iwb* (Institute for Machine Tools and Industrial Management) at the Technical University Munich. He is also the chairman of the Bavarian Cluster for Mechatronics and Automation e.V. Since July 1st 2016 he is executive director of Fraunhofer Research Institution for Casting, Composite and Processing Technology (IGCV).

Dr. Florian Hagemann obtained his doctoral degree in Mechanical Engineering from the Technical University of Munich (TUM). He is currently leading the testing center of the truck assembly line of MAN Truck & Bus AG.

Dr. Carsten Intra obtained his doctoral degree from the RWTH Aachen University after completing his academic studies in mechanical engineering and his postgraduate studies in economics there. He started his career in 2001 at MAN where from 2012 onward, Dr. Carsten Intra was the Executive Board member for Production and Logistics at MAN Truck & Bus AG, and from 2015 also the Executive Board member for Research and Development. Since 2017, he is Chief Human Resources Officer and Executive Board member responsible for employee relations at MAN SE and MAN Truck & Bus AG as well as the Head of Human Resources at TRATON AG (formerly Volkswagen Truck & Bus AG).

Modal Response of a Workpiece in an Automated Flexible Fixture for a Flexible Manufacturing System

E. Slabbert[1], G. Bright[1], A. Walker[1]
[1]Manufacturing and Robotics Research Group, Department of Mechanical Engineering, University of Kwa-Zulu Natal, South Africa

Abstract
Modal response data for a symmetric workpiece in an Automated Flexible Fixture for a flexible manufacturing environment can be used to develop a chatter reduction control system. The purpose of this research is to investigate the relationship between the fixturing and vibration of a work piece, as a mechanism to reduce vibration and hence chatter from occurring during machining. This paper addresses the mechanics of machine chatter and focuses on the simulation and physical testing results of the modal response of a symmetric workpiece. Simulations of the modal analysis showed a strong correlation between fixturing position as well as part geometry, and modal response. Both the fixturing position and geometry influenced the intensity of resonance frequencies, revealing trends that indicated certain fixturing positions yielded a more favourable response than others for different geometric features of a part. The experimental testing validated the simulation results, revealing the same trends in the relationship between fixturing position and modal response. The physical testing results showed less resolution than the simulations, due to a higher level of dampening in the system than expected, that left some trends in the geometry-modal relationship unvalidated. Overall it was concluded that fixturing has a strong effect on vibration and could be used to optimize machining processes.

Keywords
Chatter, Fixturing, Vibration

1 INTRODUCTION

Present day manufacturing is extremely competitive, with innovations in technologies driven by the customers' needs. One such consumer need is that of individualism and ultimately customization of consumer products. Current industry drives such as Factories of the Future, the Internet of Things and Industry 4.0 aim to develop the technologies necessary to accelerate current manufacturing practices towards those that are more efficient, reconfigurable and autonomous, and ultimately enable mass customization to be realized.

The Factories of the Future focuses specifically on developing Flexible Manufacturing System (FMS) technologies. One of the outcomes of Factories of the Future initiative is the design and implementation of intelligent, reconfigurable jig and fixture systems to accommodate the increase in part type, size and geometry of a FMS environment. In order to make mass customization economically viable, conditions that limit performance must be eliminated. One such condition, chatter, is the most problematic and limiting factor in machining [1]. Fixturing is a major contributor to the performance of both the individual machining system and the overall manufacturing process [2], with as much as 40% of all parts rejected in manufacturing processes being connected to poor fixture design [3]. There is a strong link between fixturing and chatter, where in practice some instances of chatter may be eliminated by reworking the fixture to hold the part more securely [4]. Thus the optimization of work-holding for the FMS is of utmost importance.

This research focuses on establishing the relationship between fixturing and chatter as the basis for introducing a layer of intelligence to a proposed Automated Flexible Fixture (AFF). This layer of intelligence will be in the form of a control system that, in conjunction with embedded sensors, will monitor vibrations in the AFF system to detect the presence of chatter, and actively adjust the fixturing setup for a specific part if necessary until the vibrations are reduced to the point that chatter is eliminated. The relationship between fixturing and chatter will be investigated through modelling of the modal characteristics of parts clamped by the AFF in different fixturing setups

2 CAUSES OF CHATTER IN MACHINING

An understanding of the relationship between vibration and chatter must be established before analysing the correlation between fixturing and chatter for the AFF. This section will give a brief outline on the fundamentals on vibration and their effect on machining.

2.1 Vibration and chatter

Chatter is a self-exciting vibration, characterized by the loud and distinctive noise emitted, that occurs during machining. Chatter causes the end mill tool piece to produce an uneven cut that appears as periodic grooves or waves on the machined surface. There are 3 main types of vibration involved in the creation of chatter, those being; free vibration, forced vibration and self-exciting vibration.

Free vibration is the frequency that a body will naturally tend to vibrate at when it is acted upon by an external excitation force [1]. It is also known as the resonant or harmonic frequency of a system. A body will have a range of natural frequencies that it will tend to vibrate at depending on the intensity of the excitation force. This is known as the harmonics of the system, with each harmonic corresponding to the number of modes of vibration at that frequency. The natural frequency of a part is determined by its material properties, mass and geometry.

The material properties of interest are Poissons Ratio and Young's Modulus, with Poissons Ratio determining the variation ratio of different mode types and Young's Modulus being directly proportional to the natural frequency. The relationship between mass and natural frequency is inversely proportional i.e. as mass increases the natural frequency decreases [4]. A change in geometry of a part can change its natural frequency by changing the stiffness of the part, where natural frequency is directly proportional to stiffness.

Forced vibration occurs due to a periodic force being applied continuously to a system. In a machining environment forced vibrations are induced in the part and fixture system by imbalances in the machine tool or the cutting forces of the tools themselves.

When the forced vibrations occur at a frequency that equals any one of the harmonic frequencies of the system, then self-excited vibration occurs. This causes excessive vibration, also known as resonance, in the system because the input pulses coincide with the peak pulses of the system, causing the system vibration to amplify. Thus when self-excited vibration occurs during machining, due to the cutting parameters matching the resonant frequencies of either the part-fixture system or the cutting-tool-machine system, it is known as chatter.

2.2 Chatter and fixturing

Present manufacturing technologies have focused on chatter reduction through cutting tool technologies whereby end mills and chucks etc. have been redesigned to increase their stiffness, and on selecting machining parameters that provide a suitable material removal rate whereby chatter does not occur. In a FMS that utilizes AFF's to accommodate mass customization, the fixturing system may play a far greater role in the resonance of the system and the generation of chatter, as the AFF may be unable to match the clamping forces of a conventional fixture [5]. In a conventional fixturing system where part is clamped with a very large force, the stiffness of the part exceeds the stiffness of the cutting tool and thus the machining parameters must be adjusted to avoid chatter induced by the tool vibration. In an environment such as the AFF where the cutting tool stiffness exceeds the fixture-part system stiffness, the effect of the fixture on the resonance of the part may be the limiting factor on machining.

3 THE AUTOMATED FLEXIBLE FIXTURE

The AFF was designed by a postgraduate student, Andrew Illedge [6], as a platform to facilitate the increase in part type in a reconfigurable manufacturing environment that supports mass customization, over a conventional manufacturing system. The AFF is a modular, automated fixturing device capable of accommodating a wide variety of part geometry, within a certain size, for machining processes, and reconfiguring itself for different production cycles.

The AFF design, shown below in Figure 1, consists of three main components: a central bed with locating pins that the part sits on, and two clamping arms on either side of the bed supported by three-degree of freedom linear actuators. The clamping arms have an added degree of rotational freedom allowing the clamping forces to be applied downwards towards the bed. The actuation is achieved by a lead screw stepper motor system allowing for the precise positioning of the clamping arms. The central bed comprises of two degrees of rotational freedom, the first allowing it to tilt towards either clamping arm, and the second allowing it to rotate the part about the vertical axis. Both clamping arms have a pin matrix field at their end that moulds around the clamped part and transfers the clamping force to the part.

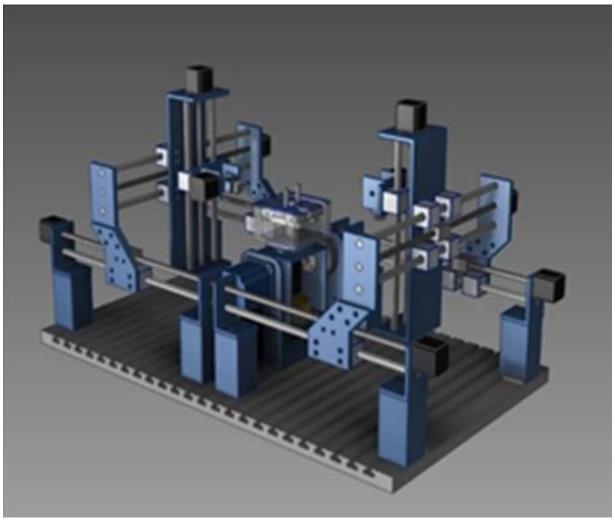


Figure 1 – CAD representation of the Automated Flexible Fixture

4 MODAL SIMULATIONS

The purpose of the simulations was to determine the effect that fixturing position has on the natural frequency of a part, for differing part geometries. Several constraints were imposed on the research in order to keep the research bounded in a large and diverse problem space. These constraints were that only parts with an overall cubic geometry were considered, the starting billet size was set and kept constant for all parts analysed, machining was restricted to the top surface of the part, the circumference of the part along its sides walls was reserved for fixturing, and the material chosen for all parts was Aluminium 6080. The modal analysis was performed using the SolidWorks simulation package.

This paper focuses on a work piece, illustrated in Figure 2, that is symmetric about two axes and whose geometric features consist of two rectangular bores at its centre, with four holes cut at each corner of the part.

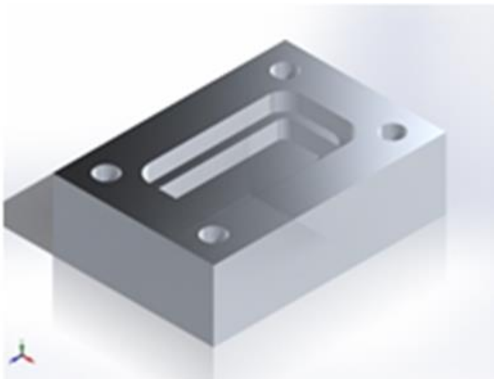


Figure 2 – Cad model in isometric view illustrating the symmetric work piece geometric features

The surfaces for fixturing were limited to the circumference of the part on its side faces. The AFF clamping arms will move in horizontally to hold the part, with the both arms being in line with each other on either side of the part, with the clamping surface of the arms parallel to the side wall surface. Figure 3 shows two examples of fixturing setups. The green arrows represent the clamping surface where the pin matrix will press against the part to fix it in place. There are 6 unique fixturing setups chosen, each correlating by name to the position of the clamping arms relative to the side walls of the part. The positions of the clamping arms for the fixturing setups are either along the length or breadth of the part, at the centre or the side of the parts side wall surface. Figure 3 shows the length-centre clamping setup on top, and the length-side clamping setup on the bottom.

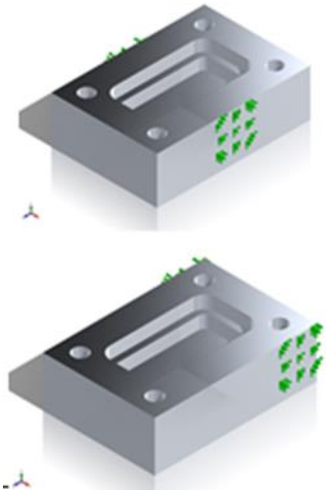


Figure 3 – Illustration of the fixtured surfaces on the workpiece for length-centre and length-side fixturing setups

Each stage of machining for the part is modelled separately to investigate the effect of geometry on the natural frequency of the part. The order of machining operations follows a logical, industry standard approach whereby the larger areas of material removal are machined first, and the holes are bored last, following a natural tool path progression. Figure 4 shows all 6 operations of machining in their order of machining. The first stage is the raw billet shown at the top left of Figure 4. Each other operation is ordered from left to right, top to bottom and are named as operation 1 to operation 6 consecutively.

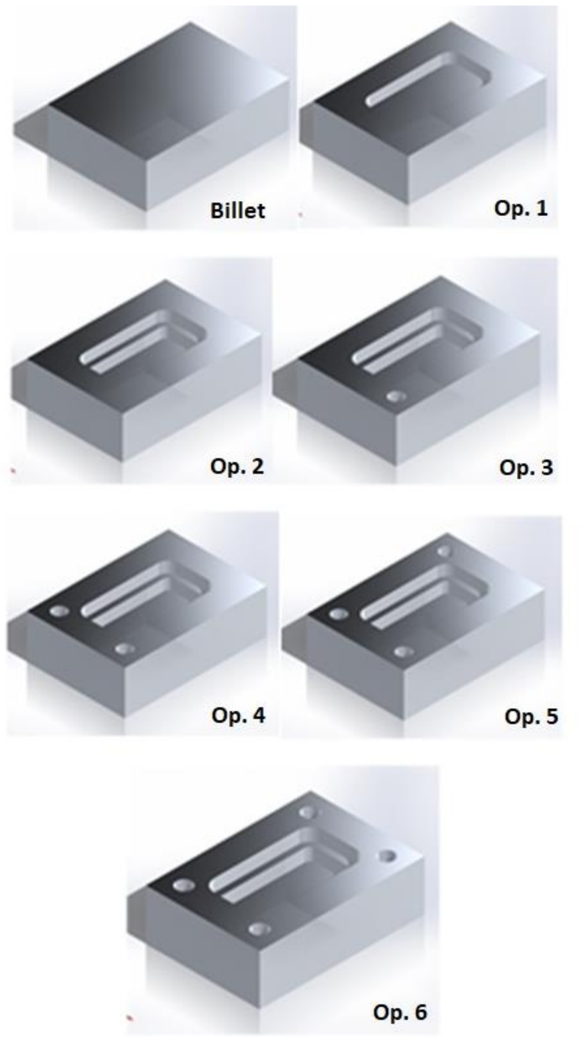


Figure 4 – Representation of the billet, and the machining operations 1 through 6 of the workpiece

5 SIMULATION RESULTS

Due to the large size of the simulation data, the results contained in this section are limited to the first three modes of vibration. The first three modes for each fixturing scenario for the raw billet and operations 1, 2, and 6 are shown in Table 1. The first trend is that for the first mode of vibration the highest frequency is achieved by the length-centre fixturing setup, for all machining operations. For the second and third modes the highest frequency is achieved mainly by the breadth-centre fixturing setup with a few exceptions. Thus when examining operation 2, the best fixturing setup is dependent on the relevant mode; for mode 1 it is the length-centre fixture, for mode 2 it is the breadth-centre fixture, and for mode 3 it is the length-side fixture setup.

	Part 1	Mode 1 (Hz)	Mode 2 (Hz)	Mode 3 (Hz)
Billet	Breadth	8166	12454	15260
	Breadth side	7618	11109	13232
	Length centre	10586	12217	12391
	Length side	6925	10035	13022
Op. 1	Breadth centre	8231	11505	13589
	Breadth side	7389	9670	12635
	Length centre	10283	10508	11187
	Length side	6048	9368	12740
Op. 2	Breadth centre	8150	10159	12413
	Breadth side	6972	8509	11677
	Length centre	9616	9753	10527
	Length side	5469	8530	12430
Op. 6	Breadth centre	8009	9971	12607
	Breadth side	6913	8711	11212
	Length centre	9954	9990	10740
	Length side	5609	8683	12541

Table 1 - Modal Frequencies for Part 1 for all Fixturing Setups

To have a clearer representation of the effect of material removal a graph of frequency versus machining operation for both length-centre and breadth-side fixturing have been plotted. The graph below in Figure 5 plots the first mode of vibration of part 1 against the stages of machining for the length-centre fixture setup. Point 0 on the x-axis refers to operation 0 which is the uncut billet. Point 1 through to 6 on the x-axis refers to operation 1 to operation 6 consecutively. The trend shows that as material is removed by operations 1 and 2 the natural frequencies decreases. As material is removed for the holes in operations 3 to 6, the natural frequency increases.

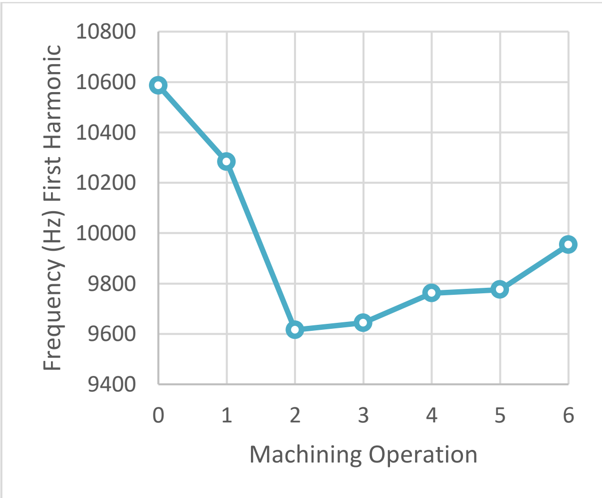


Figure 5 - Graph of frequency vs machining operation for length-centre fixturing

The graph in Figure 6 plots the first mode of vibration of part 1 against the stages of machining for the breadth-centre fixture setup. As in the graph in figure 5, point 0 on the x-axis refers to operation 0, followed by operations 1 through 6 consecutively. The trend from the graph reveals that as material is removed by operation 1 the natural frequency increases, whereas for operation 2 and the holes in operations 3 to 6, material removal *resulted i*n a decrease in natural frequency.

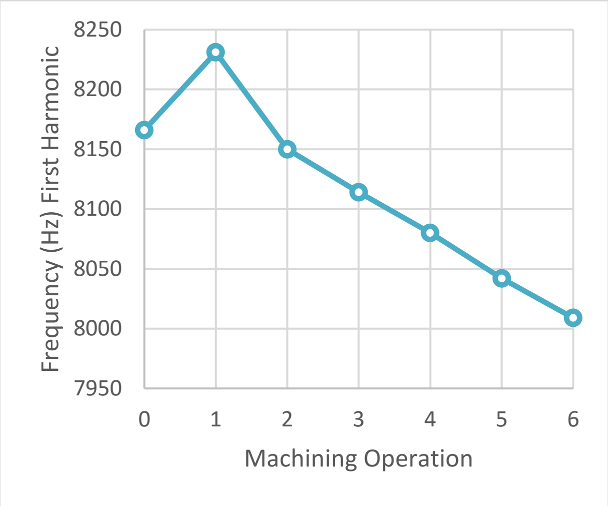


Figure 6 - Graph of frequency vs machining operation for breadth-centre fixturing

The general trends for vibrational response are that the harmonic response of a part is directly proportional to its stiffness, and inversely proportional to its mass. Thus by increasing the stiffness of a part or decreasing its mass, the natural frequency of the part will increase, which in turn may decrease the threshold at which chatter occurs.

Hence for the graph in Figure 5, showing the first modal response of the stages of machining for length-centre fixturing of part 1, the effect of material removal on mass and stiffness can be used to explain the trends of the curve. If the effect of mass reduction is greater than the decrease in stiffness of the part then the material removal will increase the natural frequency of the system. The effect of machining operations 3 to 6 in increasing the natural frequency can be attributed to this, as the material removal for the circumferential holes may be sufficient to increase the resonant frequency, but does not reduce the stiffness of the part significantly enough to decrease the resonance.

For operations 1 and 2, the central countersunk slot, the decrease in natural frequency can be attributed to the decrease in stiffness of the part being greater than the effect of reducing its mass. For the graph in Figure 6, where the part is clamped with the breadth-centre setup, the trend is opposite to that in Figure 5, and similar observations can be made regarding the effect of stiffness and mass on the resonance. This has implications for machining sequence, where dependant on the fixturing setup, the order of machining operations has an effect on the natural frequency of the system. In the case shown in Figure 5, the effect of cutting the holes last could result in a worse vibrational response, during their machining, than if they were done first.

The position of the fixture surface relative to the areas on the part where the modal peaks would natural appear may also have an effect on the frequency response. If the fixture surface was positioned at the peak of a mode it would dampen the vibration of the part to a greater degree than if it was near an anti-node. This may also play a role in the characteristics seen in the simulations.

6 EXPERIMENTAL TESTING RESULTS

The physical testing was performed to validate the trends discovered from the simulation results and to determine the actual dampening of the system. Three test pieces were machined for part 1, those being the first two operations, op.1 and op.2, and the last operation op.6. The AFF clamping arms were equipped with a load cell on each to enable a repeatable clamping force to be applied for all fixturing setups. The clamping force was kept constant at 200N exerted by the clamping arms. Testing was done by striking the fixtured part by an impact hammer, and measuring the response with an accelerometer.

The acceleration data was transposed to the frequency domain and plotted in graph form. The graph in Figure 7 shows the modal range of operation 1 for breadth-centre fixturing. The first amplitude corresponds to the impact of the hammer and is the first modal harmonic, the second and third amplitudes are the second and third harmonics respectively. The hammer impact did not have sufficient energy to reveal any higher order harmonics.

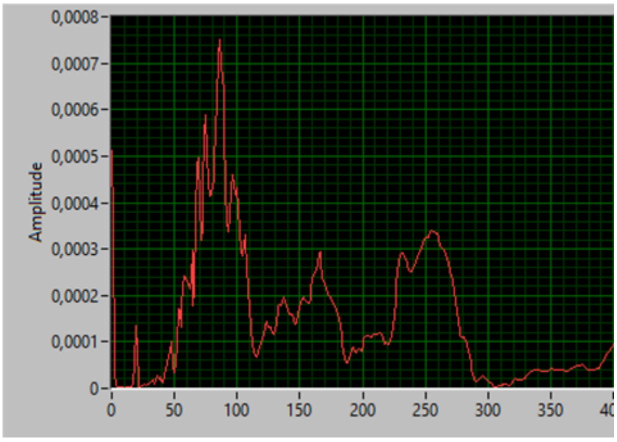


Figure 7 - Graph of natural frequencies of Part 1 operation 1 for breadth-centre fixturing

The graph in Figure 8 plots the frequency domain for the knock test of part 1 operation 1 for length-side fixturing. The First mode occurs at 40Hz, the second at 85Hz and the third at 120Hz. After the third mode there is not the expected diminishing of vibration energy, but rather a continuance of frequency data. This may be attributed to the cantilevered characteristic of the length-side fixturing setup allowing the part to continue to resonate at a lower amplitude.

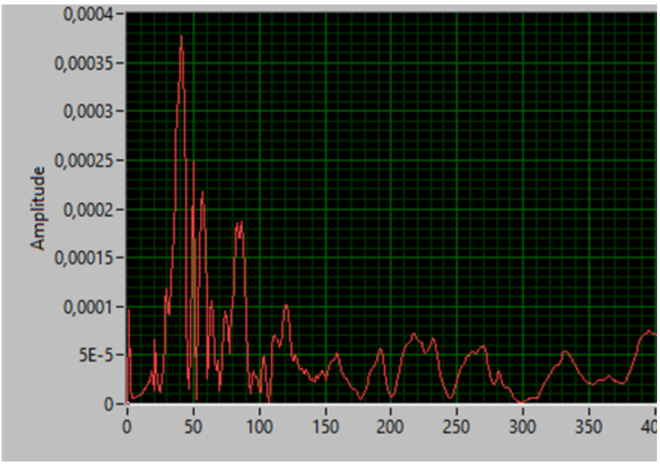


Figure 8 - Graph of natural frequencies of Part 1 operation 1 for length-side fixturing

Table 2 shows all the modal data for part 1 for each operation under each fixturing position. It is clear that the breadth-centre fixturing setup gives the best harmonic response, with the highest frequency for each operation and mode occurring for this setup. The least favourable setup is the length-side setup, which has the lowest harmonic frequencies of all the fixture setups.

	Operation	Mode 1	Mode 2	Mode 3
Breadth	Op. 1	85	165	250
	Op. 2	110	240	
	Op. 6	100	220	320
Breadth-Side	Op. 1	60	120	230
	Op. 2	100	150	230
	Op. 6	100	210	295
Length-Centre	Op. 1	48	115	168
	Op. 2	60	110	175
	Op. 6	70	140	230
Length-Side	Op. 1	40	85	120
	Op. 2	40	60	100
	Op. 6	25	70	110

Table 2 - Natural Frequencies for each Fixturing Scenario for Part 1

The Graphs in Figure 7 and 8 show a large amount of noise and interference in the frequency response. This can be attributed to the fixture itself, which due to its complexity and certain imperfections in fittings, may have inputted noise to the signal. The direction of the fixturing arms and thus clamping forces may have allowed some movement between the bed and the part, which would result in systematic errors in the measured signal.

The data in Table 2 shows that the breadth-centre fixturing setup produced the highest modal frequencies for all 3 operations and all 3 modes. In the breadth-centre fixturing setup the arms are positioned along the shortest side of the part, against the centre of the side face surface. This fixture setup is the one that would most logically be assumed the best approach, as it would place the greatest volume of material between the arms and thus would be assumed to be the stiffest. The correlation between the volume of material held between the clamping arms and the modal frequencies of each operation holds true for all four clamping setups.

When comparing the simulation results to the experimental testing data the most striking differences are the magnitudes of the natural frequencies. The Simulation modal frequencies are far higher than those recorded in the physical testing. This is because of the characteristics assumed for the fixture surface in the frequency simulations. Because the dampening characteristics of the AFF system were unknown the fixture surfaces were simulated with fixed geometry constraints. These constraints did not include any dampening effects, and revealed the highest modal frequecy for the workpiece that is possible for the differing fixture setups. Because the actual AFF has a high degree of dampening the experimental results revealed far lower natural frequencies.

The effects of geometry in relation to fixturing setup that were discovered through the simulation results are not observable in the experimental testing results. This can be attributed to the low clamping forces, due to load cells constraints, the stiffness of the system was insufficient to reveal those trends, and the effects of mass reduction on the parts resonance dominated the system. The only exception to the increase in natural frequencies for all operations and fixturing setups is the length-side fixture setup which mainly produced a decrease in natural frequency for all modes and had the lowest frequency response.

There are several possible strategies for how a theoretical chatter reduction control system might approach optimizing fixturing. The First is to determine the fixturing setup that would provide the best average response over all machining operations. The second strategy would involve adjusting the fixturing setup for each operation to provide the best response for that specific operation.

7 CONCLUSION

This research focused on establishing the relationship between fixturing and chatter as the basis for introducing a layer of intelligence to a proposed AFF. The relationship between fixturing and chatter was investigated through modelling of the modal characteristics of parts clamped by the AFF in different fixturing setups. The simulation results revealed a significant difference in frequency response for differing fixture setups. The experimental testing of the frequency response was unable to validate all the simulation results, because of the stiffness limitations of the AFF, but corroborated the general trends of the frequency response, demonstrating that fixturing has a significant effect on the resonance of a part.

Future work will involve measuring the frequency response during machining to corroborate the relationships between fixturing, natural frequency and machining quality discussed in this research. The frequency response from the machining tests will then be used in the development of an autonomous chatter control system, utilizing fixturing strategies to minimize vibration in the system, allowing for higher rates of chatter free machining, and improving the efficiency of a flexible manufacturing system.

8 ACKNOWLEDGMENTS

I would like to acknowledge the support and guidance of Dr. Anthony Walker and Prof. Glen Bright in my research, as well as the NRF Blue Sky Foundation for their financial backing of the Factories of the Future research, and the University of Kwazulu-Natal.

9 REFERENCES

[1] Harper, R., (2004), *Chatter Myths: Pieces of the Puzzle in Maximized Machining*, Available: http://www.moldmakingtechnology.com/articles/chatter-myths-pieces-of-the-puzzle-in-maximized-machining. Last accessed 15th Jan 2015.

[2] Albert, M., (2007), *Workholding that holds off chatter*, Available: http://www.mmsonline.com/articles/workholding-that-holds-off-chatter. Last accessed 15th Jan 2015.

[3] Wang, H., Rong, Y., Li, H. and Shaun, 2010, *Computer aided fixture design: Recent research and trend*, Computer-Aided Design, 42/12, 1085-1094

[4] Kennametal, (2012), *Reducing Vibration and Chatter in End Milling*, Available: http://www.kennametal.com/content/dam/kennametal/kennametal/common/Resources/Technical%20Tips/Solid%20Carbide%20Milling/TechTip_032_ChatterVibration_EndMilling.pdf. Last accessed 15th Jan 2015.

[5] H. mcCallion, "Vibration of linear Mechanical Systems" longman Group limiter, 1973

[6] Z. M. Bi and W. J. Zhang, "Flexible Fixture Design and Automation: Review, Issues and Future Direction" *int. j. Prod. Res.,* vol. 39, no. 13, pp. 2867-2894, 2001.

[7] Illedge, A., 2014, *An Automated Flexible Fixturing System for Mass Customization*, International Conferences on Intelligent Systems and Agents

10 BIOGRAPHY

Erlank Slabbert obtained his bachelors degree in mechanical Engineering at the Univeristy of Kwa-Zulu natal. He is currently pursuing an MSc.Eng in robotic manufactuign systems with the Manufacturing and Robotics Research Group.

Dr. Anthony Walker holds a MSc.Eng and PhD degrees in Mechanical Engineering from the University of Kwa-Zulu Natal.

Professor Glen Bright holds a MSc.Eng and PhD degrees in Mechanical Engineering from the University of Kwa-Zulu Natal. He is currently the Dean and head of the school of Engineering at UKZN.

21st Century Knowledge Dissemination: Adaptive Games to Address the Skills Gap

T. von Leipzig[1], K. H. von Leipzig[1], V. Hummel[2]
[1]Department of Industrial Engineering, University of Stellenbosch, South Africa
[2]Logistics Management, ESB Business School, Reutlingen University, Germany

Abstract
The 21st Century: An era where emojis and hashtags find their way into every sentence, where taking selfies, live-tweeting and mining bitcoin are the norm, and where Insta-culture dictates what we say and do. This is the era into which the digital native was born. With so many changes in every aspect of our lives, how is it that one of the most influential aspects, our education, has remained unchanged? Our education system not only fails to appeal to today's students, but more importantly, it fails to equip them with the skills required in the 21st Century. It is thus of no surprise that industries feel graduates entering the workplace lack skills in critical thinking, problem solving and self-directed learning. AI, machine learning and big data: Tools and mechanisms we so eagerly incorporate to create smart factories yet are hesitant to use elsewhere. Gamification and games have shown great results in education and training; with most research suggesting a stronger focus on personalization and adaptation. When combined with analytics and machine learning, the potential of games is yet to be realized. A real-time adaptive game would not only always present an appropriate degree of challenge for the individual but would allow for a shift in focus from the recitation of facts, to the application of information filtered to solve the particular problem at hand. South Africa, a country faced with a severe skills gap, could benefit greatly from games. If used correctly, they may just offer a desperately needed contribution toward equipping both current and future employees with the skills needed to survive in the 21st Century. This paper explores the feasibility of using such games for enhanced knowledge dissemination and the upskilling of the workforce.

Keywords
Adaptive games, 21st Century learning, skills gap

1 INTRODUCTION

In South Africa, the overall unemployment rate is 26.7%, whilst youth (aged 15-34) unemployment reached 38.2% in the first quarter of 2018. The more frightening figure is the number of youths that are not in employment, education or training (NEET). The current NEET rate is 34.2%, implying that a third of the country's youth is entirely disengaged with the labour market. [1]

These figures suggest that "South African young people face extreme difficulties engaging with the labour market", which is, in part, a result of "lack of experience" [1]. The unemployment rate of youth is high across all education levels, with the unemployment rate of graduates at 33.5%.

This problem is not exclusive to South Africa. Experts across industries state that graduates lack the skills required in the field [2], whilst some argue that today's youth are simply unemployable due to the mismatch of skills required in industry and those possessed by graduates [3]. Technological advances have changed nearly every aspect of our lives [4, 5], yet education has to a large degree remained the same, and is ineffective in preparing individuals for the dynamic workplace of the 21st century [3, 6, 7].

Furthermore, Heikkila [8] highlights that with the introduction of smart technologies, Industry 4.0 and the data explosion across industries "half of the jobs today won't exist tomorrow and … half of the jobs tomorrow don't exist yet today". The question for 21st century learning is thus: "How do we educate the unemployable for a future that does not yet exist?" [9, p. 5]. This paper explores the shortcomings of current education and training systems and proposes the use of adaptive games as an alternate approach that appeals to today's youth and aims to better equip them with the skills required in the 21st century.

2 LEARNING

2.1 Bloom's Taxonomy

One of the most recognisable taxonomies in education is that of Benjamin Bloom [10]. Bloom's Taxonomy, developed in 1956, defines six levels of cognition each with increasing complexity. Krathwohl later revised the taxonomy for easier application in educational design and introduced a second dimension to learning [10, p. 215]. The cognitive levels of the revised taxonomy are depicted in Figure 1.

The second dimension categorises knowledge into factual, conceptual, procedural and metacognitive knowledge. Table 1 defines a two-dimension matrix of learning as per the Krathwohl's revised taxonomy [10], with the cell darkness indicating the complexity of learning. The deepest and most complex learning is achieved where the metacognitive dimension and creation dimension meet.

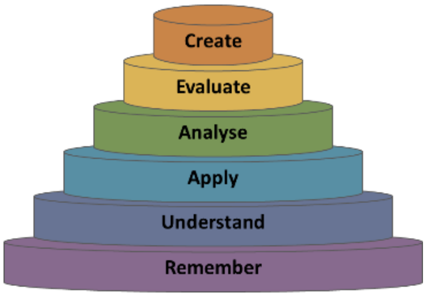


Figure 1 – Revised Taxonomy cognitive levels adapted from [10]

A learning environment in which awareness of how concepts interact is evoked (metacognitive knowledge dimension) and where learners are encouraged to formulate their own ideas based on the concepts (cognitive creation dimension) is thus an environment inducive of the deepest and most complex form of learning.

	Cognitive dimension					
Knowledge dimension	Remember	Understand	Apply	Analyse	Evaluate	Create
Factual						
Conceptual						
Procedural						
Metacognitive						

Table 1 - Classification of learning [10, p. 216]

The relevance of Bloom's Taxonomy in today's digital era has created some debates. An effective learning environment should not only induce metacognitive knowledge creation but should also take into consideration the type of student it aims to teach.

2.2 The digital era

The digital native, first defined by Prensky in 2001, refers to today's youth that are native in the digital language [11]. Prensky argues that "today's students think and process information fundamentally differently" [11, p. 1]. Whilst all generations differ slightly from previous ones, he believes that as opposed to incremental changes, the digital native has resulted from a "singularity – an event which changes things so fundamentally that there is absolutely no going back" [11, p. 1]. Technology has been introduced into almost every aspect of our everyday lives.

Greenfield's neurological studies [12] support Prensky's notion, showing that technologies have changed both the behaviour and way of thinking of the digital native. The digital native, who has grown up using technology and does not know life without it, has been "hardwired" to think digitally [13, p. 464]. In contrast, the digital immigrant who was "not born into the digital world but [has] adopted many or most aspects of the new technology" [11, pp. 1-2], has been "rewired" in part towards the use of technology [13].

Whilst Prenksy [11, pp. 2-4] defines the characteristics of the digital native, Autry and Berge [13] provide contrasting characteristics of the digital immigrant. Where the digital native prefers receiving and processing multiple fast, graphic information simultaneously in a social environment that rewards frequently and provides instant gratification, the digital immigrant prefers processing slow consistent information in a sequential manner. The strong contrasts between the digital native and digital immigrant are summarised in Figure 2.

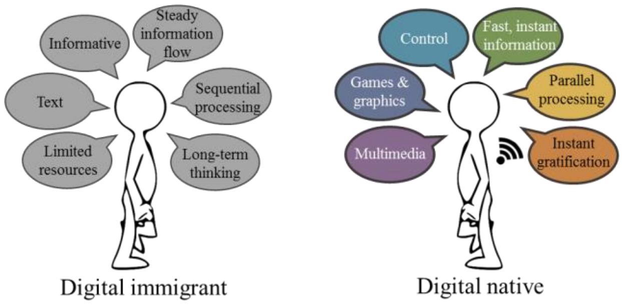


Figure 2 – Digital native vs digital immigrant [9, p. 19]

Prensky argues that "[o]ur students have changed radically [and …] are no longer the people our educational system was designed to teach" [11, p. 1]. Whilst there is some criticism of the relevance of the digital native, it is also important to consider the new demands of industry. The 21st century, with all its innovations and technological advancements, demands what has been termed '21st century skills'.

2.3 21st Century learning

The amount of data in the world has grown exponentially and is, as depicted in Figure 3, expected to continue on this path. "Data has become critical to all aspects of human life" [14, p. 2]. With so much data available at the click of a button, the ability to interpret them and apply them correctly has become far more important than the data themselves [3]. Romero et al. [7, p. 151] define 21st century skills as 'meta skills', including skills such as problem solving, non-linear thinking, social skills, creativity and the ability to learn in its definition.

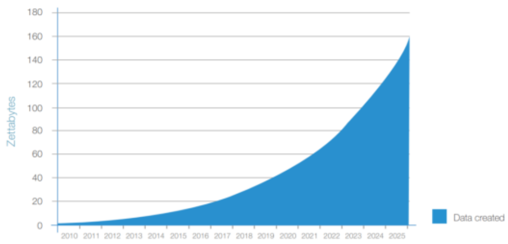


Figure 3 – Annual size of the global datasphere [14, p. 7]

Wagner [15, p. 14] defines a list of what he terms skills in which previously only the elite were proficient yet are now required for survival. These skills, termed the "seven survival skills for the 21st century", are depicted in Figure 4.

Figure 4 – 21st Century survival skills constructed from [15, p. 16] in [9, p. 21]

However important for survival, current education systems fail to equip today's students with these 21st century skills [7]. In the following section, the various shortcomings of current education are discussed.

2.4 Shortcomings of current education

Not only does current education not appeal to the characteristics of the digital native, the summative assessments used are largely ineffective [16, 17]. They do not allow for rework and neglect the actual learning process. Students are not able to learn from mistakes, nor are they given the opportunity to reflect and construct their own thoughts and applications. Current education furthermore does not equip students with the skills required to excel in the 21st century as discussed above [18]. It focuses on the "3 R's – Reading, wRiting and aRithmetic" [7, p. 149], where it should instead teach students what the internet cannot provide, namely logical and creative thinking, and how to interpret and make sense of the large amounts of data and information [3]. Traditional learning theories, which were developed before the introduction of technology, are grounded in the belief that knowledge is a destination. In contrast, Siemens [4] argues that learning is not simply the consumption of content, nor is knowledge an attainable state. Many experts support the extension of education to "connectivism", a learning theory that focuses on interpretation, internalisation and construction rather than consumption [4, 5, 19], and the "navigation through the ocean of knowledge" [5]. In South Africa specifically, but many other countries too, access to education poses a problem [20]. Not only is there insufficient space, but the ever-increasing cost of quality education makes it difficult to provide fair access to quality education for all. Finally, current education follows a 'one size fits all' approach, where all students follow identical course structures and are expected to progress at an identical pace, where it should rather consider individualism and personalised engagement [21, 22, 23]. This is especially important in an African context, where a large diversity of culture, educational backgrounds, literacy levels and language render a 'one size fits all' education ineffective [24].

These various problems could, at least in part, be addressed by technology [9, pp. 19-24]. The following section considers educational technology (EdTech) in South Africa.

2.5 EdTech

EdTech, or the incorporation of Information Communication Technology (ICT) into education holds the promise of increasing educational efficiencies, address educational shortcomings, provide more active learning, increase motivation and engagement, and, in turn, improve learning outcomes [25, 26, 27, 28]. In an analysis of the Global Information Technology Report [29], which benchmarks and reports on the ICT readiness and usage of countries, it is suggested that the implementation of technology, more specifically on mobile platforms, is a feasible option for expanding the access to quality higher education to a larger portion of the South African population [9, pp. 24-27].

To date however, ICT has, to a large extent, failed [30], with a prominent reason therefore being the incorrect application of technology [27]. Figure 5 depicts a continuum of ICT integration complexity [27, pp. 73-75]. In the emerging state, technology is purchased and the possibilities of its application in teaching explored. The applying stage refers to instances where technology is slowly incorporated yet curricula is still dominated by traditional lectures. In the infusing stage, technology is embedded throughout the curricula and finally, in the transforming stage, it is used to entirely rethink and transform the curricula into an innovative learning environment. In most instances where technology is incorporated into education, it is used to simply replace current content with digital content and make processes more efficient yet fails to actually rethink or transform the processes themselves [30, 31].

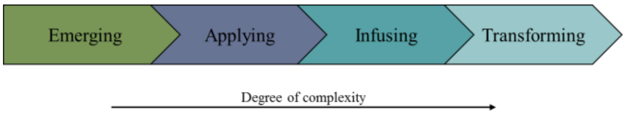


Figure 5 – Four stage continuum for ICT integration constructed from [27]

One form of EdTech that is grounded in the premise of transforming learning, is games.

3 GAMES FOR ENHANCED LEARNING

Educational games (edugames) hold the potential to transform current education and address the shortcomings of current education. Games are inherently motivating and engaging and naturally appeal to the youth of today. By providing continuous, real-time feedback, edugames allow players to reflect and reconstruct and apply ideas.

Games furthermore emphasize problem solving and intrinsically motivate students to make sense of, and form connections between information in order to succeed in the game [22, p. 170]. They thus offer not only a fun and engaging learning environment, but one in which awareness of interaction of concepts, and the construction of ideas is evoked, thus corresponding to the highest two-dimensional classification of learning as per **Error! Reference source not found.**. To assess the feasibility thereof in practice, a prototype is developed.

4 SIGHT: A TESTED PROTOTYPE

To increase the long-term transfer of skills, a game was designed and developed to enhance a seminar on shop floor management.

4.1 Game design

The target audience of the seminar, and thus the game, ranges from professional and managerial staff to shop floor employees [32, p. 46]. Since this target audience includes a great diversity in skills, education levels, and professions, as well as the possibility of large cultural and age differences, the game should include as many different motivators as possible.

4.1.1 Design motivators

The Uses and Gratifications (U&G) theory suggests that any player's motivation for playing a game and being engaged by it, can be described by one or more of seven motivations, namely control, social interaction, chellenge, competition, diversion, fantasu and interest.

Since the social interaction motivator appeals to almost every person in some way, it is selected as a design element. Many of today's shop floor employees are digital natives, and control is one of the main characteristics of the digital native. Thus it is included in the game design. Furthermore, the challenge, competition and diversion motivators are known as directional motivators [33], thus providing internal momentum and motivation to pursue a goal. These motivators will therefore appeal to the intrinsic motivation of the attendees, and are also selected. Table 2 provides an overview of how each motivator is incorporated intot the game's design.

Target motivator	**Representation in the design**
Control	The player is able to customise his/her game through the purchasing of items in the game store, as well as choosing his/her own avatar
	The player has the choice to make use of the MKO (use the help function) as required
	The player has the choice to answer selected shop floor questions to receive game points
Challenge	The levels increase in difficulty, introducing the player to a new concept as soon as one is mastered
Competition	Players compete against one another during the seminar, with the possibility to add a leaderboard where players can compare their scores against others
Diversion	The designed game provides an entirely abstract environment from the production and factory setup the players are accustomed to
	The game provides a new and abstract method of training as opposed to the traditional learning factory seminar
Social interaction	Players can discuss the game during gameplay, as well as take part in the feedback session on the day 2 of the seminar

Table 2 - Motivators incorporated into the game design

4.1.2 Game contents

The seminar for which the game is designed covers 6 shop floor management theories and concepts, namely key performance indicators, transparency, visualisation, customer-orientation and waste elimination. These concepts are thus translated into game concepts.

4.1.3 High level design

Each level in the game is designed to relate to a shop floor management concept. For each core concept, a game mechanic is included in the design which best suits and portrays the concept. The levels and shop floor concepts are ordered in such a manner, that it is logical from a gameplay perspective where each level is more difficult than the previous. Table 3 provides an excerpt of the translation of theoretical concepts to game mechanics and game level design.

Shop floor management concept	Corresponding game mechanic/ element in level	Description of game level
Customer-orientation	Level goal	• Player is given no clear goal of which coins or how many thereof to collect • Multiple types of coins lead to confusion, since not all types must be collected
Visualisation (Visual working aids, andon boards)	Visual item basket	• Player must deliver a large amount of varying coins, losing track of what has already been collected • The addition of a visual display of the player's basket aids in the successful completion
	Visual lighting on gates	• Some coins are blocked off by walls of which only some will act as gates and open • The addition of a visual colour display on the walls signals to the player which walls are gates that will open

Table 3 - Excerpt of game level derivation from core shop floor management concepts

4.1.4 Detailed design

In the detailed design stage of the game, each screen and level was sketched, afterwhich wireframes and later, mockups, were created for each. Parameters and rules were deduced for each level of the game, and in turn, translated into mathematical equations used in the development, with an excerpt depicted in **Error! Reference source not found.**.

Rules and equations		
Level 1	$numCoins_1 \geq \sum_i (goalCoins_{i,1})$	Must at least be the coins I need to collect
	$goalCoins_{i,1}$ = exist	Must be at least 1 access point (pathway could still be blocked)
	For each $goalCoins_{i,1}$ j $: \sum_x (adjacentWall_{x,j}) \leq 3$	Must at least be able to follow shortest path and collect goalCoins
	$totTime_2 \geq goalTime_2$	

Table 4 - Excerpt of rules and equations for each level's design

4.2 Game development

In South Africa, poor broadband and high costs limit online gaming, whilst mobile gaming holds a bright future. A study in 2013 showed that despite the high poverty rate, over 75% of South Africans classified into the low-income group and aged 15 years or more, own a mobile phone [34]. To best serve the purpose of the seminar, as well as allow participants the flexibility to play anywhere, anytime, the game was developed to be played on both computers and mobile devices. The game was developed using the Unity 3D game engine [35]. For initial testing, a prototype of the first four levels was developed.

When launched, the game begins with a start screen as shown in Figure 6 (a). The player is able to select an avatar and is given an explanation of the game and its controls. In each level, the player finds himself / herself in a 3D maze filled with coins. The playing field of Level 1 is depicted in Figure 6 (b). Throughout each level, the shop floor concept is intrinsically portrayed through the level objective and lesson. In addition, theoretical knowledge is conveyed through question boxes, which the player can break to answer questions for bonus points as shown in Figure 6 (c). As shown in Figure 6 (d), after successful completion of a level, the player is given feedback relating the level objective and lesson to the relevant shop floor management concept.

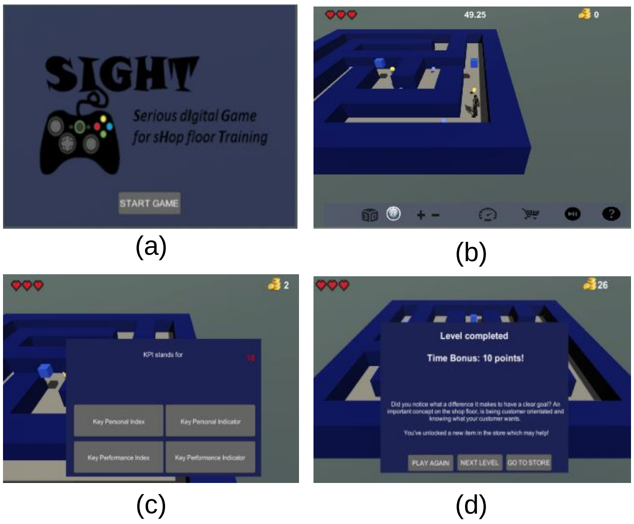
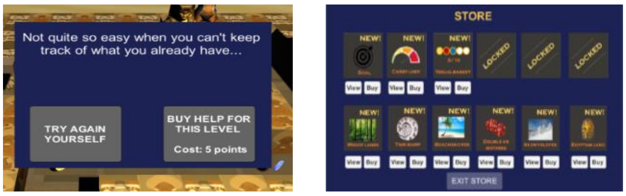


Figure 6 – Start screen (a), playing field (b), question in a question box (c) and level completion feedback (d)

A player scores points in the form of coins by completing each level. Bonus points can be gained by correctly answering theoretical questions, whilst coins are spent to receive additional help in a level as shown in Figure 7 (a). The player can spend these coins in the game store by purchasing items relating to the shop floor concepts (such as level aims and WIP baskets), items to customise their game (such as environment skins), and finally bonus items to double points or get additional time on the clock. The store is depicted in Figure 7 (b).



(a) (b)

Figure 7 – The help function with general tip (a) and game store (b)

The final level of the prototype represents the concept of visualisation on the shop floor. Some coins are surrounded by walls, with only some being penetrable. The player initially has no way of knowing which walls are penetrable and which are not but can purchase a visual aid which adds a green light to all walls through which he / she can move. Figure 8 depicts Level 4 after the visual aid is added, highlighting the importance of visualisation.

Figure 8 – Level 4 with visual help purchased

To cater for the largest possible audience, the player can choose to play the game in 3D, or in 2D with a fixed top-down view. The player furthermore has the option of adjusting the speed of the game, and is able to reduce the speed up to 25%, to ensure that lack of gaming experience and speed do not hinder the learning process in any way.

4.3 Implementation and results

The developed game was validated in 3 parts, namely by industry experts, a group of university lecturers as well as a group of university students.

4.3.1 Industry validation

The industry validation included a test group at a Cape Town-based steel manafacturing company as well as a group of experts in shop floor management and gaming. 8 Employees from the steel company were given the opportunity to test and play the game themselves. 4 Industry experts were also asked to play the game and provide feedback. The focus of this vaildation was to assess the general acceptance and support

4.3.2 Lecturer validation

To assess the acceptance of games as a teaching and training tool, 4 lecturers from 2 universities were asked to play the game and provide feedback regarding the educational support of games.

4.3.3 Student validation

Finally, a group of 10 university students were used to assess the actual learning spectrum of the implementation by means of an experiment. All students received the same pre-test, after which 5 students were randomly selected to play the game whilst the remaining 5 (the control group) received theoretical notes and a lecture relating to shop floor management. All students were then asked to complete a post-test to allow the analysis of the results.

4.3.4 Results

An unexpected yet positive outcome of the industry testing was that the attendees from the steel company, although not instructed to do so, opted to play the game in groups. Unintentionally, the game led to teamwork amongst players, an element often saught out in education. Their general feedback was also positive, with all participants agreeing that games can enhance the learning process of traditional seminars and can have a sustainable long-term effect on learning.

The responses given by the 4 industry experts were predominantly positive. All agreed that games can provide a valuable training tool for employees and that they can have a sustainable long-term effect on learning. Most believed that games can effectively enhance learning and would promote the use thereof in their own industries. Only 1 expert said he would not promote the use of games in his industry, however explained that his industry is predominantly based on programming and that this would be difficult to teach via a game.

The ratings provided by educators and students were positive. All educators and students felt that games can enhance the learning process and have a sustainable long-term effect on learning. All students also felt that playing games would be more engaging that a lecture alone.

Using the pre-test and post-test scores of the students, the effectiveness of the game was assessed. The scores of students who played the game increased significantly more than those of the control group, with the average increase in scores being 174.6% and 47.8% respectively.

4.4 Main findings and recommendations

An interesting finding that was visible in the different validation groups, is that people differ in strategies and motivations during gameplay. These differences were very prominent in the validation and were first observed at the steel company. Some participants who realised that the question boxes were not crucial to complete the level, ignored all question boxes and focussed only on collecting the coins required to get to the next level. Other participants were not as motivated to reach the next level, but were instead motivated by answering the questions in the question boxes correctly, and would spend all their time running to get to the next box, ignoring all coins on the playing field.

In the experimental validation with the students, the largest improvements were made by students with a low prior knowledge on shop floor management, suggesting that games can indeed increase the transfer of skills even in participants who already have a knowledge of the topic, but especially in participants new to the topics. The various stages of the validation suggest that industry employees and managers support the concept of gamification, lecturers strongly believe that it can improve the learning experience, and students themselves enjoy playing games.

The most influencial finding was established when participants were asked to state their favourite aspect of the game, as well as provide suggestions for improvement. Not only did participants follow different strategies, but they listed entirely different aspects as their favourite. Whilst some favoured the suspense of figuring out what the level objective was, others were more interested in the question boxes, whilst others yet enjoyed the interaction between players. Their suggestions for improving the game also varied greatly, with some suggesting more theoretical questions and introducing difficulty levels into the questions, others wanting the option of multi-player and others suggesting the addition of more randomness to a level when attempted multiple times to increase uncertainty and suspense.

From these responses, it became evident that people differ greatly and thus also respond differenty even when presented with the same game and the same pedagogical content. To be truly effective, these differences should be catered for in an edugame.

5 A FRAMEWORK FOR ADAPTIVE GAMES

A framework for the design and development of personalised and adaptive edugames was developed based on the relevant theories of adaptivity, the concepts of analytics and AI, as well as the lessons learnt and recommendations of previously-implemented edugames and case studies [9]. The framework is developed with the intention of creating edugames that may enhance learning, teaching and assessment.

5.1 High level framework

The framework for personalised, adaptive edugames consists of the following 5 stages performed in an iterative process [9, p. 95]:

- initialisation: The pedagogical, assessment and adaptivity requirements are defined and the game is designed;
- development: The game is developed and tested, with any feedback fed back into the design;
- gameplay: The game's implementation includes 3 dynamic events namely the assessment of learning, the assessment for learning and the adaptation of content, with any feedback fed back into the design;
- adjustment period: All participants (students and educators) are given sufficient time to adjust to the game and its application; and
- post implementation: The edugame's effectiveness is assessed and feedback fed back into the design of the next iteration of the framework.

Figure 9 depicts the high-level framework.

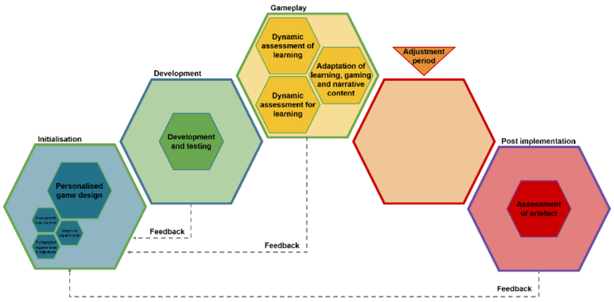


Figure 9 - A generic high-level framework for the development of personalised, adaptive edugames *[9, p. 95]*

5.2 Framework application

Each stage of the framework is expanded and dealt with in detail, explaining the steps included, as well as who should be involved and specific inputs required for each step [9, pp. 95-108]. Figure 10 depicts the detail of Stage 1, namely the personalised game design.

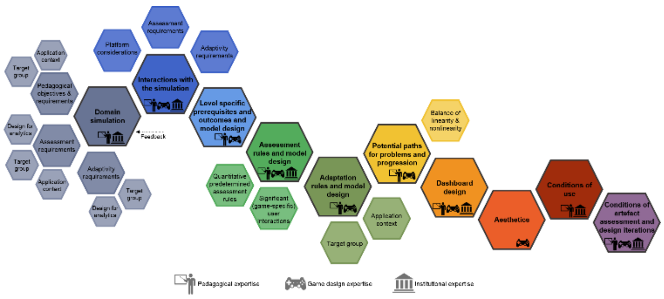


Figure 10 – Stage 1 (personalised game design) in the development of personalised, adaptive edugames *[9, p. 96]*

Furthermore, a checklist is provided for each stage of the framework to ensure correct interpretation of the framework [9, pp. 199-205]. As shown in the checklist excerpt in Figure 11, each step is broken down into sub-steps or questions, along with specific aspects to consider and the parties responsible for the step.

CHECKLIST STAGE 1: PERSONALISED GAME DESIGN		Responsible?
1. **Which domain do we want to teach in and how can we break the concepts down systematically?**		
Which concepts do we want to teach?		☐
Which learning objectives do we have?		☐
What are the required learning outcomes?		☐
	CONSIDER:	
	Who do we want to convey the content to?	☐
	Where will the game be applied?	☐

Figure 11 – Excerpt of the checklist for Stage 1: Personalised game design from *[9, p. 199]*

The framework aims to provide a practical and holistic guide for designing and developing personalised, adaptive edugames to enhance learning, teaching and assessment.

6 CONCLUSIONS

To assess the practicality and feasibility of using adaptive edugames for enhanced learning, teaching and assessment, the developed framework is to be validated by various experts, before applying the framework to design and develop an adaptive edugame. This adaptive edugame will be developed for and tested in the Industrial Engineering faculty at Stellenbosch University.

It is the hope of the authors, that adaptive edugames may just provide a desperately needed contribution toward addressing educational shortcomings and equipping both current and future employees with the skills needed to survive in the 21st Century and prepare them for the future which does not yet exist.

7 REFERENCES

[1] Stats SA, “Youth unemployment still high in Q1 2018,” Statistics South Africa, 15 May 2018. [Online]. Available: http://www.statssa.gov.za/?p=11129. [Accessed 14 October 2018].

[2] S. B. Malan, “The development of an integrated Problem-Based-Learning (PBL) approach in a post-matriculation programme at the University of Stellenbosch,” Cape Town, 2008.

[3] B. Gofron, “School in the era of the internet,” *Educ. Educ.,* vol. 17, no. 1, pp. 171-180, 2014.

[4] G. Siemens, “Connectivism: A learning theory for the digital age,” *International Journal of Instructional Technology and Distance Learning,* vol. 2, no. 1, pp. 3-10, 2004.

[5] T. H. Brown, “Beyond constructivism: navigationism in the knowledge era,” *On the Horizon,* vol. 14, no. 3, pp. 108-120, 2006.

[6] T. Trowbridge, “Learning new skills to spark creativity,” *Future of learning,* vol. 3, no. 0406, p. 7, 28 September 2016.

[7] M. Romero, M. Usart and M. Ott, “Can serious games contribute to developing and sustaining 21st century skills?,” *Games and Culture,* vol. 10, no. 2, pp. 148-177, 2015.

[8] A. Heikkila, “Jobs that don't exist yet: How to prepare for the future of work,” 6 March 2016. [Online]. Available: http://www.business2community.com/human-resources/jobs-dont-exist-yet-prepare-future-work-01484828#4BtrV67ryUC7tekL.97. [Accessed 22 May 2017].

[9] T. von Leipzig, “Real time adaptive, digital games for enhanced learning, teaching and assessment: A framework and tested prototype,” Stellenbosch University, Stellenbosch, 2018.

[10] D. R. Krathwohl, “A revision of Bloom's taxonomy: An overview,” *Theory into practice,* vol. 41, no. 4, pp. 212-218, 2002.

[11] M. Prensky, “Digital natives, digital immigrants,” *On the Horizon,* vol. 9, no. 5, pp. 1-6, 2001.

[12] S. Greenfield, Mind change: How digital technologies are leaving their mark in our brains, London: Penguin Random House, 2014.

[13] A. J. Autry and Z. Berge, “Digital natives and digital immigrants: getting to know each other,” *Industrial and Commercial Training,* vol. 43, no. 7, pp. 460-466, 2011.

[14] IDC, “Data Age 2025,” Seagate, 2017.

[15] T. Wagner, The global achievement gap: Why even our best schools don't teach the new survival skills out children need - and what we can do about it, Philadelphia: Basic Books, 2014.

[16] S. B. Malan, “Third-year student's perceptions of the use of ICT at a teacher training college in Namibia,” Stellenbosch, 2011.

[17] M. Worsley and P. Blikstein, “Towards the development of multimodel action based assessment,” in *LAK '13*, Leuven, Belgium, 2013.

[18] Y. S. Chee, S. Mehrotra and J. C. Ong, “Facilitating dialog in the game-based learning classroom: Teacher challenges reconstructing professional identity,” *Digital Culture & Education,* vol. 6, no. 4, pp. 298-316, 2014.

[19] G. Siemens, “Connectivism: Learning as network-creation,” 2005.

[20] Republic of South Africa, “White paper for post-school education and training: Building an expanded, effective and integrated post-school system,” Department of Higher

Education and Training, Pretoria, 2013.

[21] H. Imran, M. Belghis-Zadeh, T.-W. Chang, Kinshuk and S. Graf, "PLORS: A personalised learning object recommender system," *Vietnam Journal of Computer Science,* vol. 3, no. 1, pp. 3-13, 2016.

[22] J. Hamari, D. J. Shernoff, E. Rowe, B. Coller, J. Asbell-Clarke and T. Edwards, "Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning," *Computers in Human Behaviour,* vol. 54, no. 1, pp. 170-179, 2016.

[23] K. Squire, "Video games in education," *International journal of intelligent simulations and gaming,* vol. 2, no. 1, 2003.

[24] G. Gills, "Is the tablet the whiteboard of the future for corporate training?," *Heavy Chef Review: Education,* vol. 4, no. 1, pp. 29-31, 2016.

[25] N. W. Mlitwa and K. Koranteng, "Integration of ICT into curricula in Western Cape schools: The Activity Theory perspective," *The Journal of Community Informatics,* vol. 9, no. 4, 2013.

[26] European Commission, "New modes of learning and teaching in higher education," Luxembourg, 2014.

[27] M. Wilson-Strydom, J. Thomson and C. Hodgkinson-Williams, "Understanding ICT integration in South African classrooms. Research: Information and Communication Teachnologies," *Perspectives in Education,* vol. 23, no. 4, pp. 71-85, 2005.

[28] A. M. Chainda, "Third-year students' perceptions of the use of ICT at a teacher training college in Namibia," Stellenbosch, 2011.

[29] S. Baller, S. Dutta and B. Lanvin, "The Global Information Technology Report 2016," *World Economic Forum,* pp. 1-307, 2016.

[30] J. Reynolds, L. Caley and R. Mason, "How do people learn?," University of Cambridge Press, Cambridge, 2002.

[31] B. Bilbao-Osorio, S. Dutta and B. Lanvin, "The Global Information Technology Report 2014," *World Economic Forum,* pp. 1-369, 2014.

[32] N. Bachmann, "Development of a business model for consecutive seminars on shopfloor management for training of industry in a learning factory environment," Reutlingen, 2016.

[33] J. Jansz, "Serious gaming," Rotterdam, 2016.

[34] L. Peyper, "Mobile phone usage in South Africa," 22 January 2013. [Online]. Available: http://finweek.com/2013/01/22/mobile-phone-usage-in-sa/. [Accessed 2016].

[35] Unity Technologies, "Unity," 2016. [Online]. Available: https://unity3d.com/unity. [Accessed 22 September 2016].

8 BIOGRAPHY

Tanja von Leipzig holds a B.Eng degree in Industrial Engineering from Stellenbosch University and an M.Sc degree in Operations Management from Reutlingen University. She is currently a Masters student in the Department of Industrial Engineering, with her specific area of focus and research being that of adaptive games for training and education.

Konrad von Leipzig obtained both his B.Eng and M.Eng degrees in Industrial Engineering from Stellenbosch University. He later also completed a B.Comm degree. He has been a lecturer at the Department of Industrial Engineering at Stellenbosch University since 1987. He forms part of the team undertaking the initialisation and realisation of the Learning Factory in Stellenbosch.

Vera Hummel, Prof. Dr.-Ing., Dipl.-Ing., has been a professor at the ESB Business School, Reutlingen University since 2010. Previously she held leading positions at the Fraunhofer IPA in Stuttgart, the working area Industrial Engineering, and the Graduate School of Advanced Manufacturing Engineering at the University of Stuttgart. She is currently leading the expert group in logistics at Reutlingen University.

Development of a Criteria Catalogue for the Classification of Intralogistics Systems in the Context of Self-Organization

J. Schuhmacher[1], V. Hummel[1]
[1]ESB Business School, Reutlingen University, Germany

Abstract
The global demand for individualized products leading to decreasing production batch sizes requires innovative approaches how to organize production and logistics systems in a dynamic manner. Current material flow systems mainly rely on predefined system structures and processes, which result in a huge increase of complexity and effort for system and process changes to realize an optimized production and material provision of individualized products. Autonomous production and logistics entities in combination with intelligent products or logistic load carriers following the vision of the "Internet of Things" offer a promising solution for mastering this complexity based on autonomous, decentralized and target size-optimized decision making and structure formation without the need for predefined processes and central decision-making bodies. Customer orders are going to prioritize themselves and communicate directly with the required production and logistics resources. Bins containing the required materials are going to communicate with the conveyors or workers of the respective intralogistics system organizing and controlling the material flow to the autonomously selected workstation. A current research project is the development of a collaborative tugger train combing the potential of automation and human-robot collaboration in intralogistics. This tugger train is going to be integrated into a self-organized intralogistics scenario involving individualized customer orders (low to high batch sizes). To classify the application of self-organization within intralogistics systems, a criteria catalogue has been developed. The application of this criteria catalogue will be demonstrated on the example of a self-organization scenario involving the collaborative tugger train and an intelligent bin system.

Keywords
Self-organization, Intralogistics, Tugger train, Catalogue of Criteria

1 INTRODUCTION

Manufacturing companies are confronted with the increasing international competitive pressure, decreasing batch sizes due to a growing demand for individualized products and the requirement for the shortest possible delivery time [1]. Future logistics systems must enhance to adapt flexibly to the changing requirements of material flow in order to enable the factories of the future to produce personalized products in small batch sizes down to lot size one under the performance and cost conditions of today's mass production [2, 3]. The potential of evolutionary flexibilization of machine technology and production organization has already been largely exhausted. Through the networking, direct communication and collaboration of humans, machines and products aimed for in Industrie 4.0 the efficient production of individualized products will be supported through the concept of self-organization [4]. The market demand for economical production of individualized products at the cost conditions of mass production not only requires a changeable design of the factory environment, but also new methods how the production including the logistic processes are organized, planned and controlled [3].

As each product in this customized production differs from previous products in terms of the required production processes as well as the required components and its flow through the factory, real-time configuration, control and decision-making within these flexible factory environments becomes a key challenge. For the organization, planning and control of flexible material flow systems within networked production environments, new methods and systems based on self-organized or autonomously controlled systems as well as the theory of the Internet of Things offer promising solutions. The approach of self-organized factory environments has been researched for a long time, e.g. in context with Fractal Factory structures [5]. In recent years, the topic has become increasingly important due to increasing market requirements on the one hand and the now available technical possibilities in the area of networking, decentralized information processing at the field level and the development of cyber-physical systems on the other [6]. According to Ten Hompel [3] and Günthner [7] this results in a configuration and regulation of the material flow at every point of the material flow system in order to keep pace with the increased flexibility requirements of production. Self-organized, autonomously controlled decentralized material flow control systems will distribute the required decision-making and control processes to intelligent logistical units [7]. Machines and objects like intelligent products will jointly decide which tools and machines are to be used and with which conveying means components and (semi-finished) products are moving to the next production step. [8]

This paradigm shift requires profound organizational and control changes for companies, which may involve investments in new or adapted hardware and software components as well as product adaptations [9]. This required development in the direction of self-organized production systems will be evolutionary, since the investments in existing factories must first have paid off from the companies' point of view and a company-specific roadmap has to be developed for this transition [10].

2 SELF-ORGANIZATION

The theory of self-organized systems in combination with autonomous cooperation and control of logistic objects is seen as the answer for logistic systems to cope with complexity and dynamics [11]. A major aim of self-organization in logistics is to achieve changeable logistics systems. Nopper [12] describes the changeability of intralogistics systems as the "…capability of a material flow system […] to adapt to the requirements of the environment beyond the limits of the system's design. In order to do so, the system must be expandable or adaptable. The requirements for material flow systems can be described along the dimensions of conveyed goods, layout and throughput."

2.1 Differentiation of self-management, self-organization and autonomous cooperation and control

Following Windt [11] the terms "self-organization" and "self-management" are to be distinguished as concretizations of the term "autonomous control". Even though the concepts of "autonomous cooperation", "self-organization" and "self-management" describe a system's ability to create order based on its own resources, there are differences regarding the form and degree of this ability [13]. *Self-management* as the broadest concept describes the fully autonomous development of a system involving the formulation of objectives and plans, the decision on its own organization forms and required resources [14]. *Self-organization* as an element of management outlines the manner how a system creates its own structure and processes by using its own abilities [15]. Self-organization is a long-established approach for human organizations and companies, for example investigated by Probst [16] in social systems and Warnecke [5] in context with the concept of Fractal Factories. *Autonomous cooperation* only refers to the freedom degree of system elements. So based on the current situation, the system elements are able to choose among given alternatives predefined by external entities like the company management [17].

The following definition of autonomous cooperation and control has been developed by the Cooperative Research Centre (CRC) 637 "Autonomous Cooperating Logistic Processes – A paradigm Shift and its Limitations" at the University of Bremen [13]: "Autonomous Control describes processes of decentralized decision-making in heterarchical structures. It presumes interacting elements in non-deterministic systems, which possess the capability and possibility to render decisions. The objective of autonomous control is the achievement of increased robustness and positive emergence of the total system due to distributed and flexible coping with dynamics and complexity." According to Ten Hompel [18], the Internet of Things emerges within the framework of the self-organization and autonomous control of intelligent logistical objects. Within this Internet of Things logistical objects move independently through intralogistics networks like data packages in the Internet of Data. Ten Hompel [19] suggests to use the term "autonomous control", which was marked by the CRC 637, for logistics networks which has been mainly investigated within the CRC 637 and the terms "Internet of Things" and "self-organization" for intralogistics applications.

Regarding the characteristic features of self-organization and autonomous control, it can be observed that in the approach of self-organization, the characteristics at the management and organizational system level are more pronounced, whereas for autonomous control, the characteristics are more relevant for the execution system. In addition, the approach of autonomous control is more oriented towards the individual (logistical) object, while the approach of self-organization covers the system as a whole. In addition the approach of autonomous control is a more technology-oriented approach, e.g. in the discussion about Industrie 4.0 and the Internet of Things, since this approach is often associated with the use of new technologies whereas the approach of self-organization is more human focused. [11, 18, 20]

2.2 Characteristics of self-organized systems

Major characteristics that can be found in numerous self-organization approaches and definitions are the characteristics of complexity, dynamics, non-determinism, autonomy, redundancy, interaction and emergence. [13, 21, 22]

In addition, the considered systems are dynamic, complex systems no matter what type of system is involved in which discipline, since they are all based on the existence of numerous interrelationships between the system elements themselves and between the system and its environment [21, 23]. In the following, the major characteristics of self-organization in general will be discussed to set the theoretical base for self-organized intralogistics systems:

1. **Complexity**: A system is generally called a complex system, if its detailed description is hardly possible due to the variety of elements

and relations, even if all information and interactions would be known [24, 25]. The key difference between complicated and complex systems is the dynamics in the system. Complicated systems can consist of a large number of elements and connections, but these are subject to little movement or dynamics. Complex systems are characterized by a high dynamic that is present between the elements and can be described by the rate of change of the system over time. [24, 26]

2. **Dynamics**: Dynamics refers to the change of a system state or individual variables in a system in the time dimension. The intensity of the dynamics of the system processes depends in particular on the dynamics of the environment and the openness of the system to the environment. [27]
3. **Non-determinism:** A system is called non-deterministic if its behavior cannot be predicted over a longer period of time, although precise information about the states and rules of the system is available [13, 14]. The property of non-determinism should enable the system and its elements to deal more efficiently with complexity, dynamics and uncertainty in processes. [13, 28]
4. **Autonomy**: A major intrinsic feature of self-organized systems is the autonomy of the system in the sense that the relationships and interactions that unite the system only include the system itself and do not require further systems in the sense of a self-reference and operational coherence. This includes the self-design, control and development of the system, which is performed in a constant cycle of cause and effect. However, this does not lead to complete independence from other systems. No system that is part of a larger system can be completely independent or autonomous. It can only be autonomous with regard to certain criteria, as it can still be exposed to significant external influences or the environment, to which the system reacts using corresponding external signals. Thus, the concept of self-organization focuses on the active role of the system with its self-initiated actions, the design and influencing of system and environment as well as the formation and preservation of functional boundaries and identity. [16]
5. **Redundancy**: In a self-organized system, no distinction is made between the (active) organizing, designing or controlling components of the system and the (passive) components that are organized, designed or controlled. These capabilities are distributed across the system without the need for a hierarchy (heterarchy principle) and the potentials and mechanisms of system design and control are a characteristic of the system itself. The elements of the system take over the design and control of the system, which have most of the information. This redundancy in the system means that several system components can be able to do the same (design and control). This redundancy of the functions also allows internal flexibility in the system. [16]
6. **Interaction and emergence**: The development of a self-organized structure in a system is based on the interaction of different system elements, which exchange information, substances, knowledge or energy with each other [21, 29]. Through the interaction of system elements new qualitative properties are formed within the system, which are called emergences [21] or synergetics [30]. These are based on the synergy effects of the elements in interaction with each other and cannot be related to individual system elements [21, 30]. The synergetics illustrates that the interaction of system elements makes the overall system more powerful than the mere sum of its subsystems [30].
7. **Autonomous order formation:** Through a higher degree of autonomy in the logistic system and its (intelligent) units a positive emergence and autonomous order formation can be achieved. In natural systems, completely autonomous order formation is generally possible, whereas in socio-technical systems, such as companies and logistics systems, autonomous control and self-organization can only take place to a limited extent following [21]. A major reason for this is that the origin and the system design of natural systems lie in the overall context of nature, which is generally self-organized. The creation and design of companies, in contrast, is a deliberately planned and artificially created existence by man, and thus externally organized [24].

With regard to intralogistics, it can be assumed that the optimal degree of self-organization in a complex logistics system in terms of achieving defined logistical goals does not correspond to the maximum degree of self-organization [31]. In this context, a combination of external and self-organization or the self-organization of selected functions or organizations within logistics systems, which are, however, subject to defined framework conditions or logistics values, is therefore regarded as appropriate [21]. So the determination of the feasible degree of self-organization to counter complexity and to achieve defined (logistic) goals is of crucial importance.

3 RESEARCH METHODOLOGY

To investigate the potential of intralogistics systems in the context of self-organization the research methodology of a reasoning cycle starting with the hypothesis formulation, deduction of predictions, testing and observation of predictions and induction/feedback into the initial hypothesis was applied (also see figure 1). The main hypothesis to be proven is that self-organization applied, enables intralogistics systems to react more flexible and agile to changing system requirements.

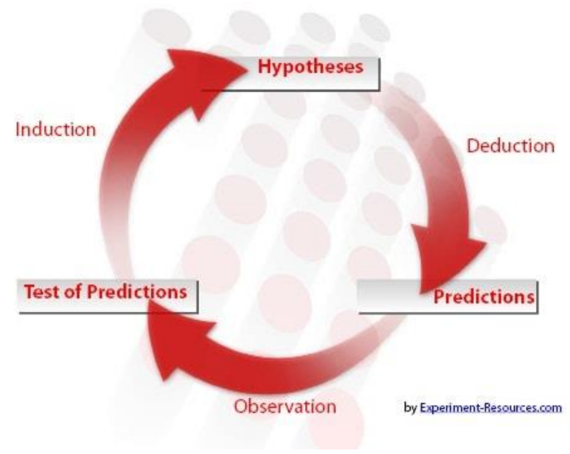


Figure 1 - research methodology of a reasoning cycle [32]

In order to do so a criteria catalogue was developed based on which the current degree of self-organization for intralogistics systems and the potentials for an increase can be determined. The catalogue was derived on the characteristics for self-organized systems with autonomous control according to Böse [33]. In addition, a generic model for intralogistics was developed which can be used to map different scenarios. Out of this generic model, scenarios were derived and then classified in the criteria catalogue developed.

4 RESULTS

4.1 Criteria catalogue

The criteria involves organizational as well as planning and control functions of intralogistics systems. To each criterion, different properties are assigned to cover the variety from conventional centralized controlled, external organized systems to autonomous controlled and self-organized systems. In order to operationalize the determination of the degree of self-organization a fulfilment value, which reaches from 0 (conventional external organized system) to 3 (entirely self-organized system), is allocated to every property of a criterion. To assess the degree of self-organization of a specific intralogistics system, all fulfilment values of the properties can be added up to the total fulfilment value. This total fulfilment value can be used to set the considered intralogistics system in relation to a completely self-organized system and so the degree of self-organization can be determined. The major criteria categories for the classification of the logistics systems are "decision making and organization", "information processing" and "decision execution". Within the category "decision making and organization" criteria such as the time behavior of the objective system, organizational structure as well as the change capability of the organization and role of the human worker are assessed. The category "information processing" covers aspects as the location of data storage and processing. The third category of "decision execution" covers amongst others the system's identification and measuring ability and flexibility.

4.2 Generic model

Following Böse [33] and Ropohl [34] the required system layers for a generic model of intralogistics are the decision, information and execution system level (also see figure 2). The decision system provides the decision-making ability. As mentioned before, in self-organized systems the decision-making functions are transferred to heterarchically organized logistic objects. The decision-making system also involves planning and control tasks enabling the system to develop and adapt in an autonomous, target-oriented manner. The information storing and processing ability on the information-processing layer forms the basis for the decision-making of the logistic objects interacting and exchanging information with each other. The decision execution layer deals with the decision execution ability of the logistic objects building up the intralogistics system consisting of sources, sinks and relations. These three layers set the basis for self-organization of intralogistics systems.

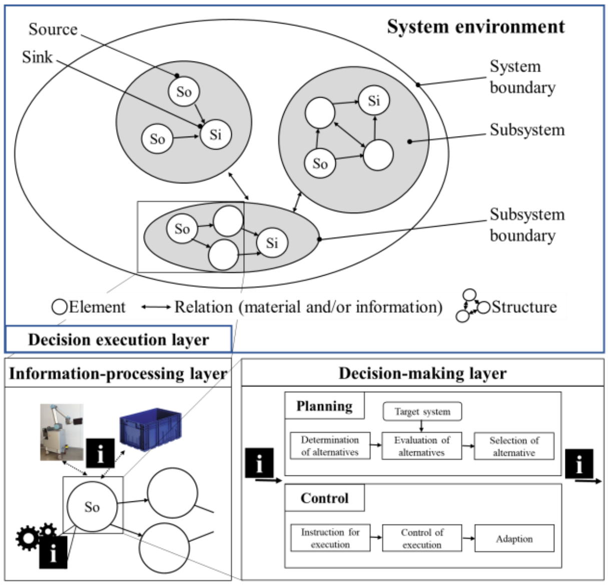


Figure 2 - Generic system model (cf.[33, 35])

4.3 Collaborative tugger train scenario

The following application scenario has been defined from the generic intralogistics model for a "Collaborative tugger train 4.0" and was classified in the criteria catalogue to visualize the existing degree of self-organization and thereby determine starting points for further enhancement. The aim of the collaborative tugger train 4.0 is to investigate the potentials of automation and human-robot collaboration for tugger train systems within changeable factory environments. The trailers of the tugger train are towed by an autonomous mobile robot platform with an articulated robot on top of the platform to manipulate small load carriers (see figure 3). The robot platform of the tugger train system is capable to drive autonomously through changing factory environments and to handle bins by using different sensor systems. So in contrast to conventional tugger train systems no human worker is required to drive the tugger train and to load and unload the small load carriers.

Figure 3 - Collaborative tugger train

Based on the developed generic model described above a material provision scenario for an assembly environment has been developed. This scenario involves the collaborative tugger train system in combination with an intelligent bin system and other manual, semi-automated and automated conveyor systems to investigate the planning, design and control as well as the potential of self-organized intralogistics systems. The transport strategy for the tugger train is going to involve methods for flexible route planning, departure time determination, vehicle scheduling and order scheduling also considering previous research done by e.g. [36, 37].

Figure 4 shows an extract of the mapping of the scenario in the criteria catalogue (Initial scenario marked blue). The decision making and organization emergence are to be done in a decentralized manner based on the local target systems of the intelligent logistical objects. In this specific scenario, the intelligent bin might have the prioritized target to be transported as fast as possible from the source (supermarket storage) to the sink (workstation). The collaborative tugger in this scenario strives for maximum utilization of transport capacity. Therefore, the tugger train is bidding on the transportation order of the intelligent bin by communicating the required transportation time to the intelligent bin. Besides the tugger train, also other transport systems of the intralogistics system, like roller conveyors or handcarts with a certain degree of intelligence, might bid on this transport in accordance with their target systems. The intelligent bin then compares all the bids coming from the transport systems and selects the most favorable transport. So an autonomous controlled material replenishment process can be initiated in a decentralized manner through a direct communication between the intelligent bin and available transport systems. The most relevant logistical objects as the transport systems and intelligent bins within the investigated system are uniquely identifiable and locatable leading to an overall medium degree of self-organization and autonomous control based on a scenario-specific total fulfillment value of 20 out of 42 with potential to increase. The maximum total fulfillment value of 42 represents in this case the maximum degree of self-organization for each criteria considered.

Criteria	criteria	Properties			
Decision making & Organization	Distribution of the target system	0 global	1 mainly global	2 mainly local	3 local
	Adaptivity of the target system	0 static	1 mainly static	2 mainly dynamic	3 dynamic
	organizational structure	0 hierarchical	1 mainly hierarchical	2 mainly heterarchical	3 heterarchical
	Change capability of the organization	0 low	1 medium	2 high	3 very high
	Number of decision alternatives	0 None	1 some	2 many	3 unlimited
	Type of decision-making	0 extrinsic (static)	1,5 intrinsic (rule-based)		3 Intrinsic (learning)
	Place of decision-making	0 system level	1,5 subsystem level		3 system element level
	Human role	0 passive, executing	1 mainly passive, executing	2 mainly aktive, co-decisive	3 Actively decisive
	Predictability of system/ element behavior	0 Elements and system deterministic	1 Elements non-/system deterministic	2 System non-/elements deterministic	3 Elements and system non-deterministic
Information processing	Location of data storage	0 central	1 mainly central	2 mainly decentralized	3 decentralized
	Location of data processing	0 central	1 mainly central	2 mainly decentralized	3 decentralized
Decision execution	Flexibility	0 inflexible	1 little flexibility	2 flexible	3 very flexible
	Identifiability	0 no elements identifiable	1 some elements identifiable	2 many elements identifiable	3 all elements identifiable
	Localizability	0 no elements can be localized	1 some elements can be localized	2 many elements can be localized	3 All elements can be localized
Increasing degree of self-organization →					

Figure 4 - Self-organized tugger train scenario mapped in catalogue of criteria (cf. [33, 38])

To increase the degree of self-organization of the tugger train system among other things the capabilities of all logistic objects (bins, tugger train, etc.) have to be further increased in the field of decentralized, intrinsic decision making, decentralized data storage and processing as well as identification and localization capabilities (Potential marked green in figure 4). Also a more active integration of the human worker, e.g. in the sense of a higher integration and responsibility for the intralogistics system, offers potential to increase the level of self-organization. By doing this, the target-oriented, autonomous organization of the intralogistics system involving the tugger train system and other intelligent logistic objects can be further improved.

5 CONCLUSION AND OUTLOOK

One of the next steps will be the implementation of the developed scenario to investigate the intralogistics system behavior in combination with changing production system requirements. First will be the implementation of the described collaborative tugger train scenario with a medium level of self-organization in the ESB Logistics Learning Factory at Reutlingen University. In line with the selected research method of the reasoning cycle the hypothesis that self-organization enables intralogistics systems to react more flexible and agile to changing system requirements will be tested. Based on the observations of this testing, a scenario with a higher degree of self-organization is going to be defined by applying the developed generic intralogistics model and classifying of the developed scenario in the criteria catalogue. Then the results of these scenarios are going to be compared with each other with respect to the logistic goal achievement and fulfillment of the hypothesis. Further scenarios with different intralogistics infrastructure will be defined, implemented and analyzed in context of self-organization and autonomous control in order to develop a method for the autonomous control of changeable, hybrid logistics systems in the end.

6 ACKNOWLEDGMENTS

The research project “collaborative tugger train 4.0” (FHprofUnt “Kollaborativer Routenzug 4.0 (KollRo 4.0)”) has been funded by the German Federal Ministry of Education and Research.

7 REFERENCES

[1] Spath D. (Ed.), Produktionsarbeit der Zukunft - Industrie 4.0, Fraunhofer-Verl., Stuttgart, 2013.

[2] Dais, S., Industrie 4.0 – Anstoß, Vision, Vorgehen, in: T. Bauernhansl, M. ten Hompel, B. Vogel-Heuser (Eds.), Industrie 4.0 in Produktion, Automatisierung und Logistik, Springer Fachmedien Wiesbaden, 2014, pp. 625–634.

[3] Hompel, M. ten, Henke, M., Logistik 4.0, in: T. Bauernhansl, M. ten Hompel, B. Vogel-Heuser (Eds.), Industrie 4.0 in Produktion, Automatisierung und Logistik, Springer Fachmedien Wiesbaden, 2014, pp. 615–624.

[4] Bürger, T., Tragl, K., SPS-Automatisierung mit den Technologien der IT-Welt verbinden, in: B. Vogel-Heuser, T. Bauernhansl, M. ten Hompel (Eds.), Handbuch Industrie 4.0 Bd.1, Springer Berlin Heidelberg, Berlin, 2017, pp. 207–217.

[5] Warnecke, H.-J., Hüser, M., Die Fraktale Fabrik: Revolution der Unternehmenskultur, second. Aufl., Rowohlt, 1996.

[6] Bauernhansl, T., Die Vierte Industrielle Revolution – Der Weg in ein wertschaffendes Produktionsparadigma, in: B. Vogel-Heuser, T. Bauernhansl, M. ten Hompel (Eds.), Handbuch Industrie 4.0 Bd.4, Springer Berlin Heidelberg, 2017, pp. 1–31.

[7] Günthner, W., Klenk, E., Tenerowicz-Wirth, P., Adaptive Logistiksysteme als Wegbereiter der Industrie 4.0, in: T. Bauernhansl, M. ten Hompel, B. Vogel-Heuser (Eds.), Industrie 4.0 in Produktion, Automatisierung und Logistik, Springer Fachmedien Wiesbaden, 2014, pp. 297–323.

[8] Kleinemeier, M., Von der Automatisierungs-pyramide zu Unternehmenssteuerungs-Netzwerken, in: B. Vogel-Heuser, T. Bauernhansl, M. ten Hompel (Eds.), Handbuch Industrie 4.0 Bd.1, Springer Berlin Heidelberg, 2017, pp. 219–226.

[9] Bildstein, A., Seidelmann, J., Migration zur Industrie- 4.0-Fertigung, in: B. Vogel-Heuser, T. Bauernhansl, M. ten Hompel (Eds.), Handbuch Industrie 4.0 Bd.1, Springer Berlin Heidelberg, 2017, pp. 227–242.

[10] Wegener, D., Industrie 4.0 – Chancen und Herausforderungen f€ur einen Global Player, in: B. Vogel-Heuser, T. Bauernhansl, M. ten Hompel (Eds.), Handbuch Industrie 4.0 Bd.2, Springer Berlin Heidelberg, 2017, pp. 391–405.

[11] Windt, K., Selbststeuerung intelligenter Objekte in der Logistik, in: M. Vec (Ed.), Selbst-organisation: Ein Denksystem für Natur und Gesellschaft, Böhlau, Köln, 2006, pp. 271-314.

[12] Nopper, J.R., Eine Methodik zur Bewertung von Wandelbarkeit in der Intralogistik am Beispiel selbstorganisierter Materialflusssysteme, Verl. Praxiswissen, Dortmund, 2011.

[13] Windt, K., Hülsmann, M., Changing Paradigms in Logistics – Understanding the Shift from Conventional Control to Autonomous Cooperation and Control, in: K. Windt, M. Hülsmann (Eds.), Understanding autonomous cooperation and control in logistics, Springer, Berlin, New York, 2007, pp. 1–16.

[14] Manz, C.C., Sims, H.P., Self-Management as a Substitute for Leadership: A Social Learning Theory Perspective, The Academy of Management Review 5 (1980) 361.

[15] Probst, G.J.B., Organisation: Strukturen, Lenkinstrumente, Entwicklungsperspektiven, first. Aufl., Verl. Moderne Industrie, Landsberg/Lech, 1993.

[16] Probst, G.J.B., Selbst-Organisation: Ordnungsprozesse in sozialen Systemen aus ganzheitlicher Sicht, Parey, Berlin, 1987.

[17] Bea, F.X., Göbel, E., Organisation: Theorie und Gestaltung ; mit zahlreichen Übersichten, second., neu bearb. Aufl., 2002.
[18] Bullinger, H.-J., Hompel, M. ten, Internet der Dinge: Www.internet-der-dinge.de, Springer-Verlag Berlin Heidelberg, 2007.
[19] Hompel, M. ten, Intralogistik — Auf dem Weg vom Prozess zum Service, in: H. Baumgarten (Ed.), Das Beste der Logistik, Springer Berlin Heidelberg, 2008, pp. 103–110.
[20] Scholz-Reiter, B., Freitag, M., Rekersbrink, H., Wenning, B.L., Gorldt, C., Echelmeyer, W., Auf dem Weg zur Selbststeuerung in der Logistik – Grundlagenforschung und Praxisprojekte, in: G. Wäscher (Ed.), Intelligente Logistikprozesse: Konzepte, Lösungen, Erfahrungen, LOGISCH GmbH, Magdeburg, 2005, pp. 166–180.
[21] Hülsmann, M., Wycisk, C., Selbstorganisation als Ansatz zur Flexibilisierung der Kompetenzstrukturen, in: C. Burmann, J. Freiling, M. Hülsmann (Eds.), Neue Perspektiven des Strategischen Kompetenz-Managements, Gabler, Wiesbaden, 2006, pp. 323–350.
[22] Wehberg, G., Ökologieorientiertes Logistikmanagement: Ein evolutions-theoretischer Ansatz, Gabler Edition Wissenschaft, Wiesbaden, 1997.
[23] Malik, F., Strategie des Managements komplexer Systeme: Ein Beitrag zur Management-Kybernetik evolutionärer Systeme. Teilw. zugl.: St. Gallen, Univ., Habil.-Schr., 1978, eleventh. Aufl., Haupt, Bern, 2015.
[24] Wycisk, C., Flexibilität durch Selbststeuerung in logistischen Systemen: Entwicklung eines realoptionsbasierten Bewertungsmodells, 2009.
[25] Gell-Mann, M., What is Complexity, in: A.Q. Curzio, M. Fortis (Eds.), Complexity and Industrial Clusters: Dynamics and Models in Theory and Practice, Physica-Verlag HD, Heidelberg, 2002, pp. 13–24.
[26] Probst, G.J.B., Gomez, P., Die Methodik des vernetzten Denkens zur Lösung komplexer Probleme, in: G.J.B. Probst, P. Gomez (Eds.), Vernetztes Denken: Ganzheitliches Führen in der Praxis, erweiterte Auflage, 1991, pp. 3–20.
[27] Beer, C.d., Untersuchung der Dynamik von selbststeuernden Prozessen in produktionslogistischen Systemen anhand ereignisdiskreter Simulationsmodelle. Univ., Diss., 2008, GITO-Verl., Berlin, 2008.
[28] Hülsmann, M., Austerschulte, L., Selbststeuerung – Ein Ansatz zur Balancierung von Flexibilität und Stabilität organisationaler Kompetenzen?, in: H. Proff, A. Eisenkopf, C. Opitz (Eds.), Strategisches Kompetenz-Management in der Betriebswirtschaftslehre: Eine Standortbestimmung, Deutscher Universitäts-Verlag, 2008, pp. 21–46.
[29] Haken, H., Die Selbstorganisation der Information in biologischen Systemen aus der Sicht der Synergetik, in: B.-O. Küppers, Ordnung aus dem Chaos: Prinzipien der Selbsorganisation und Evolution des Lebens, Orig.-Ausg, 1987, pp. 127–156.
[30] Haken, H., Erfolgsgeheimnisse der Natur: Synergetik: Die Lehre vom Zusammenwirken, Ungekürzte Ausg., fourteenth – sixteenth, 1988.
[31] Scholz-Reiter, B., Beer, C.d., Böse, F., Windt, K., Evolution in der Logistik - Selbststeuerung logistischer Prozesse, in: 16. Deutscher Materialfluss-Kongress -Intralogistik bewegt - mehr Effizienz, mehr Produktivität, VDI Verlag, 2007, pp. 179–190.
[32] Information on https://explorable.com/de/forschungsmethodik
[33] Böse, F., Selbststeuerung in der Fahrzeug-logistik: Modellierung und Analyse selbststeuernder logistischer Prozesse in der Auftragsabwicklung von Automobilterminals. Bremen, Univ., Diss., GITO-Verl., Berlin, 2012.
[34] Ropohl, G., Allgemeine Technologie eine Systemtheorie der Technik, KIT Scientific Publishing, s.l., 2009.
[35] Wiendahl, H.-P., Betriebsorganisation für Ingenieure, eighth., überarb. Aufl., Hanser, München, 2014.
[36] Klenk, E., Galka, S., Günthner, W.A., Potentiale einer flexiblen Routenzugsteuerung, ZWF Zeitung für wirtschaftlichen Fabrikbetrieb 110 (2015) 805–809.
[37] Günthner, W.A., Keuntje, C., IntegRoute - Ganzheitliche Konzeptauswahl für Routenzugsysteme zur Produktionsversorgung, TU München, Lehrstuhl für Fördertechnik Materialfluß Logistik, Garching, 2016.
[38] Böse, F., Windt, K., Catalogue of Criteria for Autonomous Control in Logistics, in: K. Windt, M. Hülsmann (Eds.), Understanding autonomous cooperation and control in logistics: The impact of autonomy on management, information, communication and material flow, Springer, Berlin, New York, 2007, pp. 57–72.

8 BIOGRAPHY

Vera Hummel, Prof. Dr.-Ing., Dipl.-Ing., has been a professor at the ESB Business School, Reutlingen University since 2010. She is a founding member of the “Initiative of European Learning Factories” and the initiator of the "ESB Logistics Learning Factory“ for education and industry training on the campus of Reutlingen University.

Jan Schuhmacher (MSc.) is a research associate at ESB Business School and part of the research, training and consultant group of ESB Logistics Learning Factory. His major field of research is the design and optimization of changeable intralogistics systems.

Game-Based Learning and Virtual Reality: Innovation in Performance Improvement

T. von Leipzig[1, 2], P.A. Gouws[2], M. H. Greeff[2]
[1]Department of Industrial Engineering, University of Stellenbosch, South Africa
[2]LTS Health, South Africa

Abstract
Diagnostic laboratories today are faced with immense pressures to consistently improve in a continuously changing environment. External factors such as new legislations, technological advances and consolidation requirements, as well as internal pressures to improve the bottom line whilst improving quality, are forcing laboratories to do more with less. Some of the biggest performance improvements can be made by improving the laboratory processes. The nature of the industry is however a silo approach, with different areas in the laboratory being acutely aware of their own requirements, but oblivious to those of others and of the laboratory value chain. LTS Health, a global leader in laboratory performance improvement, has developed courses for training staff to understand their processes holistically, to identify improvement opportunities and lead improvement projects. A commonly-experienced shortfall of such courses, other than the cost of removing a whole team from the lab, is the difficulty of visualizing processes. This is especially true in the African market, where language barriers and the lack of some fundamental educational concepts make it difficult to convey understanding through standard process maps. Virtual reality and game-based learning could provide potential enhancements by offering a visual and realistic picture of the laboratory as well as increased immersion and interaction. Furthermore, it would provide the opportunity of completing the training remotely. Beyond its use in training, virtual reality could be utilized in facility design. The increasing pressure to consolidate often brings with it the difficulties in conceptualizing the new laboratory effectively prior to consolidation. A virtual design could offer laboratories the opportunity to experience their new facility prior to development, ensuring implementation of the best possible solution.This paper presents LTS Health's most recent developments in digital training as case studies for how technology can facilitate training, facility design and process improvement in the digital age.

Keywords
Laboratory performance, technology-enhanced learning, innovative training

1 INTRODUCTION

Diagnostics, an industry currently valued at $62 billion and expected to reach a value of $87 billion by 2024 [1], plays a critical role in the medical environment. With 60-70% of medical decisions based on the results of diagnostic tests [2], diagnostics improves patient care, and contributes to consumer safety and the reduction of healthcare spending [3].

Continuous change in both the external and internal environments places pressure on stakeholders in the health care value chain to consistently reduce costs whilst improving quality [2, 4, 5]. Considering the influence of, and reliance on their results, diagnostics laboratories are ever-more forced to do more with less.

1.1 Industry Pressures

One of the most obvious pressures is inflicted by the economic environment. Drastic cuts can be seen in government research funding as well as in overall health budgets. Furthermore, a heightened sensitivity to macroeconomic changes is realised due to a significant revenue stream being industrial end markets, such as food testing, who themselves are sensitive to macroeconomic changes. Whilst these economic pressures can be expected to lead to better resource use and increased effectiveness in the long term, the short-term impact thereof is the pressure to cut costs [4, 6].

In addition, legislative changes place laboratories under further pressure to remain compliant whilst improving the bottom line [4].

Across industries, the increasingly rapid rate of technological and digital advancements is seen as one of the biggest challenges facing organisations [7]. The pace of innovation required to stay ahead increases with the rate of advancements. Most organisations however struggle to keep up, resulting in what is termed Martec's Law, in which "[t]echnology changes exponentially (fast) yet organisations change logarithmically (slow)" creating an ever-increasing gap [8].

These pressures, along with globalisation, consolidation requirements and the introduction of a new workforce, are resulting in various trends in the diagnostics industry.

1.2 Resulting Trends in Industry

With the constant pressure to do more with less, a clear trend toward automation can be seen in laboratories with new technologies offering faster, cheaper and in some cases, more accurate testing [4].

A technological advancement that completely disrupted the laboratory industry and attempted to overturn Martec's Law, was the Edison by Theranos [9]. Although it was later found that the technology did not work, and test results had been falsified, "Theranos struck a chord that needed to be played" [9]. Change is inevitable within the laboratory environment [10], making it "an area that's ripe for disruption" [9].

Another clear trend is that of consolidation, with the main drivers being a shortage of resources, the potential for multi-functional economies of scale as well as competitive aspects in the diagnostics industry [4]. Consolidation, when implemented effectively, has shown to produce significant bottom line savings, without compromising on quality [11]. With the constant pressure to do more with less, many established laboratories look toward consolidation and acquisitions to solidify their market positions.

Furthermore, nearly half of laboratory professionals will retire within the next decade, with a new younger workforce taking their place [5]. A clear need has been identified to equip the upcoming young workforce with the skills required to manage laboratories effectively in an ever-changing environment.

Since "change is inevitable for laboratories to stay abreast […] and adapt to new customer needs" [10], LTS Health aims to prepare laboratory staff for the dynamic fast-paced laboratory environment and help alleviate the skills gap through a variety of training courses. The foundations on which these courses are built, are the 8 Principles of Laboratory Performance.

2 8 PRINCIPLES OF LABORATORY PERFORMANCE

The 8 Principles of Laboratory Performance present "a new philosophy based in diagnostics, that allow[s] laboratories to be innovative through certainty" [12].

Figure 1 depicts the 8 Principles which aim to assist laboratories to "minimize risk [and] maximise impact" [12] by focusing on:

- structuring the organisation for decision-making and certainty;
- empowering the organisation to make the right decisions; and
- creating and sustaining an innovative environment to ensure success.

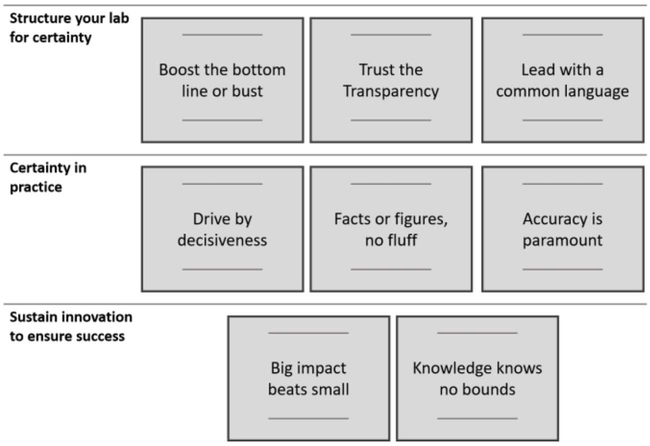


Figure 1 - The 8 Principles of Laboratory Performance [12]

Following these principles, a laboratory is most efficient and effective when [12]:

- it strives for a bottom line impact;
- its staff understand the business;
- information is shared across boundaries;
- business cases are presented honestly based on facts;
- decisions are made with certainty;
- accuracy is emphasised;
- strategies and tasks are prioritised based on expected results and anticipated requirements; and
- knowledge is shared.

To empower laboratories to implement and sustain these principles and innovation, LTS Health offers laboratory professionals the Power of Process range of courses.

3 THE POWER OF PROCESS

"The performance of individuals is only as good as the process will allow it to be" [13]. In many cases, the root cause of many process-related problems is the lack of process understanding [13, 14]. A process cannot be effectively controlled nor improved if it is not understood.

The Power of Process range of courses aims to equip laboratory professionals with an understanding of laboratory processes and lab performance, and ultimately, the tools and methodologies to plan and implement improvement initiatives.

The range consists of three courses, the Pioneer, Champion and Master. The Champion and Master courses are completed sequentially and are aimed at performance improvement teams within the laboratory, to enhance their understanding of processes and their ability to control and improve them.

A strategic problem often experienced in process improvement and management is a lack of buy-in from staff [13]. An important goal of performance improvement should thus always be to create a “process understanding mindset” for the laboratory and all its individual staff [14].

The Pioneer course is a short-course developed to provide the remaining laboratory staff, those not participating in the Champion and Master courses or not directly involved in the performance improvement teams, with a high-level overview of the laboratory processes and the importance of performance improvement.

All three courses are delivered in a practical, interactive setting, utilising case studies to reinforce theory and develop the muscle memory needed to implement the skills in the workplace post-course.

Whilst these courses have been successful in equipping laboratory professional with a much-needed skill set, and have a high client satisfaction rate, they are ultimately limited in reach and impact.

4 LIMITATIONS OF CONVENTIONAL TRAINING

A commonly experienced downside of conventional courses is the cost, both in monetary and productivity terms. These training initiatives are expensive, and require a whole team to be removed from the laboratory to attend the training [5]. Many experts are looking towards technology-based solutions and eLearning as a means of providing lower cost, scalable learning solutions. In many instances however, technology is simply used to present the same content in a different medium, rather than to enhance or transform the learning experience.

Specifically in performance improvement, there is the additional difficulty of visualising and understanding processes [13, 14, 15]. This is especially true in the African market, where language barriers and the lack of some fundamental educational concepts makes it difficult to convey understanding.

Although process maps offer a standard means for visualising processes [13], they are limited to providing a static level of detail [15]. Technology offers the ability to enhance and surpass conventional training methods.

LTS Health plans to address these limitations and enhance its learning offerings through the addition of technology. These planned enhancements are discussed in the following sections.

5 TECHNOLOGY-BASED ENHANCEMENTS

LTS Health is currently looking at incorporating a spectrum of technologies to enhance its learning proposition, including game-based learning, virtual reality, augmented reality and tangible user interfaces.

5.1 Game-based learning

Specifically for the Pioneer short-course, which is aimed at providing a high-level overview rather than practical implementation skills, a technology-based solution offers the opportunity to increase the reach of the course by enabling participation without having to remove teams from the laboratory.

Game-based learning, the use of games with an educational objective, seeks to enhance training by deepening the learning experience [16]. Games inherently apply many principles of good pedagogy as listed in Table 1.

Pedagogical principle	**Application in games**
Individualisation	Games adapt to the level of the individual
Feedback	Games provide immediate and contextualised feedback
Active learning	Games provide an active environment which leads to discovery
Motivation	Games engage users for hours of engagement in pursuit of a goal
Social	Games can be played with others (e.g. multiplayer games) or involve communities of users interested in the same game
Scaffolding	Games are built with multiple levels; players cannot move to a higher level until competence is displayed at the current level
Transfer	Games allow users to transfer information from an existing context to a novel one
Assessment	Games allow users to evaluate their skill and compare themselves to others

Table 1 - Parallels between good pedagogy and game environments [17, p. 14]

Since the short-course aims to provide an overall understanding of performance improvement in the laboratory and the importance thereof, an interactive game holds great potential.

For the Champion and Master courses, virtual reality could surpass the potential of standard process maps by offering a visual and realistic picture of what is happening in the laboratory.

5.2 Virtual reality

Virtual reality (VR) is an alternate world filled with computer-generated images creating an immersive environment [18]. The user is usually brought into contact with this environment through VR glasses or goggles, fibre-optic data gloves and/or data suits.

VR holds the potential to transform the way educational content is presented. By immersing students/trainees in the environment they are learning about, VR motivates them to fully understand the concepts, and requires less cognitive load to process the information [19]. It can furthermore be expected to increase motivation, by creating an interactive, immersive and multi-sensory learning environment [20].

Furthermore, VR would provide the opportunity of completing the training remotely, without having to remove the staff from the laboratory.

Virtual environments can be created so as to give the user the feeling of experiencing a real environment, with the ability to create various scenarios for testing, without all the real costs and resources involved.

By overlaying the virtual lab with a process simulator, it would be possible for students to experience the lab in VR, with tests and samples flowing through the laboratory in real-time. VR would thus enable the accurate representation of the complexities of processes in a comprehensible manner [15]. Figure 2 depicts a realistic VR laboratory environment.

Figure 2 – Virtual lab app [21]

In traditional learning environments, students spend a large portion of the available time carefully reading the instructions as opposed to applying them in practice [19]. Although the case studies currently used in the Power of Process courses emphasize the development of muscle memory by testing whether the student is able to apply what he/she has learnt to a case, a large portion of time is still spent reading about a static case. With VR, students could be inspired to discover things for themselves rather than reading about them [19].

To deepen the learning experience, students should be encouraged to decide for themselves which information they need to solve a particular problem, rather than being given all the facts upfront. Augmented reality could present a possible extention to the virtual case that emphasises information pull.

5.3 Augmented reality

Augmented reality (AR) is the overlay of new information on top of an existing environment [22]. Figure 3 depicts an AR instance. Whilst the existing environment is usually the real world, AR could in this case be used as an extension to the virtual world. Students would be able to move through the virtual laboratory, and request additional information on aspects including work areas, equipment, samples and staff as needed.

Figure 3 – Augmented reality example [23]

By placing the responsibility of obtaining the information required to identify and ultimately solve the problem(s) in the virtual lab in the students' hands, self-directed learning is emphasised and a deeper learning experience can be achieved.

Considering the virtual case study holistically, the technologies offer both advantages, but also one significant disadvantage. One of the main advantages of VR is that it isolates the user, allowing him/her to concentrate in the virtual environment. However, this isolation also removes the social aspect of learning in a group [20, p. 64]. Much of the current courses' successes lie in the emphasis on interaction, collaboration and group work during the course. To maintain this emphasis even with the virtual case study, the tangible user interface could be used.

5.4 Tangible user interface

A tangible user interface (TUI) is an interface which makes the interaction and manipulation of objects in the digital world possible through the physical world [24]. As depicted in Figure 4, small objects with specific tags underneath them are placed on the TUI. The TUI is able to recognise the identity and orientation of the object and position it accordingly in the digital environment.

Figure 4 – Tangible user interface [24]

To incorporate group learning and collaboration, a TUI will be used in the case study for the to-be design of the improved laboratory. Students will be able to collaborate and move the objects around on the TUI, and experience the design changes in the virtual laboratory immediately. With the simulation overlay, they will furthermore be able to experience the effects of their design on the laboratory workflow, throughput and bottom line in real-time.

In the following section, the planned developments for LTS Health courses are described.

6 DEVELOPMENTS FOR LEARNING

The Pioneer short-course is focussed on teaching the importance of the 8 Principles of Laboratory Performance (Section 2). LTS Health has designed a game in which the analogy of human cells is used. Just as viruses continuously evolve at a rapid pace (analogous to the ever-changing laboratory environment), cells must adapt and evolve in order to stay alive (analogous to the laboratory). The player's aim in the game is to adapt and evolve the human cells in order to fight off viruses, with each level of the game increasing in complexity and corresponding to one of the 8 principles. At the end of the game, the lessons learnt in the game, and the 8 principles themselves, are related back to laboratory environment for deepened understanding.

The game's development is currently underway and is expected to enhance engagement and understanding of the principles in the laboratory staff by providing a fun, interactive learning environment.

A VR-based case study will be implemented for the Champion and Master courses, with the planned development phases depicted in Figure 5.

The first phase of planned development is a VR-based case study for enhanced realism, immersion and resulting understanding of the laboratory processes. Rather than provide students/trainees with required information regarding the case in a linear fashion (information push), AR will be used to augment the case study. By providing interactive touchpoints, the VR-based case study will allow the student to request information on resources, areas or events in the process (information pull), leading to stronger self-directed learning and enhanced understanding.

In the second phase, a process and simulation overlay will be integrated into the case study, increasing realism and immersion for the student. Case studies can be adapted to represent client-specific laboratory layouts, demand patterns and resource schedules, such that students can interact with a case study based on their own laboratory environment.

In the third phase, a TUI will be introduced. Not only would this offer increased levels of customisability, with the ability to realistically and easily model client-specific layouts, but it also introduces real-time scenario and to-be modelling. Ideally, students can collaborate in designing the to-be process, moving objects around on the interface and seeing the implications of their designs on laboratory efficiency and effectiveness for given demand in real time.

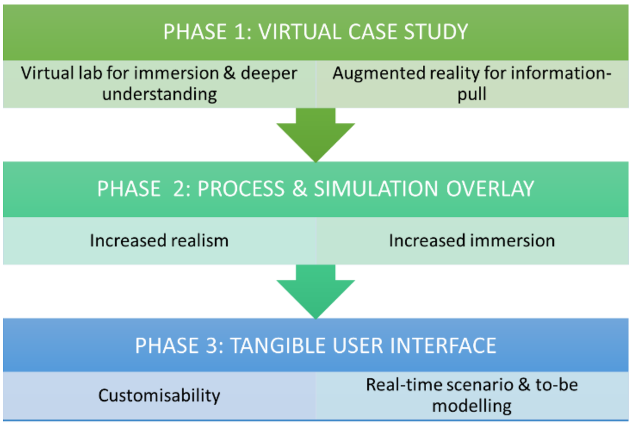


Figure 5 - Planned phases of development

Whilst the potential in training is clear, these planned developments may have additional applications.

7 POTENTIAL BEYOND TRAINING

Beyond its use in training, virtual reality could be utilised in facility design. The increasing pressure to consolidate often brings with it the difficulties of conceptualising the new laboratory effectively prior to consolidation. A virtual design could offer laboratories the opportunity to experience their new facility prior to development at a lower cost [25, 26], ensuring that the best possible solution is implemented from the start, as well as improved communication and shorter feedback loops with clients [27]. With the addition of the TUI, virtual designs could be easily modelled and updated in real-time simply by moving the objects on the interface.

8 CONCLUSION

The digital era has brought with it a plethora of new challenges and pressures. Everything needs to be cheaper, faster and of higher quality. The diagnostics environment is no exception.

LTS Health, a global leader in laboratory performance improvement, has developed courses for training laboratory staff to prepare them for today's fast-paced laboratory environment. LTS

Health plans to further enhance its learning offerings in the digital era, through the inclusion of technology.

A game has been designed for the Power of Process Pioneer course, with development currently underway. A VR-based case study for the Power of Process Champion and Master courses is planned, with planned extensions to include process simulation and a TUI for process design.

This planned approach not only holds the potential of increased realism, engagement and self-directed learning, but may provide an innovative enhancement to overcome barriers faced in traditional training. It provides a visual and more realistic means of interpreting processes and offers the opportunity of completing training remotely, without leaving the laboratory for extended periods of time. Ultimately, it aims to equip laboratory professionals with the skills needed to manage the ever-changing laboratory and in turn, assist in addressing the emerging skills gap in diagnostics.

9 REFERENCES

[1] Zion Market Research, “In-Vitro Diagnostics Market by Product (Reagents & Kits and Instruments), by Reagents & Kits (Clinical Chemistry, Immuno-Chemistry, Hematology, Microbiology Culture, Infectious Disease, Genetic Testing, and Others), by Instruments (CC/IA Systems, Micro Ar,” Zion Market Research, New York, 2018.

[2] The Lewin Group, Inc., “The value of diagnostics innovation, adoption and diffusion into healthcare,” The Lewin Group, Virginia, 2005.

[3] BioMérieux, “BioMérieux,” 2018. [Online]. Available: https://www.biomerieux.com/en/importance-diagnostics.

[4] C. Morel, L. McClure, S. Edwards, V. Goodfellow, D. Sandberg, E. Mossialos and J. Thomas, Ensuring innovation in diagnostics for bacterial infection: Implications for policy, United Kingdom: Europen Observatory on Health Systems and Policies, 2016.

[5] LABline, “Scalable learning for millennial workforce aims to narrow the lab skills gap,” Medical Laboratory Observer, 26 September 2018. [Online]. Available: https://www.mlo-online.com/scalable-learning-for-millennial-workforce-aims-to-narrow-the-lab-skills-gap. [Accessed 9 October 2018].

[6] H. Shah, “Consolidation of the Microbiology Laboratory Services: A Mini-Review of Finances and Quality of Care,” Lab Medicine, vol. 44, no. 1, pp. 86-89, 2013.

[7] CEMS, “Millennials see the rapid rate of technological and digital change as the biggest challenge facing them as future business leaders, global study finds,” CEMS, Jouy-en-Josas, 2017.

[8] S. Brinker, “Martec's Law: The greatest management challenge of the 21st century,” Martech, Boston, 2016.

[9] K. Loria, “The technology behind Theranos could still disrupt the enture lab business,” Business Insider, 2016.

[10] D. Kridelbaugh, “Leading your staff through change,” Lab Manager: Women in STEM, vol. 11, no. 11, pp. 22-25, 2016.

[11] S. Hemangi, “Consolidaion of microbiology laboratory services: A mini-review of finances and quality of care,” Laboratory Medicine, vol. 44, no. 1, pp. 86-89, 2013.

[12] LTS Health, “The 8 Principles,” LTS Health, 2018. [Online]. Available: https://www.the8principles.org/. [Accessed 3 September 2018].

[13] K. Halseth, “Process modelling and mapping: The basics,” David Thompson Health Region, Alberta, Canada, 2017.

[14] R. D. Snee, “Process understanding: Critical to successful process development, operation and improvement,” R&D Magazine, Newark, 2012.

[15] A. Pietsch, “Vritual reality process visualisation @ OTTO,” splunk, Washington DC, 2017.

[16] D. Codish and G. Ravid, “Personality based gamification – Educational gamification for extroverts and introverts,” Proceedings of the 9th CHAIS Conference for the Study of Innovation and Learning Technologies: Learning in the Technological Era, vol. 1, pp. 36-44, 2014.

[17] D. Oblinger, “The next generation of educational engagement,” Journal of Interactive Media in Education, vol. 8, pp. 1-18, 2004.

[18] P. Greenbaum, “The lawnmower man,” Film and Video, vol. 9, no. 3, pp. 58-62, 1992.

[19] N. Babich, “How virtual reality will change how we learn and how we teach,” Adobe, California, 2018.

[20] J. Niwhede and A. Lindgren, “VR from a learning perspective,” Jayway, Lund, 2017.

[21] Merck, “3D Virtual Lab App,” Merck, 2018. [Online]. Available: https://www.merckmillipore.com/ZA/en/support/mobile-apps/3d-virtual-lab/bjqb.qB.I1kAAAFUNtUSs1bx,nav?ReferrerURL=https%3A%2F%2Fwww.google.co.za%2F&bd=1. [Accessed 22 September 2018].

[22] TechTarget, “Augmented Reality (AR),” What Is, February 2016. [Online]. Available: https://whatis.techtarget.com/definition/augm

ented-reality-AR. [Accessed 22 September 2018].

[23] ARround, "Top 10 stories in the world of augmented reality you could miss," ARround, 28 May 2018. [Online]. Available: https://medium.com/@arroundsmm/top-10-stories-in-the-world-of-augmented-reality-you-could-miss-may-28-june-3-57bd1637b87e. [Accessed 22 September 2018].

[24] Tangible User Interface, "Tangible User Interface," Tangible User Interface, 2015. [Online]. Available: http://tui.umerfarooqofficial.com/. [Accessed 22 September 2018].

[25] "How does virtual reality improve industrial processes?," vr-bnb, 2017.

[26] Hero X, AR and VR making big waves in the manufacturing industry, 2017.

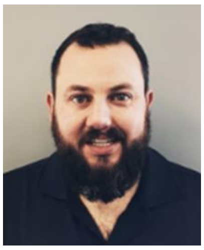
[27] M. Bennett, AIA, EDAC and LSSGB, "Enhancing healthcare facility design with VR/AR," Building Design and Construction, 2017.

[28] LTS Health, Viapath made use of LTS Power of Process Champion course to get staff ready for a major consolidation., Cape Town, 2016, p. 4.

10 BIOGRAPHY

Tanja von Leipzig holds a B.Eng degree in Industrial Engineering and an M.Sc degree in Operations Management. She is currently completing her M.Eng in Industrial Engineering at Stellenbosch University, with her specific area of focus and research being that of adaptive games for training and education. She is currently working for LTS Health where she is responsible for the integration of technology and games into the skills development training for laboratory professionals offered by LTS Health Learning.

Andre Gouws, CEO of LTS Health Learning has a strong technical background and an expert understanding of process and performance. He holds a degree in electrical engineering with an international diploma in teaching and training. He has also completed the Advanced Management Program at Wits Business School. He has a passion for skills development and the improvement of human performance. For over 25 years, Andre has managed training courses across multiple business disciplines and has been responsible for the skills development of more than 500 laboratory professionals.



Matthys Greeff obtained a B.Sc. degree in Human Life Sciences from the University of Stellenbosch. After starting his career at LTS Health he completed a Post-graduate diploma in Industrial Engineering, and is currently busy with his M.Eng., both at the University of Stellenbosch. At LTS Health he holds the position as the Principle Consultant for Innovation. He is also part of the design team to implement a secure global cloud platform, that complies with all major data security regulations.

Possibilities, Limitations and Considerations for Eye Tracking in Industrial Environments: Experience from a Case Study

A. Schramm[1], K. Kruger[2], J. Niemann[1], W. Grote[1]

[1] Department of Mechanical and Process Engineering, Duesseldorf University of Applied Science, Düsseldorf, Germany

[2] Department of Mechanical and Mechatronic Engineering, Stellenbosch University, Stellenbosch, South Africa

Abstract

Eye tracking is a measurement technique for determining the orientation and viewing direction of an individual's eyes. It allows for an evaluation of the individual's gaze and other related information. There is a wide variety of application areas for eye tracking, however the usage in engineering industry is still limited. This paper presents a case study for the use of eye tracking in a testing station for electrical circuit breakers. The case study aims to evaluate the measurement and analysis possible with the Tobii wearable eye tracking system. From the case study, possibilities, limitations and considerations for eye tracking in industrial settings are identified and discussed. The paper concludes that while eye tracking can be used to effectively capture and visualize insightful information, it faces major challenges related to subjectivity in analysis and labour intensiveness.

Keywords

Eye Tracking

1 INTRODUCTION

Eye tracking is a measurement technique for determining the orientation and viewing direction of an individual's eyes. Gaze direction, focus point and any information related to where the subject's visual attention is placed can be measured.

Modern eye trackers are mostly video based and use corneal reflections to determine the orientation of the eye-ball [1] [2]. Infrared is typically used to generate the light reflection of the cornea, which is then measured relative to the pupil centre [3]. The detection process is completely invisible to the wearer and the light source does not interfere with the person's sight. Analysis software processes the raw data to determine various events such as fixation, saccades and focus points. The data can be exported as video with depicted gaze movements overlaid or as a dataset. [1]

The system is calibrated by keeping the wearer's head stationary and having them switch their gaze between specified fixed points. The calibration software builds an accurate geometric model of the wearer's eye and the relative position of the measuring equipment. [1] [2]

There is a wide variety of application areas for eye tracking. Duchowski [3] lists some of the main fields of interest:

- Neuroscience and psychology – vision during reading, visual search, perception of art as well as vision in natural and virtual environments are common psychological investigations of vision.
- Marketing and advertising – eye tracking is mostly used for the evaluation of advertising effectiveness, but there are many opportunities within marketing. One example is the improvement of advertising materials through analysing advertisement placement. In this context, eye tracking has been extensively used with online applications, e.g. to investigate the effects of customer reviews and product image placement on an online social commerce platform [4].
- Computer Science – eye tracking is used for interactive applications, either as an input device or as a display system.

Eye tracking has also been used extensively on drivers [5] and pilots [6] [7]. These studies have aimed to extract information on the situational awareness, cognitive processes, performance indicators and the effect of experience, from the gathered eye tracking data.

Recently, there has been some exploration into the application of eye tracking in the manufacturing industry. Various engineering sectors, such manufacturing, value the monitoring of the safety, speed and quality of manual labour. The analysis of eye movements can give access to visual, cognitive and attentional aspects of human performance. [3]

One of the world leaders in eye tracking equipment, Tobii, used their eye tracking glasses for a case study to increase efficiency, and improve training and safety within a metal foundry [8]. Niemann *et al* [9] performed a case study to improve the final quality control process of a paint shop within a transporter production line, and Yousefi *et al* [10] tested if users' satisfaction of construction design is related to their visual attention. Lušić *et al* [11] investigated the impact of productivity and added value within a manual assembly. For other interesting examples and further reading on this topic, the reader is referred to [12].

Despite the potential value, the usage of eye tracking in industrial environments is still rare. This paper presents a discussion on the possibilities and limitations of eye tracking in industrial environments, based on the experience obtained from an industry-inspired case study. The experience obtained from the study is used to identify important quantitative and qualitative aspects that must be considered for future studies.

Section 2 presents the case study with a discussion of the experimental setup and procedure, and some results from the analysis. This provides a platform for a discussion of the possibilities and limitations of eye tracking in industrial environments, and considerations for future studies, which follows in section 3. The paper concludes with section 4.

2 CASE STUDY: INSPECTION OF ELECTRICAL CIRCUIT BREAKERS

A South African circuit breaker manufacturer incorporates an electrical test station in its production line. At this electrical test station, a worker places circuit breakers into a rack of automatic test devices and removes them again upon test completion. The worker activity at the electrical test station is used as a case study.

The case study aimed to analyse the eye tracking data from a worker at the station to obtain information related to worker performance (specifically, speed and accuracy). The electrical test station was replicated in a laboratory environment, which was used for conducting the eye tracking experiments.

For the analysis, the implications of following different detection strategies was investigated. Two participants were instructed to adopt different detection strategies; Participant 1 (P1) performed a continuous circular scanning path strategy (shifting focus between the testing devices in a clockwise direction), and Participant 2 (P2) focussed on a central point between the devices. The difference in detection strategy is clear in the gaze plots shown in Figure 2.

2.1 Experimental Set-Up

The test rack consists of a panel, shown in Figure 1, which is mounted to a frame. The panel holds a 2x4 array of testing devices, where each device simulates the testing of a circuit breaker. Next to each testing device are red and green LEDs that indicate the test status (idle/empty or busy) and result (pass or fail).

The LEDs are controlled by a Raspberry Pi running a Python script. The Python script simulates the execution of every testing device, switching the LEDs to indicate the simulated status and results. The simulated tests are set to execute for 30 ± 10 seconds (randomly varied), where after a test result is also randomly generated. At this stage, a pass result is then indicated by the green LED and a fail result by the red LED – the worker then removes the circuit breaker from the device and places it in "successful" and "unsuccessful" bins. During the test process (i.e. when a circuit breaker is placed in the slot of the testing device), both LEDs are switched on to indicate that the device is busy performing a test. If a testing device does not have a circuit breaker inserted, both LEDs are off to indicate the idle state.

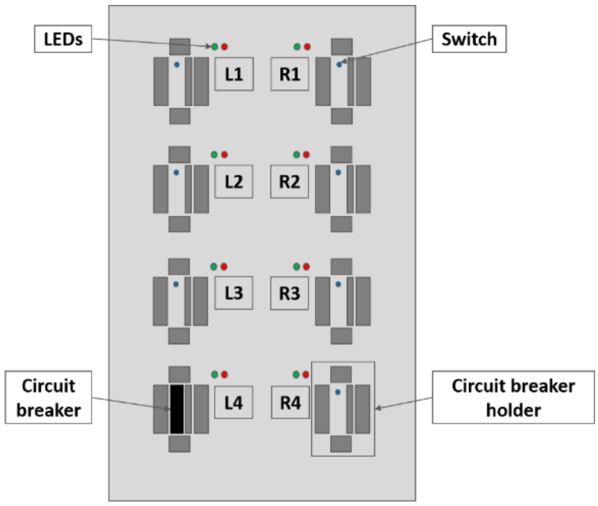


Figure 1 - Testing rack setup

The gaze point data of a worker during the testing activity is captured by a wearable eye tracking system – specifically, the Tobii 1 glasses. The analysis is performed using the associated software, Tobii Studio 3.2.1. Along with capturing the gaze information from eye movement, the eye tracking glasses also record the field of vision of the participant with a forward-facing scene camera and sound using a built-in microphone.

2.2 Experimental procedure

Before conducting the experiment, the participants were fitted with the eye tracking glasses and performed a standard calibration procedure [2]. After calibration, they were instructed to stand in front of the panel at a distance of 1.5m – this distance was used to ensure that the entire test rack fits within the field of view of the participant for the duration of the experiment.

The experiment was then started – the simulated testing process was started with a circuit breaker placed in all testing devices (i.e. all testing devices start the process simultaneously, at the start of the experiment). The participants were instructed to detect the result of each testing device on the panel, as they are indicated by the switching of the LEDs, and then immediately communicate the test result verbally. The completion of testing at all the testing devices is referred to here as a single *test run*. The experiment was conducted for 30 minutes of successive test runs.

All the participants were given music to listen to over ear-phones during the recordings in order to avoid distraction from the relays that switch the lights on and off.

To simplify the study and to adapt it to the limitations of the eye tracking glasses (as discussed further in section 3), the circuit breakers remained in the testing rack throughout the experiment. When inspecting the eye tracking recordings for the activity of inserting and removing circuit breakers from the testing devices, it was observed that resulting the head movement was detrimental to the data quality and made further analysis impossible.

2.3 Analysis

The analysis only considers data captured from two participants and is not the focus of this paper. Instead, an analytical approach is described for demonstrative purposes – providing a platform for the discussion in section 3.

The data captured from the experiments was analysed using the Tobii Studio software. The software presents the captured data as a video of the field of vision (obtained from the scene camera), overlaid with the corresponding gaze points of the participant. The sound recording, obtained from the microphone, is also synchronised with the video.

The analysis of the recordings is aimed at obtaining two measures:

- The reaction time of the participant, i.e. the time between when the testing device indicates a result and when the participant's gaze focusses on the device's LEDs.
- The accuracy of the participant in identifying and communicating the result of a test device.

The analysis of reaction time starts with the manual insertion of events in the captured data recording, in Tobii Studio. From inspection of the recording, an event is inserted at the moment of time that a testing device finishes a test (i.e. when the result is indicted by switching the LEDs). Areas of interest are then defined within the video recording, which is done using functionality provided by the software. The software then detects when the gaze point of the participant enters these areas of interest – this detection is then automatically recorded as an event. The timestamps of the aforementioned events are exported to Microsoft Excel, where the reaction times are calculated.

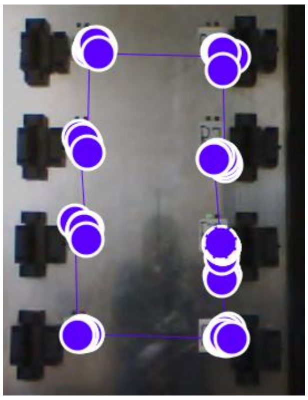
To determine the accuracy of the participants in identifying and communicating test results, the audio recording of the participant vocalizing the observed result is compared to the video recording. The analyst compares these sources for the entirety of the experiment recording.

2.4 Preliminary results and observations

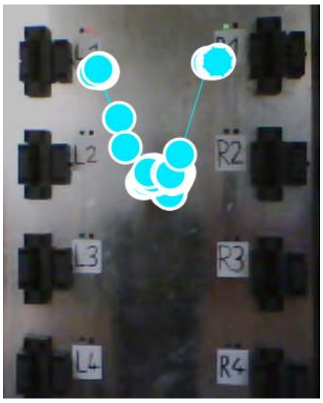
After analysing the recorded data, it was found that the average reaction time of P1 decreases within the first 30 minutes In contrast, there is a slight increase of the reaction times calculated for P2.

In analysing the accuracy of detection, it was found that P1 misinterpreted 4 results (two of which was immediately corrected) and P2 incurred eleven misinterpretations (of which all but one were corrected immediately). These misinterpretations occurred when either the corrected station is recognized but the verbal confirmation of the result was incorrect, or when the result was correctly communicated but the testing devices was wrongly identified.

In the data used to calculate reaction time, a high number of outliers were observed – specifically for P1, who followed the circular scanning strategy (as is shown in Figure 2(a)). The outliers of significantly longer reaction times occurred when the a testing device generated a result right after the participant shifted their focus to the next device – the result was then only noticed when the full scan path was completed. Alternatively, when a testing device generated a result just as the participant shifted focus to it, resulted in significantly shorter reaction times. It was also found that the participants were occasionally confused with the order of the inspection when two stations finished testing nearly simultaneously. The time to notice the result of the second station was then significantly higher compared to the station which was recognized first.



(a)



(b)

Figure 2 - Gaze plots for participants (a) P1 and (b) P2.

Considering the presented results and the discussion on data outliers above, it is clear that the study is not conclusive. In this context, the following observations are offered:

- Training/familiarization and fatigue may have an effect on reaction time. The reaction time of P1 improved over the duration of the experiment, while that of P2 deteriorated – further experiments are needed to establish the circumstances where these factors come into play.
- There are several factors that contribute to outliers in the data: scanning strategy, area of interest placement, randomization of testing time, etc. This indicates that the experiment must be further refined to reduce these outliers and produce more reliable data.
- While verbal confirmation of a detected test result indicates that the participant identified the result correctly, it does not ensure that the participant would place the tested circuit breaker in the correct output bin (i.e. the “successful” or “unsuccessful” bin). The experiment should thus be extended to include the removal activity.

3 POSSIBILITIES AND LIMITATIONS OF THE EYE TRACKING GLASSES AND SOFTWARE

The following section identifies some possibilities, limitations and considerations for future studies on the use of eye tracking in industrial settings, based on the observations and experience obtained from the case study of section 2.

3.1 Possibilities

While the presented case study uncovered several challenges, it also illustrated the possibilities for using eye tracking in industrial settings. It is evident that eye tracking can extract interesting information regarding the behaviour of workers when performing their activities. While the study focussed on investigating the effect of scanning strategy on worker performance, it could easily be adapted to explore the effects of lighting, workspace configuration, worker fatigue and many other aspects.

In conducting the case study, the authors identified two factors that should encourage further exploration with eye tracking systems: the ease of use and the functions to informatively represent data.

3.1.1 Ease of use

After reading through the manuals, the use of the Tobii glasses and software is simple and intuitive. It is easy to insert events in the recordings, which enables some automatic analysis. The extracted timestamps can then be exported to an Excel file, so that further computational analysis can be performed.

3.1.2 Data representation

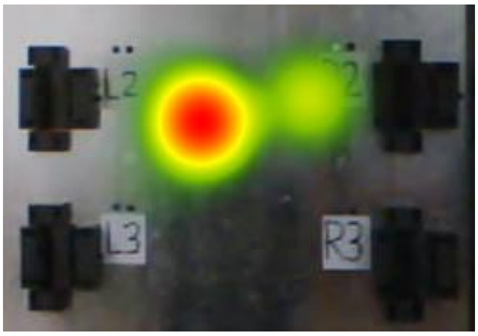
The software is able to visualize the recorded data, e.g. providing functions for generating heat maps and gaze plots.

Gaze plots, which overlay the gaze point data on the recorded field of view, are especially useful to visualize the scan paths of participants (as demonstrated in Figure 2). With this, the actual path can be compared to anticipated (or instructed) gaze paths, which can be useful to improve the procedures of visual inspections.

Heat maps represent the number and position of gaze fixations in the field of view, over the recorded time [13]. Figure 3(a) illustrates that P1 keeps the gaze fixed at unfinished stations approximately for the same duration before continuing the scan path. Figure 3(b) clearly reflects the central focus point scanning strategy of P2.

This data representation can provide clear and intuitive visualizations, which should facilitate better communication and improve understanding.

(a)



(b)

Figure 3 - Heat Map examples from (a) P1 and (b) P2.

3.2 Limitations

The experience from conducting the case study has led to the identification of several limitations to the use of eye tracking in industrial settings. The following sections elaborate on some of most important limitations.

3.2.1 Calibration issues

The calibration of the glasses proved to be cumbersome and unreliable. The glasses must be calibrated every time that it is put on again and the process requires the aid of an assistant (i.e. the participant struggles to perform the calibration on his/her own). The calibration process requires the participant to focus their gaze on several points with specified coordinates. It was found that the lower left focus point was often not detected during this procedure, requiring the calibration to be re-done. Changing the light conditions, as well as the background, did not solve this problem. It was found that the calibration could be improved by instructing the participant to open their eyes as wide as possible during the procedure.

3.2.2 Restriction on head movement

In the early stages of the case study, the insertion and removal of circuit breakers was included in the experiment. However, the analysis showed that the fast head movements made the recordings almost unusable, since the pupils would not be tracked accurately during movement.

3.2.3 Field of view

The scene camera, which records the field of view, was observed to have a slight downward-angling orientation. Initially, this offset led to recordings where the desired field of view was not captured. To ensure that the camera records the desired area, it was necessary to continuously monitor the recording via the display of the controller of the Tobii glasses. The participants could then immediately be instructed to correct their pose so that the desired field of view would be recorded.

3.2.4 Labour intensiveness

All the recordings needed to be analysed manually and carefully watched multiple times to insert events and observe abnormalities in the recording, and verify the results of the analysis. For every 30 minute recording used in the case study, 6-7 hours of analysis was required.

3.2.5 Subjectivity and lack of benchmarks

As the events have to be put in manually, the analysis dependents on the analyst – this inherent subjectivity may affect the results as every human has a different perception. The same problem influences the identification of special occurrences (or abnormalities) in the recording. The detection and interpretation of these occurrences are then also subjective. The possible inconsistency can be negated by employing only one analyst for an entire case study, but it does not facilitate a reliable comparison between case studies.

The procedure of the analysis, as well as the results, are very case specific and a unique approach must be developed for every new case. In general, eye tracking research suffers from a lack of standardizations for the reporting of key figures. It is thus challenging to evaluate the quality, reliability and repeatability of results [2].

3.3 Considerations for future studies

As mentioned, the case study had to be adapted to accommodate the limitations of the eye tracking system. As already mentioned in section 3.2.2, the head movements had to be restricted, so that the eyes can get tracked properly. In addition, the participants needed to stand a certain distance away from the testing panel in order to record the whole array of stations. The time consuming manual analysis meant that the recordings were limited to 30 minutes.

In general, the two most important aspects to consider for future eye tracking studies are related to the nature of the study and the capabilities of the technology.

3.3.1 Suitability of study

The paper has already highlighted the adjustments that were made to the case study to allow for a sensible analysis. The limitations of the eye tracking system to be used should thus be considered when proposing of selecting a case study.

It is suggested that the suitability of studies be evaluated according to the following aspects:

- Limited head movement is required.
- The position of the participant should allow the area/object of interest to fit into the scene camera's field of view.
- Studies should have a narrow scope to enable an effective analysis.
- Experiments should be short and focussed.
- The objectives and key measures of the analysis should be well defined to reduce analyst subjectivity.

3.3.2 Eye tracking technology

The Tobii glasses 1 system is quite dated and newer technology offer some improvements and additional functionality. For example, the Tobii glasses 2 system includes a high-resolution scene camera, an additional eye tracking camera and built-in accelerometer-gyroscope sensor. The recording device can also stream the recording data over a wireless network, and the software includes some new functionality. However, new systems require a substantial capital investment. The requirements and specifications of the study should be carefully considered when selecting the eye tracking system to use.

4 CONCLUSION

Despite the potential benefits, like the evaluation of person's performance while completing a task, the usage of eye tracking in industrial environments are still rare. Therefore, this paper documents a case study of the inspection of electrical circuit breakers and points out the possibilities and limitations of the used eye tracking technology. The case study shows that the eye tracker is very easy to use, the software provides informative visualization of the recorded data and the analysis results are easily exported to common data formats. However, several limitations were identified – the labour intensiveness and subjectivity of the analysis are especially significant. The paper also suggests that the suitability of the experiments and the choice of eye tracking technology be considered in future studies of eye tracking in industrial settings.

5 REFERENCES

[1] C. Balke, "Eye-Tracking: Grundlagen und Anwendungsfelder", in *Handbuch standardisierte Erhebungsverfahren in der Kommunikationswissenschaft*, W. Möhring e D. Schlüz, Edits., Hannover: Springer, 2013, pp. 367-387.

[2] K. Holmgvist, M. Nyström, R. Andersson, H. Jarodzka and J. van de Weijer, *Eye tracking: A Comprehensive Guide to Methods and Measures*, Oxford: Oxford University Press, 2011.

[3] A. Duchowski, *Eye Tracking Methodology - Theory and Practice*, 2nd ed., London: Springer, 2007.

[4] P. Mikalef, K. Sharma, I.O. Pappas and M.N. Giannakos, "Online Reviews or Marketer Information? An Eye-Tracking Study on Social Commerce Consumers", in *Conference on e-Business, e-Services and e-Society,* Cham: Springer, 2017, pp. 388-399.

[5] B. Kapitaniak, M. Walczak, M. Kosobudzki, Z. Jozwiak and A. Bortkiewicz, "Application of Eye-Tracking in the Testing of Drivers: A Review of Research", *International Journal of Occupational Medicine and Environmental Health*, 2015.

[6] G. Ziv, "Gaze Behavior and Visual Attention: A Review of Eye Tracking Studies in Aviation", *The International Journal of Aviation Psychology,* 2016, Vol. 26, No. 3-4:75-104.

[7] S. Peißl, C.D. Wickens and R. Baruah, "Eye-Tracking Measures in Aviation: A Selective Literature Review", *The International Journal of Aerospace Psychology*, 2018, pp. 1-15.

[8] Tobii, "Press release: New Eye Tracking Research Tackles Safety and Performance in Manufacturing", 2017.

[9] J. Niemann, M. Schlösser and C. Fussenecker, "Eye Tracking usage as a Possible Application to Optimize Processes in the Engineering Environment", in *International Conference on Competitive Manufacturing*, Stellenbosch, South Africa, 2016.

[10] M. V. Yousefi, E. P. Karan, A. Mohammadpour and S. Asadi, "Implementing Eye Tracking Technology in the Construction Process", in *51st ASC Annual International Conference Proceedings*, Texas, 2015.

[11] M. Lušić, C. Fischer, K. Schmutzer Braz, M. Alam, R. Hornfeck and J. Franke, "Static versus Dynamic Provision of Worker Information in Manual Assembly: A Comparative Study using Eye Tracking to Investigate the Impact on Productivity and Added Value Based on Industrial Case Examples", in *49th CIRP Conference on Manufacturing Systems*, Stuttgart, Germany, 2016.

[12] J. Niemann and C. Fussenecker, "Analysis of Eye Tracking Usage in Different Domains and Possible Applications in the Engineering Environment", in *International Conference on Production Research - Africa, Europe and Middle East 3rd International Conference on Quality and Innovation in Engineering and Management*, 2014.

[13] Tobii, “Tobii Studio User`s Manual Version 3.4.5”, 2016.

6 BIOGRAPHY

Alexandra Schramm holds a BEng degree in industrial engineering from the Baden-Wuerttemberg Co-operative State University. She is currently a postgraduate student at the University of Applied Science, Düsseldorf.

Karel Kruger obtained his PhD from Stellenbosch University, where he is now a lecturer at the Department of Mechanical and Mechatronic Engineering and co-leader of the Mechatronics, Automation & Design research group.

Jörg Niemann obtained his PhD in manufacturing engineering from the University of Stuttgart. He is currently a Professor at the University of Applied Sciences, Düsseldorf, and head of the FLiX Centre for Life Cycle Excellence.

Wolfgang Grote received his PhD in mechanical engineering from Ruhr-Universität Bochum. He worked for MAN Energy Solutions SE and is now a Professor at the University of Applied Sciences, Düsseldorf.

Let's Put the V in Smart Factory: Empowering Employees to Shape a Safety Culture for Industry 4.0

C. Digmayer[1], E.-M. Jakobs[1]

[1]Chair of Textlinguistics and Technical Communication, Human-Computer Interaction Center, RWTH Aachen University

Abstract

Industry 4.0 is the concept for the smart factory of the future: Digitalization will not only revolutionize the way in which industrial work is carried but also entail new types of risks that are to date difficult to predict. Current transformation approaches towards Industry 4.0 lack concepts that offer employees an active role in the transformation process – especially with regard to risk management. This paper presents a study investigating the potential of empowering employees in shaping a safety culture for Industry 4.0 for small and medium enterprises. The present work builds on the business transformation approach by [1] which suggests that transformations towards Industry 4.0 can only be successful if employees are integrated in the change process with top-down and bottom-up approaches. The study combines expert interviews with insights from industrial process modelling methods. In a metalworking company, a partly digitalized production process chain was examined. The focus was on perceived risks in the context of the production chain and the potential and willingness of workers to improve the safety and security of their work. The results indicate a high reflectiveness as well as a high willingness to contribute to transformation processes. The participants name a broad range of need for action. A highly ranked risk area concerns information flows. The participants emphasize the need for means that enhance information flows horizontally (e.g., along the production process) as well as vertically (e.g., between the management and the workers) and suggest various measures to improve both, safety in production processes and information flows.

Keywords

Industry 4.0, industrial digitalization, safety culture, employee empowerment

1 INTRODUCTION

Digitalization rapidly changes the modern industrial world. ‚Industry 4.0' has been intensely discussed in Germany as future concept for modern factories. Similar concepts are known in other countries (such as ‚Industrie du futur' or ‚Smart Factory'). A vast body of literature focuses on Industry 4.0 and the underlying change process. However, concepts for safety cultures in Industry 4.0 have so far been addressed only marginally. The existing research about safety and safety cultures mainly relates to the situation before Industry 4.0. As the transition towards Industry 4.0 revolutionizes technological production concepts (cyber-physical systems) and assigns novel tasks to employees, the main challenge is to establish a safety culture for Industry 4.0 following the basic assumption that transformation processes and Industry 4.0 will only succeed if the people are taken along [1]. In this context, the role of the employee will change significantly towards a decision-maker responsible for one's own work.

The interdisciplinary research project SiTra4.0 ("Nachhaltige Sicherheitskultur als Transformationsansatz für Industrie 4.0 in KMU") addresses this issue by developing a transformation approach for the establishment of a preventive, participatively acquired and lived safety culture concept as a decisive success factor for the implementation of Industry 4.0 in small and medium enterprises (SME). The approach is to develop behavior-influencing means of communication and participation that allow SME to initiate a transformation of the corporate culture with regard to aspects such as value structures, perceived safety or informal behavior. Substantial sub goals are the development of methods that enable SME to monitor and improve the current status of 'lived' safety cultures as basis for the development of a safety culture 4.0, e.g. by empowering employees in taking an active part in the transformation process towards Industry 4.0.

This paper focuses on the following research questions:

1. Are employees willing to contribute to the development of a safety culture?
2. Are employees capable of reflecting their work environment and thus able contribute to the development of a safety culture?

Selected results of a study investigating the willingness and potential of employees to shape their work environments for a digitalized industry are presented in this paper. The results exemplify how the systematic integration of top-down (means initiated by the management) and bottom-up (means of employee empowerment) approaches foster the establishment of safety cultures in Industry 4.0.

2 THEORETICAL BACKGROUND

2.1 Occupational safety in Industry 4.0

Various publications examine consequences of industry 4.0. A large body of research addresses the question which potential risks will emerge in Industry 4.0. Several risks are suggested to be consequence of the interconnection of systems and machines [2,3], the increasing use of digital systems [4,5] or issues of managing vastly increasing amounts of digital data that needs to be integrated and analyzed in industrial production process [6,7].

The transition towards a digitalized industry imposes significant changes on the interplay between humans and machines [8] resulting in novel demands on the way how work is carried out in safe and secure ways [9], it requires a rethinking of the current understanding of occupational safety (e.g. safety as the absence of undesirable events and consequences [10]) and related concepts such as health protection. It has to be considered that occupational safety is only one facet embedded in a complex framework of mutually influential effects in socio-technical systems such as RAMSS - reliability, availability, maintenance, safety and security [11].

Consequence of the progressing digitalization are novel challenges for occupational safety and health protection that especially hinder SME at economizing digitally. What is missing are transformation approaches that combine technological changes with forms of responsible self-organization by employees in the everyday professional life that consider the various degrees of ‘digital maturity’ in industrial domains.

2.2 Safety cultures

‘Safety cultures’ comprise as attitudes and behaviors towards safety aspects in organizations as part of the corporate culture [12]. The concept of safety cultures is regarded as key element for the further development of the ‘classical’ understanding of occupational safety and health protection. Future concepts for safety need to consider which type of safety culture is required that fits the various demands of Industry 4.0.

Existing definitions conceptualize safety culture as ongoing social process, as result of interaction and mutual negotiation processes [13,14] or as manageable characteristic of companies with various gradations [15,16]. Safety cultures have been object of numerous investigations. However, few studies examine whether safety-culture concepts are adaptable in the transformation towards Industry 4.0.

Scientific publications addressing industrial transformation processes emphasize the meaning of *change* for releasing the full potential of technological innovations in enterprises [17]. Organizational change is a controlled process to increase the competitiveness of enterprises. Transformation approaches stress the significance of the employee and its potential to successfully initiate change in organizations [1]. Safety has to date been a less considered topic in this field.

In this paper, safety culture is conceptualized as approach that merges top-down and bottom-up processes. Bottom-up processes include empowerment approaches which offer employees “[…] the ability to identify and express needs, establish goals or expectations and a plan of action to achieve them, identify resources, make rational choices from various alternative courses of action, take appropriate steps to pursue objectives, evaluate short-and long-term results (including reassessing plans and expectations and taking necessary detours), and persist in the pursuit of those goals.” [18].

Essential for the consolidation of an organizational safety culture is its linguistic representation, e.g., in documentation) as well as its negotiation in discourses. Informal communication plays an important role in managing operational processes, e.g., behavioral patterns adhering to safety demands [19]. However, few studies address communicative and behavioral aspects for Industry 4.0 and related transformations processes.

Of major importance are options for actions as experienced by the employees [20]: Safety is not only a result from knowledge and insights, (apparent) rational behavior and high transparency [13] but also consequence of intuitive understanding and perceived involvement in decision-making processes. Employees need to participate actively in developing, discussing and ratifying safety concepts [21]. Employee empowerment presupposes self-responsibility, a common way of handling occupational stress, uncertainty and errors. Required are room for novel ideas and solutions as well as ‘shared language’ for both, known and new issues. Thus, safety results from the constellation of ‘want to, can do, allowed to" [22].

According to [13], organizational (safety) cultures manifest on three layers: the layer of artifacts (observable behavior), the layer of expressed values, and the layer of basic assumptions. The third layer represents a safety culture’s core for the influence on behavior: Basic assumptions are implicit and direct the behavior of employees, the results of such behavioral patterns become observable on layers one and two.

[13] recommend that processes of mutually understanding safety in an organization should focus on developing a shared language first to make implicit standpoints explicit. Secondly, standpoints need to be covered linguistically and transferred into documents. To achieve this, novel multi-dimensional

strategies and methods are required that combine the qualitative with quantitative means of analyzing safety cultures and consider technological changes over time.

3 METHODOLOGICAL DESIGN

3.1 Use case

The study was conducted with an industrial partner - a metal manufacturing SME currently implementing Industry 4.0 principles. For the study, a partly automatized and digitalized process was selected as use case (production of a particular structural element used in ventilation systems) that comprises 11 steps and 7 departments.

3.2 Methodology

The study used a combination of methods: participatory observation, document analysis, and interviews with employees. This paper focuses on outcomes of the interviews that included a 'worker's journey' adapting methods of industrial process modelling to survey the current state of production chains.

Data collection: An interview guideline was developed. It contains four parts addressing the following topics: expertise and tasks of the interviewees, perceived information flows and perceived level of information, perceived problems (in the production chain, in their daily work, at the interfaces between steps of the production chain), willingness to contribute to a safety culture. 10 interviews were conducted with representatives from the departments involved in the process chain (8 male, 2 female; age: 29 to 63 years, ∅ 45.6 years; experience in the company ranges from 7 to 38 years, ∅ 22.4 years). The interviews took on average one hour and were audio recorded with the agreement of the interviewees. Interviewees were shown a flowchart of the process chain and asked to situate themselves and their tasks within the chart.

Data preparation: The audio files were transcribed and anonymized. The anonymized transcripts were converted into MAXQDA format.

Data analysis: The MAXQDA files were analyzed qualitatively (content analysis).

4 RESULTS AND DISCUSSION

In the following, selected results are presented and illustrated by comments of the interviewees. The abbreviations are to be read as follows: Interview number_department of the interviewee.

4.1 Willingness to contribute to transformation processes

The results indicate that the interviewees are highly motivated to contribute to improvements of the existing safety culture. The reasons for employee engagement differ and include extrinsic (e.g., to adapt their work place to changing conditions and requirements, and by doing so to create a safe and well-functioning work environment) as well as intrinsic motives (e.g., to contribute to the company's future and thereby safeguard jobs). Another outcome is a general interest in co-creation: The interviewees state that they would like both, to contribute own ideas as well to reflect and comment on the ideas of other colleagues. However, they expect to be involved actively (by superiors and the management) and that their potential contributions are valued (perceived interest, listening, and appreciation). Thus, willingness is not enough – the company must be able and interested to respond to contributions and to sustain this motivation over time. Motivation is easily lost – especially if employees get the feeling that their ideas are ignored or do not lead to results:

I1_manufacturing: Recommendations for improvement are a fine thing, but someone needs to be interested in them. It is no use for me to repeat the same thing for five years: I want this and that to change. But these things are not changed in five years and put aside.

Others report missing feedback. They don't know whether or not, how, and why their ideas are put into practice and what is the outcome of the idea realization. Other reasons for frustration and resulting des-interest to contribute arise due to perceived inaccessibility of decision makers (the worker is not able to discuss his idea with superiors or managers) or if an idea is prevented by others (e.g., superiors):

I2_manufacturing: Several proposals were handed in from other departments: These things need to be improved. But they told us that it is not paying off or that things are carried out in a different way in this department. In the end, nothing changed.

For safety cultures it is especially important to make employees feel like being listened to and taken seriously. As studies show, changes in behavior are expectable if they are initialized by the employees themselves and if the employees are motivated by the management to act on their own responsibility [23].

I5_assembly: We contribute if we are asked to do so. At the moment, we work overtime to move our company forward. We really like to do that. Not for us. We work for our company.

4.2 The potential to contribute to transformation processes

The results indicate that the majority of workers has high potential to make a difference in the development of a safety culture for Industry 4.0. Their comments addressing work safety indicate a high level of reflection and sensitiveness regarding changing work environments and their consequences as well as the ability to create solutions for a new safety culture. Their understanding of safety exceeds 'traditional' concepts of occupational safety and health protection and comprises aspects such as the handling of data as well as the quality of information and communication flows. They are able to describe gaps between intended corporate safety issues and actual behavior patterns as well as organizational roles and regulations.

The interviewees address different safety areas, e.g. the quality of work instructions (Which instructions fit current requirements, which ones need to be improved?). In the following example, an interviewee recommends informing workers systematically about customer complaints and by doing so to help them to avoid errors:

I3_assembly: Regarding customer complaints it is necessary to tell us which problems have occurred so far and to which things we need to pay attention to so that such errors do not happen again.

Other interviewees reflect on how work is organized and what can be improved:

I2_manufacturing: There are some things that need to be changed. For instance regarding the arrangement of storage areas. Nothing is regulated to 100 percent and often things are handled differently from necessity. It does not run the way it should. To improve this, the production department should be informed about the planned sequence of components – which components need to be taken out first, which ones second, and so on. This requires clear arrangements between the departments.

Other interviewees reflect on boundary conditions influencing how work is carried out (e.g. weather conditions, order situation). Others relate to the collaboration of departments and the need for a better inter-divisional communication:

I6_work preparation: There are several possibilities to improve communications in our company. We have the necessary tools such as e-mail or instant messengers. We should use them to speed up the communication between departments. For instance, if there are postponements, changes to components, production stops or shortened cycle times which could lead to potential problems.

The broader understanding of the interviewees regarding safety goes beyond their tasks, roles, and areas of responsibility. The ability to reflect on the work environment at large scale has high potential for the implementation of a safety culture – especially with regard to the improvement of information flows, which has been a frequently discussed problem area in the interviews.

4.3 Information flows as subject of safety cultures

Many participants emphasize the need for means that enhance information flows horizontally (e.g., along the production process) as well as vertically (e.g., between the management and the workers).

4.3.1 Horizontal information flows

Horizontal information flows span along production chains and address the organization of work. Several interviewees state a high interest in improving the existing management and quality of information flows. The interviewees emphasize that this aspect is highly relevant for the quality and outcome of their daily work, e.g., in terms of speed and safety.

I1_ manufacturing: Relevant information in adequate quality need to be communicated along the production chain. From the beginning to the end. It does not suffice if the work preparation of one department just communicates with the work preparation of the other department – every worker needs to get such information.

One often mentioned shortcoming of the horizontal information flow is a perceived low level of information. Often the options to influence the efficiency of one's own work are perceived to be decreased by missing forecasts and missing follow-up information.

I1_ manufacturing: My problem is that I do not have an overview over the incoming orders for the next days. Thus, I frequently produce small quantities of components. Always one piece, one piece, one piece. I need to retool the machine for each piece which also entails a higher amount of paperwork.

The workers expect to be informed about errors in the production chain to get a better understanding of the current state of work organization. Missing follow-up information leave workers in uncertainty whether errors occurred in the following steps of the production chains leading to reclamations and changes in production schedules.

The interviewees not only reflect on weak points of horizontal information flows but also on means to change the current situation. Such means include actions initiated by the management (top-down) and actions initiated by the employees (bottom-up). From a top-down perspective, the management should initialize efforts to gather information in just one 'single point of truth'. From a bottom-up perspective, employees should commit to use one location for all information flows and discontinue alternate communication practices (e.g., passing handwritten information to colleagues). Studies show that the design and implementation of highly accepted information platforms require a deeper understanding of the work tasks, use situation and user requirements [24] as well as the willingness of employees to contribute to the design of an information platform and its functions. The interviewees have a clear understanding of how such a system should be designed:

I8_sales: A system should offer the possibility to collect the information for all of our products and store attachments. For instance: A link to the technical drawing. Who created the product description? When has the description been created? How does the *parts list look like? How does the drawing of previous versions look like? How is the processing time for this product organized? How is it packaged, how does it look like once it is packaged? Which issues can occur in the production process?*

A main function of the information and communication platform is to provide particular information for steps in production chains and related roles. Die interviewees demand one system as central storage location where they can find all relevant digital data for their work (instead of various dispersed information 'silos'). The digitalization is perceived as chance for gathering work-related information:

I3_assembly: Work places should be digitalized. I would like to have a work place with a display or a tablet which allows me to click on an order and receive all relevant information: What do I need to take care of, how do I need to package components, which reclamations occurred so far.

4.3.2 Vertical information flows

Vertical information flows connect the management and employees. They comprise different topics, e.g., the development of the company. Again, most of the interviewees show a high interest in improving vertical information flows in terms of an ongoing dialog and a stronger organizational listening.

I4_ manufacturing: Several colleagues wish that experienced employees are being listened to about what went wrong in the past and that they are not just told: 'Go to work!'

The employees want to be informed about ongoing processes and the company as a whole, e.g., decision-making processes about future developments of the company.

I8_work preparation: Who has been hired, who is leaving, what is the company currently doing, which things work well, which ones do not, etc. We do not need such information every month, but at least every half year. And the information should not be published by anyone – this needs to be done by the management and should not be delegated.

The interviewees propose a combination of bottom-up and top-down approaches to address this issue: From a bottom-up perspective, employees should request innovation workshops in which representatives of different departments and the management work on previously defined tasks (e.g., safety innovations) and by doing so develop a shared understanding of important topics. From a top-down perspective, representatives of the management should act as mediator and decision-maker to ensure that workshop results are put into practice.

I8_sales: Negotiations in such workshops will often not succeed unless someone from the management is involved and decides: We solve this problem in this particular way. Decision. We do it like this. Even if you do not agree, three other people think it is a good idea and we do it like this.

4.4 Implications for the development of a safety culture 4.0

Employee empowerment can be put into practice because employees are both willing and able to contribute to transformation processes. Willingness to contribute is the foundation for developing and establishing an adequate safety culture for Industry 4.0. In the case study, comments on ignored willingness indicate that bottom-up approaches are alleged but not 'lived' throughout the company. To develop an empowerment approach, emotional factors need to be considered such as identifying with one's own work (proudness) and with one's own company by fostering a perceived high level of information.

As the examples show, isolated attempts to solve safety issues emerging by the progressing industrial digitalization fall short – the development of a safety culture 4.0 requires top-down *and* bottom-up approaches to interlock systematically.

Of major importance are measures at the layer of communication and interaction that control and foster such efforts. Executives are expected to create an environment that fosters and supports the creation and discussion of shared values, e.g., values regarding safety. Codes of conduct need to be defined and communicated in adequate formats. The foundation for putting the vision of a safety culture 4.0 into practice needs to be laid (e.g., by providing spaces for co-creation activities). Employees need to actively facilitate safe attitudes and behaviors as well as to contribute to transformation processes.

The combination of expert interviews and worker's journey methods of industrial process modelling seems to be a promising approach to identify gaps between intended corporate safety issues and actual behavior patterns as well as shortages in information flows providing valuable insights for possible integrated top-down and bottom-up approaches that can be applied in transformation processes.

5 CONCLUSIONS

The transition towards an Industry 4.0 goes hand in hand with a changing understanding of safety: The range of familiar issues of work safety is extended by the progressing digitalization leading to novel problems that go beyond the traditional understanding of occupational safety and health protection. Such issues require carefully designed empowerment solutions that consider top-down and bottom-up approaches as identified in the present study.

However, there is need for further research, e.g., studies investigating other companies, production chains, and industries. Open questions relate to the influence of factors such as the given industry, the corporate culture, the level of digital maturity or the size of the company. Value creation chains of other domains and related process chains, roles, tasks, and work contexts need to be investigated. Insights may be gained by examining role-specific value creation chains [25]. Further research should consider the question how changing safety conditions in an Industry 4.0 impact professional communication as well as documentation requirements and formats and which consequences result for the education and training of employees, e.g., with regard to human-machine learning (see [8]).

6 ACKNOWLEDGEMENTS

This research and development project is funded by the German Federal Ministry of Education and Research (BMBF) within the "Innovations for Tomorrow's Production, Services, and Work" Program and by the European Social Fund (ESF) (funding number 02L15A002) and implemented by the Project Management Agency Karlsruhe (PTKA). The author is responsible for the content of this publication.

7 REFERENCES

[1] Gudergan, G., Feige, B. A., Krechting, D., 2017, Ordnungsrahmen für den Prozess der Business-Transformation, Pollety, W., Blaeser-Benfer, A., Digitalisierung. Betriebliche Handlungsfelder der Unternehmensentwicklung, Frankfuter Allgemeine Buch, Frankfurt, 155-172.

[2] Sopori, D., Pawar, T., Patil, M., Ravindran, R., 2017, Internet of Things: Security Threats, International Journal of Advanced Research in Computer Engineering & Technology, 6/3:263-267.

[3] Jaradat, O., Sljivo, I., Habli, I., Hawkins, R., 2017, Challenges of Safety Assurance for Industry 4.0, Proc. EDCC, 103-106.

[4] Benias, N., Markopoulos, A.P., 2017, A review on the readiness level and cyber-security challenges in Industry 4.0, Proc. SEEDA-CECNSM, 1-5.

[5] Waidner, M., Kasper, M., 2016, Security in industrie 4.0 - challenges and solutions for the fourth industrial revolution, Proc. DATE, 1303-1308.

[6] Gökalp, M.O., Kayabay, K., Akyol, M.A., Eren, P.E., Koçyigit., A., 2016, Big Data for Industry 4.0: A Conceptual Framework, Proc. CSCI, 431-434.

[7] Zhou, K., Liu, T., Zhou, L., 2015, Industry 4.0: Towards future industrial opportunities and challenges, Proc. FSKD, 2147-2152.

[8] Ansari, F., Erol, S., Sihn, W., 2018: Rethinking Human-Machine Learning in Industry 4.0: How Does the Paradigm Shift Treat the Role of Human Learning? Proc. Manufacturing, 23:117-122

[9] Digmayer, C., Jakobs, E.-M., 2018: Employee Empowerment in the Context of domain-specific Risks in Industry 4.0. Proc. ProComm.

[10] Aven, T., 2014, What is safety science?, Safety Science, 67:15-20.

[11] Avizienis, A., Laprie, J.C., Randell, B., 2000, Fundamental Concepts of Dependability, Proc. of the 3rd IEEE Information Survivability Workshop, 7-12.

[12] Hale, A.R., 2000, Culture's confusions, Safety Science, 34:1-14.

[13] Banse, G., Hauser, R., 2008, Technik und Kultur. Das Beispiel Sicherheit und Sicherheitskultur(en), Rösch, O., *Technik und Kultur,* Wildauer Schriftenreihe Interkulturelle Kommunikation, Berlin, 61-83.

[14] Guldenmund, F.W., 2007, The use of questionnaires in safety culture research – evaluation. Safety Science, 45/6:723-743.

[15] Parker, D., Lawrie, M., Hudson, P., 2006, A framework for understanding the development of organisational safety culture, Safety science, 44/6:551-562.

[16] Hudson, P.T.W., 2007, Implementing a safety culture in a major multi-national, Safety Science, 45/6:697-722.

[17] Schuh, G., Stich, V., Jakobs, E.-M., Ziefle, M., 2015, Zukunft gestalten: Soziale Technologien in Organisationen in Zeiten des Wandels. Wissen – Innovation – Demographie. FIR-Edition Forschung Band 15., FIR e.V., Aachen.

[18] Fetterman, D. M., Empowerment Evaluation, Evaluation Practice, Feb. 1994: 1-15.

[19] Brünner, G., 2000, Wirtschaftskommunikation, Niemeyer, Tübingen.

[20] Judge, T. A., Bono, J. E., 2001, Relationship of core self-evaluations traits—self-esteem, generalized self-efficacy, locus of control, and emotional stability—with job satisfaction and job performance: A meta-analysis, Journal of applied Psychology, 86/1:80-92.

[21] Hartmann, A., 1995, Ganzheitliche Sicherheits-IT, Proc. 4. Deutscher Sicherheitskongress, 1-13.

[22] Gardner, D. G., Pierce, J. L., 1998, Self-esteem and self-efficacy within the organizational context an empirical examination, Group & Organization Management, 23/1:48-70.

[23] Rich, B.L., Lepine, J.A., Crawford, E.R., 2010, Job Engagement: Antecedents and Effects on Job Performance, Academy of Management Journal, 53/3:617-635.

[24] Digmayer, C., Jakobs, E.-M., 2014, Corporate Lifelong Learning 2.0: Design of Knowledge Management Systems with Social Media Functions as Learning Tools, Proc. ProComm, 121-131.

[25] Fesidis, B., 2013, Erfolgsfaktoren von

Fachmedien. Eine nutzerorientierte Untersuchung, Apprimus Verlag, Aachen.

8 BIOGRAPHY

Dr. phil. Claas Digmayer is a research assistant at the Human-Computer Interaction Center at RWTH Aachen University (Germany), department of text linguistics and technical communication. He graduated in technical communication and computer science at RWTH Aachen University. Main research interests are: risk perception and communication, text mining, industry 4.0, open innovation.

Prof. Dr. phil. Eva-Maria Jakobs is a full professor of text linguistics and technical communication at RWTH Aachen University, Germany. She is co-director of the Institute for Industrial Communication and Business Media, the Human-Computer Interaction Center at RWTH Aachen University, and the study program technical communication. Main research fields are: technical and digital communication, text technologies, usability research and writing at work.

Stream C

You're My Mate – Acceptance Factors for Human-Robot Collaboration in Industry

V. Lotz[1], S. Himmel[1], M. Ziefle[1]
[1]Human Computer Interaction Center (HCIC) / Chair for Communication Science, RWTH Aachen University, Germany

Abstract
One major step to improve assembling and manufacturing in the age of digitalization could be the implementation of human-robot collaboration (HRC) in industrial workplaces. While there are obvious benefits in using robots, there are several barriers for companies employing robots: On the one hand, safety regulations are the key show stopper for the investment from the employers' point of view, especially when robots physically collaborate with human workers and do not just replace a machine tool, installed in a safety cage. On the other hand, workers' acceptance is crucial for a successful implementation of HRC in the company. In our study, we therefore examined the main influencing factors for HRC acceptance from the employees' point of view. In a two-step procedure, we first gained qualitative results in expert interviews with industrial workers. The results helped to identify the diverse and individual acceptance factors for HRC. In a second approach, we applied a quantitative questionnaire study, adapting the Technology-Acceptance Model (TAM3) on HRC for (1) validating the qualitive results and (2) compare participants with and without an industrial background. Results showed that there are three key anxieties from the employees' point of view: 1st being injured by the robot, 2nd losing the job, 3rd robot anxiety. Addressing the workers' usage intention for the implementation of HRC, we furthermore gained two key factors fostering HRC acceptance: improved output quality from the industrial workers' perspective and, as a generic outcome for all participants, the enhanced enjoyment when collaborating with a robot.

Keywords
Human-Robot Collaboration, Technology Acceptance, User Diversity

1 INTRODUCTION

With autonomous robots as one of nine technologies transforming industrial production to Industry 4.0 [1], making their way into spaces shared with human workers and the disappearance of safety fences – once separating workers from manufacturing robots – the efficient collaboration between humans and robots becomes vital [2]. However, in order to exploit the individual strengths of humans and robots and form hybrid teams that provide added value for all parties involved, new challenges must be solved. On the one hand, safety regulations (DIN EN ISO 10218-1&2) are one of the key show stoppers for the investment from the employers' point of view. This is especially true when working spaces overlap and the robot does not just overtake a task and work autonomously towards task completion. On the other hand, workers' acceptance – and the go from the works committee – is crucial for a successful implementation of HRC in the company. In this study, we focused on the latter and examined the main factors influencing HRC acceptance from the employees' point of view. In a first approach, we conducted a qualitative interview study with industrial workers to gain new insights into acceptance factors. In a second quantitative step, these factors were aggregated with the Technology-Acceptance Model [3], comparing four user groups concerning age, gender, robot-, and manufacturing experience.

2 REFERENCE WORK

2.1 User Diversity and Technology Acceptance

To gain insights into the adoption and usage intention of present and future technologies, the broad research field of Technology Acceptance provides manifold theories and methods, focusing on both the technology and the user himself. On the one hand, the technology itself, as the object of research, plays a crucial role. However, a technology not used by anyone is expensive and unnecessary. Therefore, it is important to understand why or why not users adopt technologies. First models, e.g., Theory of Planned Behavior (TPB) or Technology Acceptance Model (TAM), try to explain behavioral intention and actual usage were developed in the late 80s, examining the use of the new information technologies in the working context [4]. However, considering user diversity, on the one hand, the attempts to explain acceptance with a model-based approach led to several adaptions with more influencing factors. On the other hand, the use of one model for different, established, and especially new technologies, led to controversial discussions about the use of technology acceptance models [5]. As we examined technology acceptance in the working context and could constitute the main acceptance factors in our qualitative pre-study, we considered a model-based approach using TAM3 [3] for this research study.

2.2 Challenges in Human-Robot Collaboration

From the theoretical point of view, several benefits encourage the implementation and use of robots in industrial processes: non-recurring investment (plannable maintenance but neither labor nor non-wage labor costs), 24/7 human mood independent work schedule, integration plus automation in Industry 4.0 processes, high accuracy, etc. Even though human workers are the better and cheaper alternative for processes needing intuition and flexibility, there are three key anxieties from the employees' point of view. When workers have to collaborate with a robot, we need to address and resolve the workers' concerns.

Although **safety in human-robot collaboration** was neglected by research for a long time [6], it is one of the most important aspects for HRC acceptance in the industrial sector. For one, by law, safety standards must be guaranteed by any employer, at least in high-wage countries. For another, workers' acceptance decreases with higher safety risks and the go from the works committee and labor union is inevitable. Latest developments in collaborative robots and sensor technologies show possible solutions. However, the risks are still high and cost-benefit considerations often make HRC solutions inefficient [7]. As safety was considered an imminent factor, we did not analyze its importance for acceptance in this paper but assumed it as a basic prerequisite for our evaluation, well aware that it is a goal hard and expensive to achieve.

As a global challenge, automation in Industry 4.0 will have a deep impact on manpower markets and workplaces. While automation just changes requirements of job-profiles and creates new jobs, especially the low-skilled workers in manufacturing sectors have reasonable fears to lose their jobs. Although collaborative robots are implemented to support workers with the aim to increase production, in the long run, the **fear of job loss** remains [8].

Apart from existential threats influencing physical and financial conditions, the collaboration with a robot can be refused just because of the interaction itself. **Robot anxiety** may sound irrational and irrelevant when considering the implementation of HRC workplaces in manufacturing companies, but for the acceptance and motivation from the co-workers perspective, it can play a crucial role [9].

3 STUDY PART 1 – QUALITATIVE PRE-STUDY

3.1 Methodology

In this study, an exploratory research approach was chosen to determine relevant issues concerning the use of collaborative robots for the manufacturing industry. Since the main focus was to investigate the workers' perspective and their primary concerns, usage conditions, and attitudes, a qualitative approach provides reliable in-depth insights [10]. Hence, guideline-based semi-structured interviews were conducted. Dependent on the participants' answers, the guideline questions were adjusted. In total, five employees were interviewed: three workers in small and medium-sized enterprises (SME's) and two expert interviews at a multinational control and automation company who were head of department and answered as representatives for their department.

The interview guideline was based on newspaper articles as well as literature concerning recent insights into human robot interaction, computer and robot anxiety, and the implementation of technology into the workspace, in consultation with industrial project partners. Apart from general questions about the employees' daily work routine, questions concerning their perception of robots and machines and how those two differ from each other were included. This was done to evaluate if or how the participants' mental models of those technical devices differ from each other. After that, a short definition of a robot was given to ensure that all participants talk about the same concept. The next interview addressed how a robot would alter the work routines and included the employees´ expectations and concerns linked to the use of a robot.

3.2 Results – Interviews

3.2.1 Factors of HRC acceptance

In general, the employees expressed mixed attitudes towards collaborative robots. Concerns were most frequently related to *job loss* as a consequence of increasing digitization and *safety issues* were expressed. However, safety regulations were perceived as a general requirement for all technologies implemented at the workplace and hence not a unique concern in case of collaborative robots. Apart from the above-mentioned major concerns, some participants also noted issues and requirements such as small *relevance* of robots, *usability* factors, *perceived control*, and the perceived *hierarchical relationship* between human and robot. The employees expressed no difference in the perception of traditional tools for manufacturing and collaborative robots which were perceived as an additional tool with a higher degree of autonomy.

Above all, the mixed attitudes and varying requirements suggest a high impact of user factors on the acceptance of collaborative robots. Alongside this result, the interviews revealed a high diversity of employees in the manufacturing industry. Hence, for the next research part, known technology acceptance models were used to further analyze which factors are the strongest predictors of HRI acceptance.

4 STUDY PART 2 – QUANTITATIVE STUDY

4.1 Methodology

4.1.1 The survey

Based on the pre-study insights, an acceptance model that takes already existing knowledge about human-computer interaction in a workplace context into consideration was developed. In the first section, *demographics* such as age and gender were queried. Additionally, the first section contained *work related* questions (duration of occupation, vocational training, current occupation) to compare participants with and without manufacturing background. Finally, the participants' *robot related experience was surveyed*. Another diversity factor was measured by a short version of the *locus of control* (*LOC*) [11]. As introduced before and as previous research [12] revealed, the factors used in TAM3 are suitable for the context of human-robot interaction. TAM3 items regarding *Computer Playfulness* (CPLAY), *Computer Anxiety* (CANX) , *Perceived Usefulness* (PU), *Job Relevance* (REL), *Subjective Norm* (SN), *Output Quality* (OUT), *Computer Self- Efficacy* (CSE), *Perception of External Control* (PEC), *Perceived Enjoyment* (ENJ), *Voluntariness* (VOL), *Image* (IMG), and *Perceived Ease of Use* (PEOU) [3] were measured, as was LOC, using a 6-point Likert-scale (from 1=high approval values to 6=high rejection values). Other than *Computer Playfulness* and *Computer Anxiety*, all were asked after giving an HRC scenario.

The reliability of the used scales was mostly above α= .7. Only CPLAY and VOL had smaller Cronbach´s Alpha values, VOL was not considered for correlation measurements.

4.1.2 Sample description

In total, n=159 people took part in the online survey, of which n=131 where used for the statistical analysis. The survey was distributed as part of a bachelor thesis, hence most of the participants come from a university-related environment.

The respondents' age ranged from 19 to 60 (*M*=30.52, *SD*=12.00), with a slightly higher ratio of male participants (n_{male}=79, 58.0%, n_{female}=55, 42.0%). The level of education was comparatively high with 35.1% holding a university degree, 42.0% with a higher education entrance qualification, and 18.4% with a secondary education degree. Half of the participants (52.7%) were employed and the rest were either university students (43.5%) or still in school (2.3%). 44.3% were working or had worked in production and 55.7% had no previous work experience in production. Slightly more than half of the participants (56.5%) had no experience with manufacturing robots, 32.1% rarely encountered robotic systems, 9.2% worked with robotic systems on a regular basis, and 2.3% reported to work with a robotic system on a daily basis. Concerning attitudinal variables, the participants reported above average *Locus of Control* (*M*=2.06; *SD*=.87) and a similarly high mean for *Behavioral Intention* (*M* = 2.16; *SD*=.93). Additionally, the participants' *Computer Anxiety* was on average negative (*M*=5.37; *SD*=.68) (Figure 1) which means they had no fear of using computers.

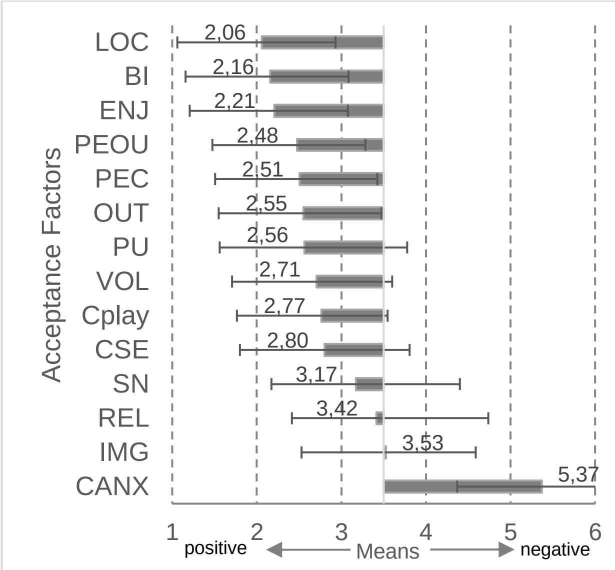


Figure 1 – Overview mean values of queried factors.

Since user diversity plays an important role in the acceptance of new technologies, the participants were divided into groups (*age, gender, robot experience,* and *manufacturing experience)* for further analyses (Table 1). Two age groups were formed by performing a median split. The younger group´s age range matches the average age of German students [13].

Factor	Group	Valid responses [%]	n
Age	19-26	54.2	71
	26-60	45.8	60
Gender	Male	58.0	79
	Female	42.0	55
Robot Experience	Yes	43.5	57
	No	56.5	74
Manufacturing Experience	Yes	44.3	58
	No	55.7	73

Table 1 – User diversity groups**.**

4.2 Results – Survey

In a first step, the whole sample was analyzed by using correlation (Spearman´s rho) with a level of significance of α= .05. The resulting collaborative-robot acceptance model is based on TAM3 [3] and extended by the factors *Locus of Control*, *Age*, *Robot Experience*, *Manufacturing Experience,* and *Gender*.

Subsequently, a linear regression analysis was performed to determine which factors predominantly impact *Behavioral Intention*. Finally, the queried factors were analyzed with respect to user diversity by using a linear regression analysis.

4.2.1 Correlations TAM3

The final model is presented in Figure 2. Almost all of the queried factors were correlated with either *PEOU* or *PU* or both. Only *Image*, *Computer Self Efficacy* and *Gender* had no detectable impact on *Perceived Usefulness* and *Perceived Ease of Use* in the present sample. Overall, the three factors with the strongest influence on *Perceived Usefulness were Subjective Norm* (*M*=3.17; *SD*=1.23), *Output Quality* (*M*=2.55; *SD*= .92), and *Job Relevance* (*M*=3.42; *SD*=1.32). Regarding *Perceived Enjoyment* (*M*=2.21; *SD*=.87), *Perception of External Control* (*M*=2.51; *SD*=.92) and *Output Quality* (*M*=2.55; *SD*=.92) were the most influential factors. *Computer Anxiety* (*M*=5.37; *SD*=.68) and *Age* (*M*=30.52; *SD*=12.00) were negatively correlated with both *PEOU* and *PU*.

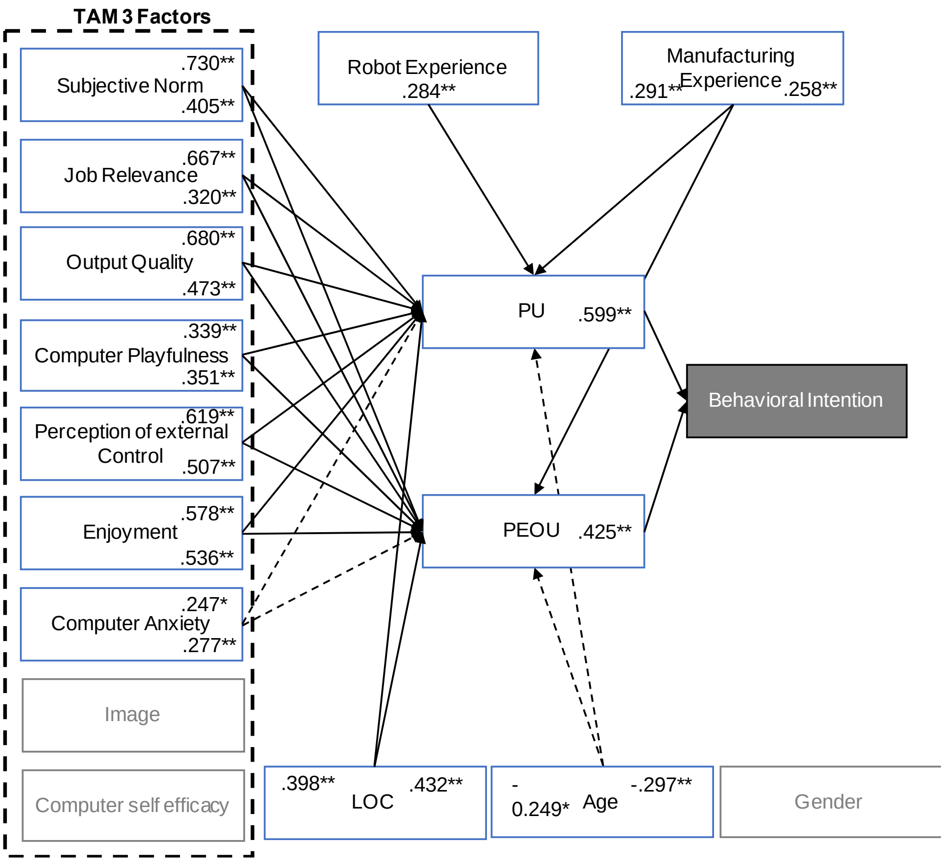


Figure 2 – Collaborative-Robot Acceptance Model – derived from TAM 3 [3] – with correlation coefficients values (left/right, up/down) referring to connecting line in direct proximity (*p<.05, **p<.001).

4.2.2 Overall acceptance factors on HRC

In order to identify the factors with the strongest influence on the participants' *Behavioral Intention*, a step-wise linear regression analysis was performed, with *BI* as the dependent variable. Since we assume that in the present case not only *PU* and *PEOU* alter BI, but *BI* is also directly influenced by other factors of the model, we set all remaining factors of the acceptance model (Figure 2) as independent variables.

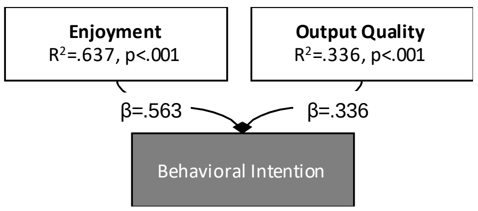


Figure 3 – Stepwise regression: Main influencing factors on *Behavioral Intention*.

As a result, two significant models were revealed. The first model could account for 56.3% (adj. R^2=.563) of the variance in acceptance and the second for 63.7% (adj. R^2=.637). In the first model, acceptance of the robotic system was solely based on *Enjoyment* (β=.754). The second model (Figure 3) also includes *Output Quality* (β=.336).

4.2.3 Influence of user diversity on acceptance

Age – The calculation resulted in three significant models for the younger group of the sample and in two significant models for the older group.

Group	Mod.	1st	2nd	3rd	Adj R^2
Age: 19-26	1	Enj β=.734			.530
	2	Enj β=.606	Out β=.292		.592
	3	Enj β=.521	Out β=.281	Vol β=.216	.624
Age: 26-60	1	Enj β=.765			.575
	2	Enj β=.457	PU β=.428		.656

Table 2 – Stepwise regression: age-groups.

The first two models, regarding the younger participants, were based on the same factors as the models for the sample as a whole (Model 1: adj. R^2=.530, Model 2: adj. R^2=.592). However, the final model explained +3.8% variance and additionally included the factor *Voluntariness* (β=.216). All values are listed in Table 2.

The first model for the older participants was likewise based on *Enjoyment* and can account for 57.5% (adj. R^2=.575) of the variance. The second model additionally included *Perceived Usefulness*.

Gender – For both genders, two significant models were revealed. While the first model only differs marginally between men (Enj β=.743, adj R^2=.542) and women (Enj β=.779, adj R^2=.596), the second factor in the final model is different. For men, the second model includes *Output Quality* and for women *Perceived Usefulness*. The values are listed in Table 3.

Group	Mod.	1st	2nd	Adj R^2
Men	1	Enj β=.743		.542
	2	Enj β=.545	Out β=.349	.618
Women	1	Enj β=.779		.596
	2	Enj β=.547	PU β=.402	.699

Table 3 – Stepwise regression: gender.

Robot-Experience – The linear regression analysis in regard to Robot-Experience revealed the same results as the regression for the whole sample. In Table 4, all relevant values are listed.

Group	Mod.	1st	2nd	Adj R^2
No Exp.	1	Enj β=.763		.574
	2	Enj β=.554	Out β=.335	.637
Exp.	1	Enj β=.696		.469
	2	Enj β=.579	Out β=.351	.569

Table 4 – Stepwise regression: robot-experience.

Manufacturing Experience – Regarding Manufacturing Experience, the calculation resulted for both groups in two significant models. However, in contrast to the previous results, the first factor differs between the two groups.

Group	Mod.	1st	2nd	Adj R^2
Manuf. Exp.	1	Out β=.583		.320
	2	Out β=.422	Enj β=.374	.421
No Exp.	1	Enj β=.815		.657
	2	Enj β=.637	PU β=.307	.716

Table 5 – Stepwise regression: manufacturing exp.

Looking at the final models, the most influential factor for participants with manufacturing experience is *Output Quality* (OUT β=.422, adj R^2=.421), followed by *Enjoyment* (β=.374). Nonetheless, model two can only account for 42% of the variance. For those of the sample with no manufacturing experience, *Enjoyment* is the strongest predictor of *BI* (ENJ β=.637, adj R^2=.716). This model can explain 71.6% of the variance and additionally includes *Perceived Usefulness (*β=.307). All results are listed in Table 5.

5 DISCUSSION

5.1 Acceptance Considerations Beyond Workplace Safety and the Fear of Job Loss

As previously reported, nearly all TAM3 constructs affect *Perceived Usefulness* and *Perceived Ease of Use* and thus also influence the *Behavioural Intention*. With *REL*, *SN*, and *OUT* as the most influential constructs for the *Perceived Usefulness* and *ENJ*, *PEC* and *OUT* for the *Perceived Ease of Use*. While these results are congruent to the findings of Bröhl et al. [12] concerning *PU*, they diverge from the earlier study in regard to *PEOU*. Nonetheless, a further analysis of the data revealed that two of these constructs are sufficient to explain 63.7% of the observed variance in *Behavioral Intention*, namely *Enjoyment* and *Output Quality*. This implies that, in case of collaborative robots, *PU* and *PEOU* are important but not the fundamental concepts modeling acceptance.

Concerning user diversity, *gender* and *robot-experience* had little to no impact on *BI*. As manufacturing experience and age correlated with acceptance factors (*PU*, *PEOU*), one could interpret that younger people without industrial background were compared to older industry workers. Regarding the regression analysis results, this hypothesis can be falsified.

The repeated occurrence of *Enjoyment* as an important factor for *BI* across all user groups leads to the assumption that the enjoyability of such a system fundamentally shapes its acceptance. Consequently, apart from the expected requirements in a workplace context – such as a reliable output quality and the perceived usefulness of the system – it can be crucial to develop a robotic system which users enjoy working with. The importance of *Enjoyment* is particularly surprising when comparing this result to the pre-study insights. In those, the employees seemed to view the robot as any other tool used for their work routines and like any other tool, the robot mainly needs to serve its function. It remains debatable if, in case of established manufacturing tools, enjoyment likewise alters their acceptance.

5.2 Limitations, Recommendations, and Future Research

The qualitative study confirmed the main known concerns about HRC from the workers' point of view. However, the number of participants was quite low and more interviews or focus groups might have revealed other influencing acceptance factors.

The quantitative study validated the factors revealed in the qualitative pre-study. Especially the model-based approach to explain technology acceptance was applicable as nearly all influencing factors had an impact on the traditional acceptance dimensions (perceived ease of use, usefulness, and intention to use). Regarding user diversity, the evaluation of HRC for industrial workers vs. non-experts resulted in significant differences (see 4.2.3). In contrast, the influence of robot experience did not reveal any differences which might be due to the measurement of robot experience, on the one hand, and the overall quite low expertise for HRC. After all, the use of TAM3 should include not only a scenario and the measurement of intentions to use a technology but finish the validation with a usability study. Hands-on experience is crucial to understand the actual usage of technologies and their acceptance, which has to be included in future research.

As recommendations for the implementation and acceptance of industrial HRC, our study revealed especially one generic component to be considered: enjoyment. We did not measure and therefore cannot recommend gamification for the increase of acceptance, although this could be possible. However, as our results indicate, the fun of working with a robot colleague plays a vital role for acceptance. In conclusion, our research suggests, as soon as the most urgent issues – safety regulations and the possibility of job loss – are solved, comparatively small variations in the design of a robotic systems can have a great impact on the successful implementation of HRC in industry.

6 ACKNOWLEDGEMENTS

We acknowledge the profound input of our industry partners and all participants for their time. We also want to thank Fabian Comanns and Sean Lidynia for research support. This project was supported by the German Federal Ministry of Education and Research (BMBF, Funding number: 01FG14001A).

7 REFERENCES

[1] Rüssmann, M., et al., 2015, Industry 4.0: The future of productivity and growth in manufacturing industries, Boston Consult. Group, 9.

[2] Rozo, L., Amor, H. B., Calinon, S., Dragan, A., Lee, D., 2018, Special issue on learning for human-robot collaboration, Auton. Robots, 42/5:953-956.

[3] Venkatesh, V., Bala, H., 2008, Technology acceptance model 3 and a research agenda on interventions, Decis. Sci., 39/2:273-315.

[4] Mathieson, K., 1991, Predicting user intentions: comparing the technology acceptance model with the theory of planned behavior, Inf. Syst. Res., 2/3:173-191.

[5] Röcker, C., 2010, Why traditional technology acceptance models won't work for future information technologies, World Acad. Sci. Eng. Technol., 65:237-243.

[6] Bauer, A., Wollherr, D., Buss, M., 2008, Human-robot collaboration: a survey, Int. J. Humanoid Robot., 5/1:47-66.

[7] Zanchettin, A. M., Ceriani, N. M., Rocco, P., Ding, H., Matthias, B., 2016, Safety in human-robot collaborative manufacturing environments: Metrics and control, IEEE Trans. Autom. Sci. Eng., 13/2:882-893.

[8] Arntz, M., Gregory, T., Zierahn, U., 2016, The risk of automation for jobs in OECD countries, OECD Social, Employment and Migration Working Papers, 189.

[9] Nomura, T., Kanda, T., Suzuki, T., Kato, K., 2004, Psychology in Human-Robot Communication: An Attempt through Investigation of Negative Attitudes and Anxiety toward Robots, ROMAN 2004, 13th International Workshop on IEEE:35-40.

[10] Courage, C., Baxter, K., 2005, Understanding your users: A practical guide to user requirements methods, tools, and techniques. San Francisco, CA, Morgan Kaufmann Publishers.

[11] Ajzen, I., 2002, Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior 1, J. Appl. Soc. Psychol., 32/4:665-683.

[12] Bröhl, C., Nelles, J., Brandl, C., Mertens, A., Schlick, C. M., 2016, TAM reloaded: a technology acceptance model for human-robot cooperation in production systems, International Conference on Human-Computer Interaction: 97-103.

[13] Statista, 2016, Mean age of German University Graduates, study 189237.

8 BIOGRAPHY

Vivian Lotz holds a BSc degree in Technical Communication and Materials Science. She is a Master student and research assistant at the Human-Computer Interaction Center (HCIC) in Aachen, Germany.

Simon Himmel has a master's degree in Technical Communication with Mechanical Engineering from the RWTH Aachen University. He is Ph.D. student at the Human-Computer Interaction Center (HCIC) in Aachen, Germany. His research focuses on technology acceptance and user-centered design.

Martina Ziefle earned her PhD degree in psychology from the University of Fribourg, Switzerland. She is full professor, at the chair for Communication Science at RWTH Aachen and a director of the Human-Computer Interaction Center at RWTH Aachen University, Germany

Can the Concept of Reengineering be Applied to the Human Voice?

P. Hoyer[1], J. Schwiebacher[1], S. Graf[2]
[1]Fraunhofer Head Department, Munich, Germany.
[2]Department of Otorhinolaryngology, TU Munich, Munich Germany

Abstract
We discuss a novel approach for adjusting the acoustic surroundings of the vocal folds using the concept of reengineering based on reverse engineering. In this approach, without knowing vocal tract shape, adaptation of selected acoustic parameters allows vocal capacity to be increased. In the proposed workflow, formant frequencies and bandwidth data from literature are selected. Within a vocal training procedure, an external loudspeaker placed in front of the open mouth enables transfer of the selected parameters. The participant reengineers the vocal tact by intentionally increasing the volume of the external soundwaves without using the vocal folds. This increase in intensity during adjustment at resonance frequency constitutes a decrease in acoustic damping and may help to obtain an optimised acoustic surrounding for a healthy voice production.

Keywords: Reengineering, Reverse engineering, Acoustic phonetics, Vocal training

1 INTRODUCTION

The human voice is fundamental to social interaction and self-expression. The capability of human beings to articulate a combination of vowels and consonants is the basis of all languages. Successful vocal development relies upon learning new — or re-learning previous — vocal motor abilities, yielding optimised output qualities [1, 2].

The variable shape of the vocal tract influences the output parameters of human voice. In particular, vocal articulation is related to changes in the positions of the lips, tongue, jaw, and larynx [3]. For this reason, functional dysphonia is common when the vocal folds are used with high mechanical pressure on the one hand and with low support from the acoustic properties of the vocal tract on the other [4].

The acoustic properties of the vocal tract resemble those of an open/closed tube. Well-defined resonant modes (formants) support high output values and low damage, even during prolonged use. The first two formants define which vowel is produced, while the higher formants allow a pervasive voice.

Reverse engineering [5] is commonly used to infer the functional features of existing systems and thus facilitate the development of new products with the same features. The concept is usually restricted to technical devices. In the present study, we explore whether a reengineering approach with the help of reverse engineering can be applied to stabilise vocal fold vibration and support durable voice production.

As previous studies [6,7] indicated that externally applied soundwaves induced changes in acoustic output of the participants, it may be possible to transfer selected acoustic properties without any specific knowledge of the individual vocal tract shape. We hypothesised that active adjustment of the vocal tract shape using a reengineering approach enhances the resonance of the acoustic system.

The aim of the study was thus to estimate whether the concept of reengineering might be applied in the context of vocal training.

2 THEORY

2.1 Excitation of vocal tract resonances

Fig. 2 shows some of the basic articulatory elements of the human vocal tract used for resonant amplification and vowel formation at the first and second formants. This complex system can be simplified as an open/closed tube [8, 9].

The particle velocity of the first two formants is in an idealized characterization near to zero at the bottom of the pharynx (high impedance) and high at the mouth opening (low impedance). A vibrating loudspeaker membrane placed directly in front of the mouth couples with the particle velocity anti-node and matches the natural resonance conditions.

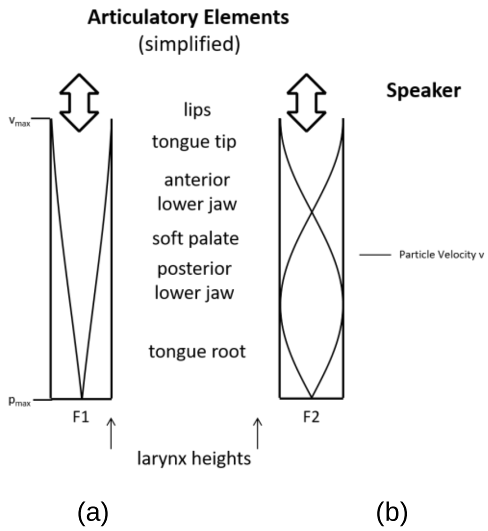


Figure 1 - Simplified model of vocal tract excitation with an external sound source. The resonance conditions of the first and second formants (F1, F2) of the vocal tract are indicated, as are the basic articulatory elements that tune the vocal tract resonances and adjust the acoustic boundary conditions for resonance (adapted from [8,10,11,12]).

The resulting radiation impedance of the vocal tract upon external excitation can be characterised as:

$$Z_{VT} = R + iX \qquad \textbf{Eq. 1}$$

With Z_{VT} being the radiation impedance, R = real part, related to converstion of acoustic energy to heat, the reactance X and $i = \sqrt{-1}$.

X is related to the difference between the sinusoidal frequency of the speaker and the resonance frequency of the vocal tract. When the resonance frequency of the vocal tract equals the frequency of the source, the complex part of the impedance ('iX') is no longer relevant. Equation 1 can then be simplified to

$$Z_{VT} = R \qquad \textbf{Eq. 2}$$

The participant in the present experiment amplified the sinusoidal soundwave by adjusting the vocal tract resonance frequency. Applying Eq. 2, the amplitude at the resonance condition ('A_{max}') is directly related to a low loss factor ('ɑ') within the vocal tract, which is correlated with a small bandwidth ('Bw') and a high-quality factor ('Q').

2.2 Reengineering of the human voice

Reengineering and reverse engineering are widely used for fast product development [13]. The approach analyses existing designs to integrate key elements of functionality into new products without full knowledge of the underlying primary structures. The concept is only common within the development of technical systems. However, we propose that the underlying idea can also be applied to human activities.

Instead of a knowledge of the geometrical shape of the vocal tract, a usage of easily available acoustic properties preferably derived from professional voices is proposed. Fig. 2 shows the resulting process flow applying reengineering to vocal enhancement.

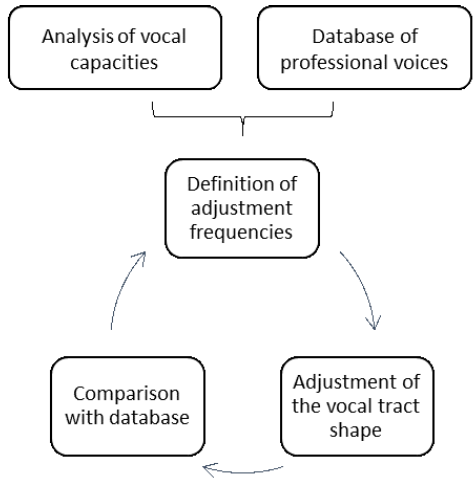


Figure 2 - Overview of the proposed reengineering cycle for adjustment of the human voice.

The formant frequencies and bandwidth of the voice define the starting position of the individual vocal capacities and are compared to standard values of professional voices from literature. For example, in the case of so-called backward voices, the formant frequencies may be low and the bandwidth high. In such cases, the participant may raise the formant frequency and narrow the bandwidth via specific training with external soundwaves.

In the literature, broad data on trained voices are available [3,11]. These data include characteristics such as intelligibility, volume, and the economical use of air.

2.3 Vocal tract filter/resonator function

The vocal tract can be regarded as an acoustic filter of the primary sound source at the glottis. Its resonating aspects enhance the acoustic energy emitted at the resonance frequencies.

Fulop et al. [14], as well as Ladefoged and Fant [15], proposed a semi empirical formula to describe the vocal output in relation to formant frequencies and bandwidth.

Following this approach, the primary sound source S(f) is assumed to decay as described by Eq. 3.

$$S(f) = g * 20 * log_{10}(2 * \frac{\frac{f}{100}}{1+(f/100)^2}) \qquad \textbf{Eq. 3}$$

Where *'g'* represents differences in phonation type and *'f'* the frequency. In standard phonation, *'g'* will be equal to 1.0.

The contribution of each formant to the spectrum envelope is shown in Eq. 4, where *'f'* is the frequency, 'F_n' is the resonance frequency of the respective formant *'n'*, and 'BW_n' is the corresponding bandwidth of the formant *'n'* [13]. The transfer function 'H_n' is calculated in Eq. 4:

$$H_n(f) = -10\log_{10}((\frac{1-f^2}{F_n^2})^2 + (\frac{BW_n^2}{F_n^2}) * (\frac{f^2}{F_n^2}))$$ **Eq. 4**

Finally, a correction factor ('K_{r4}') for the contribution of higher formants with *'f'* as the current frequency can be calculated as shown in Eq. 5:

$$K_{r4}(f) = 0.54 * (\frac{f}{500})^2 + 0.00143 * (\frac{f}{500})^4$$ **Eq. 5**

Including these elements, the output spectre 'L(f)' can be expressed as follows:

$$L(f) = \sum H_n(f) + K_{r4}(f) + S(f)$$ **Eq. 6**

Fig. 3 depicts the different parameters of the procedure.

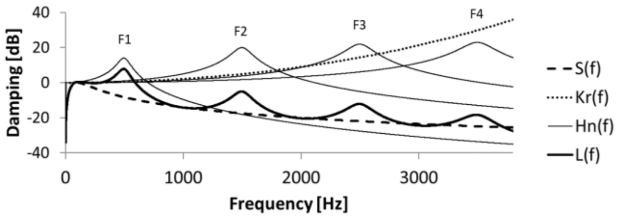


Figure 3 - Calculation procedure to simulate vocal tract transfer function based on formant frequencies and bandwidths, according to Funlop et.al. [14].

3 EXPERIMENTAL SECTION

3.1 Acoustic measurements

The acoustic measurements were performed with a G.R.A.S. 46AF (G.R.A.S Ltd., UK) microphone connected via a preamplifier (Nexus 2690; Brüel & Kjaer, Denmark) to a Steinberg UR242 (Yamaha Inc., Japan) as an external sound card. The distance between the microphone and vocal source was 20 cm. All recordings were performed at 41.100 Hz/16 bit using the software Audacity (D. Mazzoni, Mountain View, CA, USA).

3.2 Voice analysis

Fast Fourier transform (FFT) Spectra were obtained using WaveSurfer version 1.8.8.p4 software (KTH Department of Speech, Music, and Hearing, Sweden). The FFT analysis during phonation was performed at wavelength steps of 5 Hz.

3.3 Resonant amplification

The sound waves of a frequency generator app (ee-toolkit, Thomas Gruber), together with the inbuilt speaker of an iPhone 5s (Apple Inc., Cupertino, CA, USA) were used to excite the air column within the vocal tract. In order to apply two frequencies simultaneously, the dual frequency mode was used, with the phase difference of the two frequencies set at 180°.

The procedure to enhance the amplitude of the external signal has been described in detail in a previous work [6]. After voice analysis and recording of vowels, the soundwaves in front of the participant's open mouth were amplified via a variation of the vocal tract shape. During this stage, the participant produced no sound from the vocal folds. The application time was less than 30 seconds, with similar relaxation times in between and the procedure repeated several times. The overall adjustment period was less than 5 min. Next, the participant was asked to sustain vowels for at least 5 seconds. The fundamental frequency of these vowels was within the range of the spoken voice.

3.4 Participant

All experiments were conducted with one male participant who had a healthy voice: a non-professional with > 10 years' choral experience, an average fundamental speech frequency of 99 Hz (44 dB), and a sound pressure maximum of 104 dB, measured using a XION Voice Analysis System (XION GmbH, Berlin, Germany).

The experiments were approved by the Ethics Committee of the Technical University of Munich in January 2017.

4 RESULTS

4.1 Calculation of the filter/resonator function

Fig. 4 shows the primary sound source and the simulated filter function, calculated using the data of the first four formants. While the primary source showed a steady decay in intensity, the spectrum of the radiated voice depended on the bandwidth of the formants.

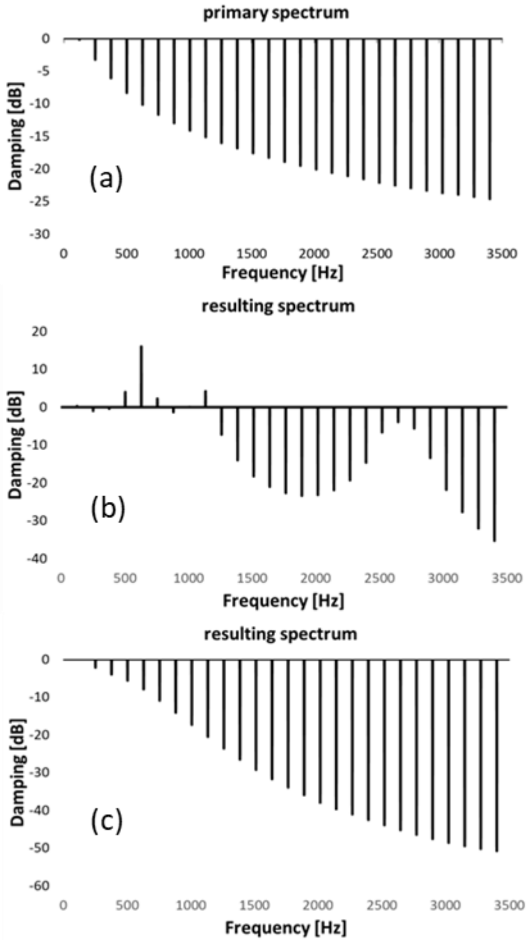


Figure 4 - (a) Basic spectrum at the entrance of the vocal tract. (b, c) Filtered spectra. (b) Formant 1 to 4 frequencies/bandwidths (Hz): 620/50, 1120/100, 2580/150, and 2800/180. (c) Filtered spectrum with formant frequencies (f_n/Bw_n = 1.0).

The fundamental frequency was 126 Hz.

In addition to the bandwidth and formant position, the output level was also sensitive to the position of the formant frequencies.

4.2 Sound measurements

Fig. 5 shows the results of reengineering the vocal output parameters. To observe the spectral changes, the sound level was normalised to the intensity of the fundamental frequency. The upper plot 5 a shows the natural spectral response of the participant's vowel 'a'. The first formant was at 511 Hz, while the second one at 1100 Hz. Subsequently, two different interventions were carried out. One had similar frequencies compared to the natural formant frequencies (500 Hz and 1000 Hz, respectively) and one with raised formant frequencies (650 and 1300 Hz). Figs. 5b and 5c show the peaks of the higher harmonics after vocal tract shape adjustment. When the given frequencies were comparable to the formant frequencies, the harmonics were clearly enhanced.

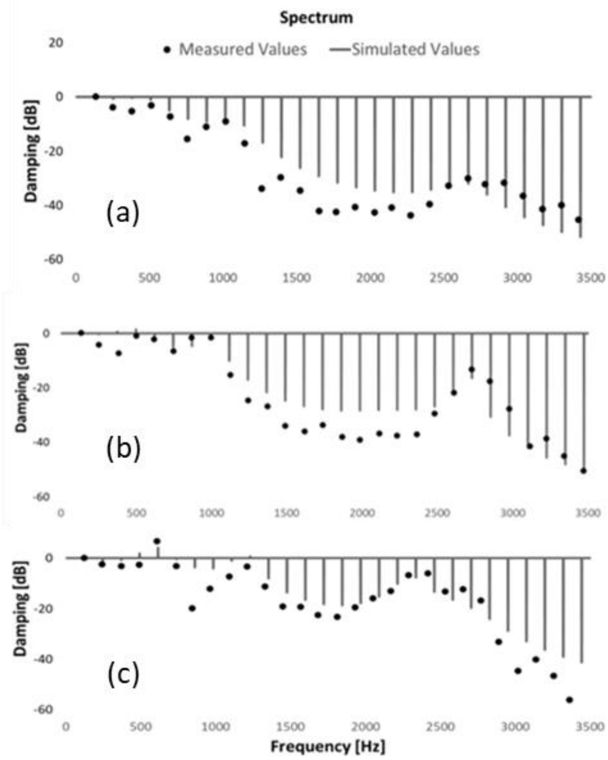


Figure 5 - Simulated (lines) and measured spectra (dots) of the vowel 'a' before and after reengineering of the vocal tract shape. (a) Without adjustment of the vocal tract shape, (b) with adjustment to frequencies of 500 Hz and 1000 Hz, and (c) with adjustment to frequencies of 650 Hz and 1300 Hz.

The following parameters gained the demonstrated spectra in Fig. 5:

(a) F1=500 Hz, BW1=300 Hz; F2=1100 Hz, BW2=250 Hz; F3=2200 Hz, BW3=2200 Hz; F4=2660 Hz, BW4=300 Hz

(b) F1=500 Hz, BW1=250 Hz; F2=1000 Hz, BW2=150 Hz; F3=2200 Hz, BW3=800 Hz; F4=2700 Hz, BW4=80 Hz

(c) F1=600 Hz BW1=150 Hz; F2=1200 Hz, BW2=130 Hz; F3=2300 Hz, BW3=160 Hz; F4=2700 Hz, BW4=450 Hz

The reengineering approach to reducing the bandwidth without changing the formant frequencies is shown in Fig. 5b. Although only the first two formant frequencies were addressed, a reduction of the bandwidth was observed at all four formants. The amount of energy within the frequency range where the human ear is most sensitive—from about 500 to 2500 Hz—was enhanced after adjustment of the vowel ‘a’.

In Fig. 5c, the frequencies of the first two formants were raised. They are still within the range of the vowel ‘a’ in German pronunciation. The reduced bandwidth was conserved even at higher adjustment frequencies. However, unlike the previous case, the higher harmonics were more intense.

The calculated and simulated output of the produced sounds were similar, but not equal—the deviation being about +/- 5 dB. Larger deviations occurred in the spectrum shown in Fig. 5a at 1300 Hz and in Fig. 5c at around 880 Hz and 3000 Hz, with a singular decrease in intensity.

5 DISCUSSION

5.1 Inverse engineering and reengineering of the acoustic parameters of the vocal tract

The amplified sinusoidal frequencies used in the reengineering approach are similar to the ones of the first and second formant. While the acoustic transfer function can be calculated on the basis of given geometric data, given acoustic data may result from different geometric arrangements. Therefore, an analysis with the help of simple reverse engineering tool that stresses individual acoustic properties (Fig. 4) may be flexibly adapted by users with different vocal tract geometries.

The formant positions and corresponding bandwidths are essential to the characteristics of the individual voice. In the case of high bandwidths, as shown in Fig. 4, frequencies below the first formant show decay characteristics similar to those of the source. However, narrow bandwidths reduce damping and enable higher harmonic radiation and may lead to a more persuasive voice.

Fig. 5 shows stages of the proposed reengineering cycle (Fig. 2). Starting from a given vowel, the inverse engineering procedure starts with a comparison of formant frequency and formant bandwidth with healthy or professional voices. In the present case, the formant frequencies of 511 and 1100 Hz were rather low for the vowel ‘a’ which are in average at 639 and 1225 Hz respectively. Furthermore, the bandwidth of the first and second formants were broader than those found in the literature [12**Error! Bookmark not defined.**]. In a first step, a narrowing of the formant bandwidths should enable more efficient voice production and a more clearly structured voice. Fig. 5b shows that vocal tract adjustment can lead to reduced bandwidth while maintaining the formant frequencies to about 10% of their original value.

Further optimisation may involve an induced change in the formant frequencies. For the vowel ‘a’ in men, the usual frequency of the first formant lies in the 529–674 Hz range, while that of the second formant is in the 1224–1386 Hz range [16]. Since higher formant frequencies support the radiation of higher harmonics, raising the formant frequency can lead to a more brilliant voice. However, in the present study, care was taken to remain within the formant frequency ranges of 650 Hz and 1300 Hz of the vowel ‘a’ (German pronunciation). After repeated intentional amplification of sinusoidal soundwaves, the spectral range increases (Fig. 5c). The strong intensity of the 4^{th} harmonic at 615 Hz was close to the external sound wave used. This suggests that impedance of the vocal tract is reduced at the given adjustment frequency, as proposed in Section 2.1. Furthermore, the second format position was at 1200 Hz, which is 8% lower than the adjustment frequency.

In Fig. 5c the recorded signals of the 7^{th} and 8^{th} harmonics (880 and 1000 Hz) are highly damped, while any formant value chosen within the simulation would create less congruence of the overall spectrum. Since 1300 Hz is the first harmonic of 650 Hz, an anti-resonance is likely to occur between both. In the semi-empirical approach used in the present study, only resonances were included.

5.2 Resonant amplification and vocal tract shape

The proposed procedure placed no restrictions on the geometry of the vocal tract. Instead, the reengineering approach addresses changes in geometry via a transfer of acoustic data.

It should be noted that the deviations between adjustment with an open glottis compared to phonation are much smaller than those induced by a whispering voice [17].The resonant frequencies are close to the calculated formant frequencies, but not equal. The participant adjusted the vocal tract with free breathing through the open mouth, without using the vocal folds as voice source. If the vocal folds started to oscillate, a change in the acoustic parameters would be expected. The findings might be explained with the simulation results of Takemoto et al. [8] who found the particle velocities of the first and second formant to be almost zero at the passage between the epilaryngeal tube and pharynx, which is positioned about 2 cm above the vocal folds.

The reengineering approach used in the present study did not directly address the third and fourth formants. However, we did find a reduction of these formants bandwidths as well. Although the term **R** (Eq. 2) is reduced within the experimental setup at the specific wavelengths used, the value is regarded to be less sensitive to frequency and may thus help to decrease at all bandwidths observed. Further work is necessary to elucidate the process further.

6 CONCLUSIONS

By applying the concept of reengineering, the vocal tract shape can be adjusted on the basis of induced output parameters.

The concept of reengineering or reverse engineering is not commonly applied to human activities. However, the underlying ideas may be used for vocal training or to maintain a healthy voice. Specifically, with only limited information about the vocal tract shape, vocal output can be optimised.

The current experiment demonstrates that externally applied frequencies can be used to alter formant position and bandwidth. However, the present experimental setup could not assess the universal validity or long-term effects on voice quality.

If confirmed in further studies, this approach could be used to transcribe selected vocal properties of professional voices to untrained voices. More work will be necessary to further evaluate the practical applications of the present study within the field of vocal training.

7 REFERENCES

[1] Fant G., 1970, Acoustic Theory of Speech Production, Mouton, Hague, The Netherlands

[2] Lindblom B., Sundberg J., 2007, The Human Voice in Speech and Singing, in Springer Handbook of Acoustics, Ed. T.D. Rossing; Springer-Verlag New York, 669–712

[3] Lidblom B., Sundberg J., 1971 Acoustical Consequences of Lip, Tongue, Jaw, and Larynx Movement, The Journal of the Acoustical Society of America 50, 1166–1179

[4] Maryn Y., De Bodt M., Van Cauwenberge, P., 2006, Effects of Biofeedback in Phonatory Disorders and Phonatory Performance: A Systematic Literature Review, Applied Psychophysiology and Biofeedback 31 (2006) 65-83

[5] Eilam E., Secrets of Reverse Engineering, Wiley Publishing, New York, 2005

[6] Hoyer P. and Graf S., 2018 , Adjustment of the Vocal Tract Shape via Biofeedback – a Case Study, Journal of Voice in press, DOI: 10.1016/j.jvoice.2018.01.018

[7] Graf S., Schwiebacher J., Richter L., Buchberger M., Adachi S., Mastnak W., Hoyer P., Journal of Voice in press, DOI: 10.1016/j.jvoice.2018.10.007

[8] Takemoto H., S. Adachi S., Kitamura T., Mokhtari P., Honda K., 2006 Acoustic roles of the laryngeal cavity in vocal tract resonance; The Journal of the Acoustical Society of America 120 2228–2238

[9] Kitamura T., Takemoto H., Adachi S., Mokhtari P., Honda K., 2006 Cyclicity of laryngeal cavity resonance due to vocal fold vibration, The Journal of the Acoustical Society of America 120, 2239–2249

[10] Wolfe J., Garnier M., Smith J., 2009, Vocal tract resonances in speech, singing, and playing musical instruments, Human Frontier Science Program Journal 3, 6–23

[11] Sundberg J., 2017, Vocal Tract Resonace, in: The Professional Voice: The Science and Art of Clinical Care, 3-Volume Set. Plural Publishing, 4th ed., Salatoff R.T. (Ed.), 309 – 328

[12] Patzold M., Simpsons A.P., 1997 Acoustic analysis of German vowels, Kiel Corpus of Read Speech, Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung Universität Kiel, 32, 225

[13] Sokovic M., Kopac J., 2006, RE as necessary phase by rapid product development; Journal of Materials Processing Technology, 175. Jg., Nr. 1-3, S.398-403

[14] Fulop S.A., Kari E., Ladefoged P., 1998, An acoustic study of the tongue root contrast in Degema vowels. Phonetica, 55 (1-2), 80-98

[15] Ladefoged P., Fant G., 1997 Fieldwork techniques for relating formant frequency, amplitude and bandwidth Fifth European Conference on Speech Communication and Technology

[16] Patzold M., Simpsons A.P., Acoustic analysis of German vowels, Kiel Corpus of Read Speech, Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung Universität Kiel, 32 (1997) 225

[17] Swerdlin Y., Smith J., Wolfe J., The effect of whisper and creak vocal mechanisms on vocal tract resonances, The Journal of the Acoustical Society of America, Volume 127, (2010) 2590–2598

8 BIOGRAPHY

Patrick Hoyer obtained his PhD degree in Physical Chemistry from the Technical University of Berlin. He is currently department head in research planning at Fraunhofer headquarters.

Simone Graf received her PhD in Medicine at Erlangen University. She is head of the phonetics department of the TU Munich and chair of the German Society of Dysphagia.

Digitization of Ergonomic Planning and Design Processes - Results from the Validation Process of Motion Capture as Ergonomic Assessment Tool

R. Hensel[1], M. Steinmetz[1], L. Hausmanninger[1], M. Keil[1], M. Benter[2], P. Kuhlang[2]

[1]Dept. Industrial Engineering Methods, AUDI AG, Ingolstadt/ Germany

[2] MTM-Institute, German MTM-Association, Hamburg/ Germany

Abstract

In Germany, employers are legally obliged by labor legislation to assess the conditions at workplaces that could cause harm to identify hazards and control any health risks. In the context of demographic change the emerging burden of workers' chronic diseases caused by physical workload grows in importance. Hence, by carrying out risk assessments by means of ergonomic assessment tools, the physical workload is evaluated in order to take measures of ergonomic workplace design. The Ergonomic Assessment Worksheet (EAWS) of the MTM-Association as well as Audi's so-called Workplace Structure Analysis (APSA) are validated assessment tools, which can be used to evaluate and improve the employee's working conditions. By help of motion capture the worker's body postures and motions can be recorded and converted into a data stream of ergonomic influencing factors, which can be interpreted automatically based on either ergonomic assessment method, with the aim to increase both efficiency and objectivity of the ergonomic analysis. An inertial motion capturing system is currently under validation with the purpose of evaluating its potential as an ergonomic assessment tool. The present paper will give an insight into the validation process and present first results.

Keywords

Motion Capture, Ergonomic Assessment, Digitization

1 INTRODUCTION

Competitive manufacturing is not only related to productive production systems, but also to efficient production planning processes. In order to increase both the velocity and the quality of the planning results knowledge valorization is a strategic competitive advantage. The age of digitization can help to leverage its full potential by combining existing Industrial Engineering (IE) know how with data and information on the product, the work process as well as the worker. The current conditions of overwhelming digitization represent a paradigm change and require a digital transformation process of the IE in manufacturing companies, especially in the automotive sector. Therefore, new planning methods, digital planning tools and automated evaluation routines need to be developed and implemented into the production planning process.

At AUDI AG, the IE department is responsible for designing efficient and ergonomic production processes and workplaces. To fulfil this obligation, capturing of working motions is essential in order to gather data on body postures, forces and loads for manual material handling processes as well as information about the workload on the upper limbs. This data can not only be used for the ergonomic assessment of workplaces, but also for (temporal) planning of production processes. Currently, ergonomic information and process data are gathered separately using various methods. Audi investigates the potential of motion capture to gather all relevant data from recorded working sequences by help of an inertial sensor system and force gloves with piezo pressure sensors in order to generate ergonomic (and perspectively time) analyses automatically. In a comprehensive validation process firstly the hardware, the sensor network as well as the force gloves, are evaluated. Secondly, the conformity with the rules of the ergonomic assessment tools of Audi and German MTM Association are verified. The results of the validation process are presented in this publication.

2 METHODS

The following paragraph gives an overview of the methods used in the validation study, firstly the ergonomic assessment tools, and secondly the methodology of the validation process itself.

2.1 Ergonomic Assessment Tools

With the purpose of meeting labor law requirements, employers are obliged to conduct risk analyses by help of ergonomic assessment tools. In the following, chapter, the tools used in this study are introduced.

2.1.1 Ergonomic Assessment Worksheet (EAWS)

The Ergonomic Assembly Worksheet (EAWS) is a screening method to assess physical workload, which is used by major European companies of the automotive sector and was validated in the field [1]. It was developed by a consortium with partners from occupational safety and health institutes, universities (namely the IAD of Darmstadt University of Technology), industrial companies and MTM. EAWS grants load points for unfavourable physical workload in four different sections, each one covering a specific risk area: body postures, action forces, manual material handling and upper limbs in repetitive tasks. As a 1st level risk assessment tool, for each section the risk score for ergonomically unfavourable conditions for given work tasks is calculated based on the corresponding 2nd level risk assessment methods, such as: OWAS for body postures, RULA as well as Snook & Ciriello for action forces, NIOSH for manual material handling, and OCRA Index for upper limb assessment in repetitive tasks [2]. An overall risk score is calculated and its status is indicated based on a traffic light scheme, which not only gives a clear and immediate picture of the potential risk occurring, but also helps to determine actions for improvement.

2.1.2 Arbeitsplatz-Struktur-Analyse (APSA)

At Audi the so-called Arbeitsplatz-Struktur-Analyse (Workplace-Structure-Analysis), which is the result of an in-house development, is used for the ergonomic evaluation of the industrial workplaces [3]. It was implemented in 2006 and is based on the Automotive Assembly Worksheet (AAWS) that is actually the predecessor of EAWS [4]. By help of APSA the physiological workload at a specific workstation is measured. The workload is assessed for a given working method executed according to a given production plan (quantity and production mix), depending on the respective conditions, such as workplace geometries, postures, equipment or parts. APSA considers, firstly the body posture (trunk, arm & leg posture), secondly forces (finger, hand, arm/ whole body forces and awkward wrist postures) and thirdly manual material handling (carrying, repositioning, pushing/ pulling). The evaluation of the individual ergonomic criteria coalesces to an overall risk score, which is also represented by a traffic light scheme and helps firstly to identify the needs for workplace design by pointing out ergonomic deficits, and secondly to determine organizational measures (employee deployment, job rotation) [5].

2.1.3 Motion Capturing

Motion Capture (Mo-cap) in general refers to the process of recording motions of objects or people. This could be realized either with optical (e.g. marker based) or non-optical systems, such as inertial or magnetic systems for example [6]. In the context of computer-aided ergonomics [7], motion capture techniques are already applied frequently to obtain the dynamic and natural motion information for human simulation tools [8]. Apart from that, motion capturing systems can also help to record the worker's body postures and motions, which can be evaluated automatically based on either ergonomic assessment method, with the aim to increase both efficiency and objectivity of the ergonomic analysis.

In the presented study, an inertial motion capturing system (of AXS Motionsystem Kft., Hungary) is used that was specifically developed for use in industrial environment. The sensors are robust and are neither influenced by magnetic fields caused by machinery, nor do they interfere with the technological equipment on the production floor, since they are not conntect via WiFi, but hard-wired. AXS' motion capturing system consists of 18 inertial measurement units (IMU) that are placed on defined body parts, as shown in figure 1.

Figure 1 - Sensor positions on distinct body parts (Source: AXS Motionsystem Kft.)

The IMU sensors measure velocity and orientation and typically use two different kinds of sensors. Gyroscopes measure angular velocity, whilst accelerometers are used to measure linear acceleration. Rotation angles can be measured by gyroscopes (pitch, roll and yaw) and accelerometers (pitch and roll) [9]. Both sensors have three degrees of freedom (DOF) defined for x, y, and z-axis. In combination of both sensors six DOF are provided. The AXS human body reference system is illustrated in figure 2. Besides that, the system uses piezoelectric sensors, which are integrated in gloves to measure hand forces. By help of insole pressure sensors the walking distance is recorded.

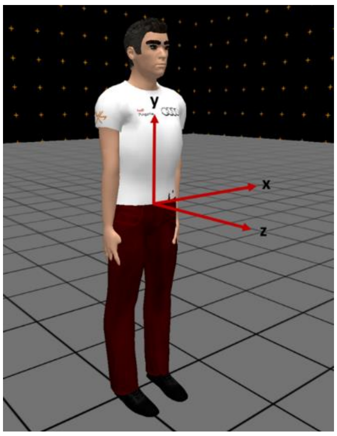


Figure 2 - Human body reference system

The motion data recorded is being pre-processed by the system and converted into a data stream of ergonomic influencing factors, which can be interpreted automatically based on either ergonomic assessment method, such as APSA or EAWS.

2.2 Measurement Validation

With the aim to validate both the hardware and the software of AXS' motion capturing system, a multi-stage validation concept was set up.

2.2.1 Validation concept

Measurement systems have to fulfil four main criteria in order to ensure their functionality and the quality of the measured data in particular: precision, repeatability, reproducibility and stability. *Precision* describes how close repeated measures are to each other and constitutes a random error component of a measurement system. *Repeatability* defines the variation in measurements gathered with one measuring tool when used several times by one assessor while measuring identical attributes. It shows within-system variations and the capability or potential of a measurement system. *Reproducibility* states the average variation of the measurements data made by different assessors while using the same measurement system and when measuring one identical attribute. Measurement variation may be caused by assessor, environment or method. *Stability* describes the change in bias over time respectively the so-called drift [10]. This is very important, since it is known that gyroscopes have drift errors over long periods of time, whilst acceleration sensors are sensitive to acceleration in any direction when objects show translational acceleration or fast rotation [11,12].

In order to address all these single aspects, a multi-stage validation process was set up. First of all, the hardware, both the IMU sensors and pressure gloves, need to be verified. Apart from that, the software itself is proved, not only the conformity with the rules of the respective ergonomic assessment tools (mentioned before), but also the software functionalities as well as its usability. This paper mostly focuses on the hardware validation. Firstly, every single IMU sensor will be checked in one-dimensional motions, secondly the interaction of the sensors applied on the respective body parts to track complex motions of the extremities for example, thirdly by assessing real working situations the parallel-test reliability towards observer-based EAWS- and APSA-analyses will be checked.

2.2.2 Scientific test set-up

The so-called Human Body Angle Assessment Model (HASS) was designed to ensure objective angle adjustments (see figure 3). Its body part proportions are equivalent to the 50 percentile male (according to DIN 33402-2). Angle adjustments were measured by using a goniometer.

Figure 3 - Human Body Angle Assessment Model (posterior view)

With the objective to validate measurement *precision* of AXS' measurement system with regard to the motions and postures of the upper limbs three different angles with ten repetitions where measured at the shoulder joints. Measured angles were defined referring to the frontal plane around the transverse axis and according to the threshold angle of EAWS, 20° retroversion and 60° as well as 90° anteversion (see Figure 4). One experiment was conducted with an inclined sensor of 30° in order to investigate the influence of deviating sensor positions. Additionally, elbow joint flexion and hand wrist extension were recorded for 90°, 45° and 20°. To ensure *repeatability* of the measurements, 24 motions of each limb were executed within a time frame of 15 sec for each repetition, evaluating the angle for 10 sec in steady-state and 5 sec of transition to reassume the initial position. Since it is known that gyroscopes (such as the IMU sensors) have drift errors over long periods of time, *stability* was checked by measuring three times 60° left shoulder joint anteversion angle for 15 min. *Reproducibility* is proved by conducting multiple measurement series with individual calibration measurements under identical conditions.

Furthermore, conformity with APSA and EAWS is validated using test cases for every single workload factor represented in the respective ergonomic assessment tool as well as every single rule. The firstly measured position was an above shoulder height position, as considered in EAWS Section 1 and in APSA's line 19b(A). Therefore, a fixed angle of 80° in shoulder joint was pre-set. The elbow joint was moved 24 times with a direction of 90° flexion. This position was held for 10 sec. After 10 sec, 90° extension was executed to reach the initial position again with a rest of 5 sec (as briefly described before). The evaluation is based on a cycle time of 1.5 min, which is representative for the automobile industry. The results of the risk scores were compared to the results of analysis with EAWS and

APSA using the parallel test method to ensure inter-method reliability [13].

3 RESULTS

The following paragraphs give an insight into the results of both the hardware and the software validation.

3.1 Hardware Validation

The *precision* of the measuring system was evaluated at various pre-set angles for the joints of the upper limbs. The pre-set shoulder joint angle of 90° varied with a measurement error of 1.7% ($\bar{x}$ = 88.51°, SD = 0.93°). The measurement error was 1.1% for a pre-set angle of 60° ($\bar{x}$ = 59.32°, SD = 0.16°). The 20° shoulder retroversion angle showed a measurement error of 8.1% ($\bar{x}$ = -18.37°, SD = 0.15°). The results are shown in Table 1.

Pre-set x-rotation	Sensor position	Mean	SD	Measurement error
90°	correct	88.51°	± 0.93°	1.7%
60°	correct	59.32°	± 0.16°	1.1%
60°	30° tilted	59.26°	± 0.17°	1.2%
-20°	correct	-18.37°	± 0.15°	8.1%

Table 1 – Measurement results of shoulder joint for x-axis rotation (N = 10)

Figure 4 exemplarily illustrates repeated measurements of shoulder anteversion with an expected x-rotation of 60°. The measurement error was 1.1% ($\bar{x}$ = 59.32°, SD = 0.16°). The data shown in the graph attests the measuring system's *repeatability*.

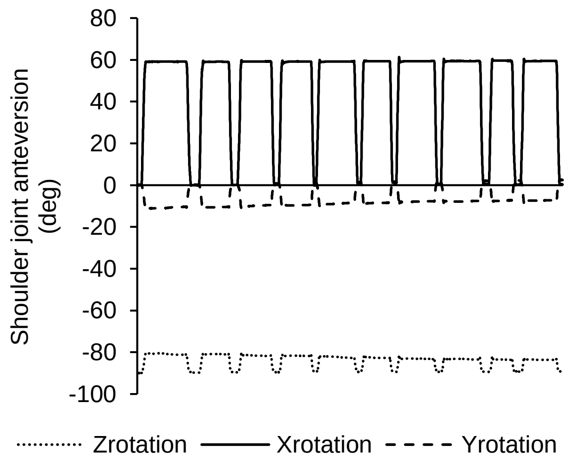


Figure 4 - Repetitive measurements of shoulder anteversion (x-rotation) (N = 10)

As shown in Table 2, the measurement error was 4.5%, when the pre-set elbow flexion angle was 90° repetitively ($\bar{x}$ = 85.94°, SD = 0.88°). The pre-set elbow angle of 45° varied with a measurement error of 3.9% ($\bar{x}$ = 43.23°, SD = 0.79°) and the 20° angle showed a measurement error of 1.2% ($\bar{x}$ = 19.77°, SD = 0.23°).

Pre-set x-rotation	Sensor position	Mean	SD	Measurement error
90°	correct	85.94°	± 0.88°	4.5%
45°	correct	43.23°	± 0.79°	3.9%
20°	correct	19.77°	± 0.23°	1.2 %

Table 2 - Measurement results of elbow joint for x-axis rotation (flexion) (N = 10)

The pre-set wrist extension angle of 90° varied with a measurement error of 2.6% ($\bar{x}$ = 87.70°, SD = 1.12°) and the 45° angle showed a measurement error of 6.2% ($\bar{x}$ = 42.22°, SD = 0.43°). The measurement error was 12.0% when the pre-set wrist angle of 20° was measured repetitively ($\bar{x}$ = 17.60°, SD = 0.54°) (see Table 3).

Pre-set x-rotation	Sensor position	Mean	SD	Measurement error
90°	correct	87.70°	± 1.12°	2.6%
45°	correct	42.22°	± 0.43°	6.2%
20°	correct	17.60°	± 0.54°	12.0%

Table 3 - Measurement results of wrist joint for x-axis rotation (N = 10)

As shown in figure 5, over a time period of 15 min the values drifted in positive direction first. After 8 min ($\bar{x}$ = 58.87°, SD = 0.02°), a deviation of 0.29% occurred compared to baseline ($\bar{x}$ = 58.70°, SD = 0.02°). A measurement drift of 0.15% was recorded after 15 min ($\bar{x}$ = 58.61°, SD = 0.03°) compared to baseline. A maximum difference of 0.44% was determined between 8 min and 15 min, proving the systems high *stability*.

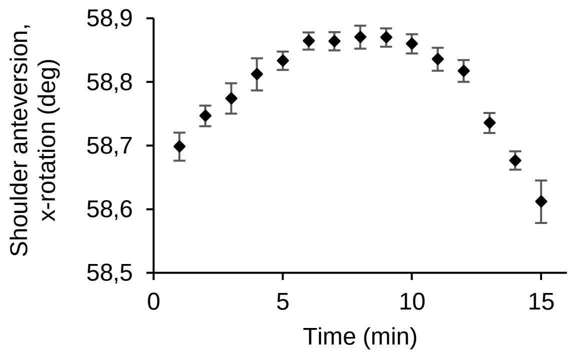


Figure 5 - Measurement drift of left shoulder joint angle in x-axis over a period of 15 min (values in mean ± SD) (N = 24)

Software Validation

The main purpose of the software validation was to check the conformity of the results with the rules of the ergonomic assessment tools of Audi and German MTM Association are verified. As briefly described above, this was done by help of specifically designed test cases for every single workload factor represented. A first measured position was a simulated above shoulder height position, which is defined as position of the wrist respectively the elbows above the shoulder (but not above head) in APSA's line 19b(A) and in EAWS Section 1.

For the ergonomic assessment, workload situations are always evaluated based on intensity and duration or frequency of the respective workload factor. Hence, with regard to arm postures, both the type of posture (arm above shoulder) and the proportion of the cycle time, during which the worker is exposed to this workload, are considered.

In APSA, line 19b(A) is assessed, resulting in a risk score of 44 points for 2/3 of the cycle (6 times 10 sec, cycle time of 90 sec) (see figure 6).

19b – Arm posture

Arm posture	Time proportion [% of cycle]					
	[1] ≤ 10	[2] ≤ 20	[3] ≤ 33	[4] ≤ 50	[5] ≤ 67	[6] ≤ 100
[A] above shoulder	6	14	24	32	44	64
[B] above head	10	22	38	52	68	100

Figure 6 - APSA risk score for the simulated above shoulder work

In EAWS, line 6 is assessed. Thus, the result is a risk score of 51 points for 2/3 of the cycle (see figure 7).

		[%]	5	7,5	10	15	20	27	33	50	67	83
		[s/min]	3	4,5	6	9	12	16	20	30	40	50
		[min/8h]	24	36	48	72	96	130	160	240	320	400
Standing (and walking)												
1		Standing & walking in alteration, standing	0	0	0	0	0,5	1	1	1	1,5	2
4		b with suitable support	2	3	5	7	9,5	12	18	23	31	38
5		Upright with elbow at / above shoulder level	3,3	5	8,5	12	17	21	30	38	51	63
6		Upright with hands above head level	5,3	8	14	19	26	33	47	60	80	100

Figure 7 - EAWS risk score for the simulated above shoulder work

The results calculated by the AXS software differ slightly, as shown in table 4. The deviation of the Mo-cap results from the evaluation with APSA and EAWS is caused by a misinterpretation of the exposure time as the AXS system's calculation is based on the real working pace. According to the recordings, the exact time the overshoulder position is maintained is 61.8 sec in total. With this said, the results calculated based on the real-time recording have to be considered as correct for both evaluation methods, APSA and EAWS. Anyhow, this shows the problem of the results' dependency on the worker's individual working method and his individual working pace. Instead, the ergonomic assessment is conducted for a workplace, so that the calculation of the results should rather be prorated to a standard performance.

	Pen & Paper method	AXS Mo-cap
EAWS	51	53.5
APSA	44	54

Table 4 – Comparison of mo-cap results with the respective risk scores from eaws and apsa

4 CONCLUSION

The results presented above focused on the validation of the left hand-arm-shoulder system in one dimension (x-axis) by checking the IMU sensors precision, repeatability, reproducibility and stability. In a next step, the sensors for the hand-arm-shoulder system need to be validated for the remaining dimensions (in y- and z-axis), before those for the lower extremities and the upper body are evaluated. The *precision* can be considered as very high, but although the difference of the measured angle is mostly below 1°, for smaller angles the relative deviation is becoming more crucial in terms of accuracy. The *repeatability* of the measurements was also proved, since the precision of the results remained relatively stable throughout the repeated posture changes. It was also found out that even the alignment of the sensor position only has a small impact, which underlines the practicability of AXS' Mo-cap systems in terms of robustness towards user deviation. The *stability* was assessed in multiple repetitive measurement series over a duration of 15 min. Although, it was found out that the sensor drift is also quite low (maximum of 0.44%), it can be assumed that this will have an impact on more complex movements. In this case, the measurement errors of the IMUs alongside the sensor chain of the extremities sum up, so that the motions and postures of distal body parts will most likely become inaccurate after a certain time (probably after about 7 mins). The reproducibility was not assessed yet, but will be checked in a next validation step by conducting multiple measurement series with individual calibration measurements under identical conditions.

Knowing that the sensors are working correctly, in the next validation step the interaction of the IMU sensors needs to be validated in order to evaluate the reliability of complex motions that are being tracked, as shown by help of the simulated above shoulder height working situation. Last but not least, real working situations will be assessed to find out about the reliability of the recording of complex three-dimensional motions that are recorded.

Although the presented paper is focusing on the inertial Mo-cap system from AXS, which has been validated in the study, the introduced validation process can be considered as general approach to validate motion capturing systems.

In conclusion, the results show the potential of motion capture systems to digitize ergonomic workplace analyses in order to increase the efficiency of the assessment processes as well as the objectivity of the evaluation results. With this, as part of Audi's digitization strategy, Mo-cap systems can help to improve both the velocity and the quality of the production planning process and to speed up the time-to-market with the purpose to exploit the competitive advantage.

5 REFERENCES

[1] Schaub, K., Caragnano, G., Britzke, B., Bruder, R., 2012, The European Assembly Worksheet, Theoretical Issues in Ergonomics Science (14), 1-23.

[2] Lavatelli, I., Schaub, K., Caragnano, G., 2012, Correlations in between EAWS and OCRA Index concerning the repetitive loads of the upper limbs in automobile manufacturing industries. Work (41), 4436-4444.

[3] Bogus, T., Dorn, R., 2010, Berücksichtigung des Arbeitsplatzbedarfs für einsatzkritische leistungsgewandelte Mitarbeiter im Planungsprozess, In: Proceedings of 57th Frühjahrskongresses der Gesellschaft für Arbeitswissenschaft, 265-268, GfA-Press, Dortmund.

[4] Schaub, K., Ghezel-Ahmadi, K., 2007, Vom AAWS zum EAWS – ein erweitertes Screening-Verfahren für körperliche Belastungen, In: Proceedings of 53rd Frühjahrskongresses der Gesellschaft für Arbeitswissenschaft, 601-604. GfA-Press, Dortmund.

[5] Bös, G., 2007, Strategisches Management alternder Belegschaften bei der AUDI AG. In: M. Holz & P. Da-Cruz (Eds.), Demografischer Wandel in Unternehmen. Herausforderung für die strategische Personalplanung. Springer, Berlin.

[6] Foxlin, E., Inc, I., 2002, Motion Tracking Requirements and Technologies. Handbook of virtual environments: Design, implementation, and applications.

[7] Karwowski, W., Genaidy, A., Asfour, S., 1990, Computer-aided ergonomics. Taylor & Francis, Abington.

[8] Ma R., Chablat D., Bennis F., Ma L., 2011, A Framework of Motion Capture System Based Human Behaviours Simulation for Ergonomic Analysis. In: C. Stephanidis (Eds.), Proceedings of HCI 2011. Communications in Computer and Information Science. Springer, Berlin.

[9] Ahmad, N., Ghazilla, R. A. R., Khairi, N. M. Kasi, V., 2013, Reviews on Various Inertial Measurement Unit (IMU) Sensor Applications, International Journal of Signal Processing Systems 1(2), 256-262.

[10] AIAG, 2010, Measurement Systems Analysis, MSA. Automotive Industry Action Group, Michigan.

[11] Filippeschi, A., Schmitz, N., Miezal, M., Bleser, G., Ruffaldi, E., Stricker, D., 2017, Survey of Motion Tracking Methods Based on Inertial: A Focus on Upper Limb Human Motion. Sensors, 17, 1257.

[12] Zhao, H., Wang, Z.Y., 2012, Motion measurement using inertial sensors, ultrasonic sensors, and magnetometers with extended kalman filter for data fusion, IEEE Sensors Journal, 12(5): 943-953.

[13] Davidshofer, Kevin R. Murphy, Charles O., 2005, Psychological testing: principles and applications. Pearson/Prentice Hall, Upper Saddle River.

6 BIOGRAPHY

Ralph Hensel obtained his PhD degree in Work Science (Industrial Engineering) from the Chemnitz University of Technology. Since 2016 he is Senior Expert for Ergonomics and Industrial Engineering at Audi's Industrial Engineering Methods Department.

Micha Steinmetz studies M.Sc. Human Technology in Sports and Medicine at German Sport University Cologne. He currently works at Audi's Industrial Engineering Methods Department in Ingolstadt.

Lukas Hausmanninger obtained a B.E. in Technical Orthopaedics and a M.Sc. degree in Sports Technology from German Sport University Cologne. He currently is a PhD student at the Industrial Engineering Methods Department at AUDI AG in Ingolstadt.

Mathias Keil is currently the Head of the Industrial Engineering Methods Department at AUDI AG in Ingolstadt. He holds a PhD degree in Work Science (Industrial Engineering) from Chemnitz University of Technology.

Martin Benter is research assistant at the MTM-Institute of the German MTM-Association. There he is involved into the development of the MTM process building block systems.

Prof. Peter Kuhlang is head of the MTM-Institute and the MTM-Academy of the German MTM-Association. He also coordinates the international developments within the IMD (International MTM directorate).

Exoskeletons as Human-Centered, Ergonomic Assistance Systems in Future Competitive Production Systems

R. Hensel[1], M. Keil[1]
[1]Dept. Industrial Engineering Methods, AUDI AG, Ingolstadt/ Germany

Abstract
In addition to technical and organizational measures for ergonomic workplace design, exoskeletons can be considered as innovative ergonomic assistance systems to support workers in production and logistics in order to reduce their physical workload. With this, exoskeletons cannot only help to prevent musculoskeletal diseases, but also to augment workers with physical limitations in cases of inclusion or occupational reintegration. Currently, numerous open questions hinder their implementation in industrial applications. For the successful implementation of occupational exoskeletons, they have to fulfil several ergonomic, work physiological and safety-related requirements. Based on the experiences in piloting and implementing occupational exoskeletons at AUDI AG, practice-oriented aspects for their usage in industrial application scenarios will be compiled based on the results of a field study with a passive industrial exoskeletons for lower-back support. The goal is to provide assistance for piloting and implementation of occupational exoskeletons in industrial practice.

Keywords
Exoskeletons, Ergonomic Assistance Systems, Field Study

1 INTRODUCTION

Exoskeletons were developed for usage in medical rehabilitation and are also applied for military purposes. With first application scenarios for the utilization of active and passive exoskeletons in industrial practice a disruptive change in the field of ergonomic assistance systems is ahead [1]. In the context of industrial application, an exoskeleton can be defined as body-worn ergonomic assistance system that acts mechanically on the worker's body to reduce his physical workload. Particularly in situations in which the potential of technical and organizational measures for ergonomic workplace design is limited, exoskeletons can be considered as innovative ergonomic aids and support workers in manufacturing as well as logistic processes. In times of demographic change, exoskeletons can also assist workers with physical limitations in cases of inclusion or occupational reintegration processes. With this, exoskeletons potentially give a cutting edge in competitive manufacturing processes.

At AUDI AG, about 43% of the musculoskeletal workload is affecting the lower back, due to (dynamic) manual material handling tasks as well as static, forward-bending postures. Audi focusses on ergonomic workplace design, since work physiological research proves that workers, which are often subjected to repeated load lifting and carrying activities or required to maintain postures of trunk flexion over longer periods, are increasingly exposed to the potential for long-term back injury with a lifetime prevalence of up to 80% [2]. In the context of demographic change the emerging burden of workers' chronic diseases caused by physical workload grows in importance [3].

Employers are obliged by labor legislation to take measures to improve safety and health of workers, by assessing their workload and eliminating all risks through collective measures. In contrast, individual protective measures, such as exoskeletons, are of lower priority, but are promising where other measures are technologically impossible, interfering with the production process or too expensive.

Whilst the technological readiness level of active exoskeletons available is still relatively low [4], a few passive devices are commercially available (e.g., SuitX BackX models S and AC and Laevo), as well some "soft" exoskeletons (e.g., Hunic and Rakunie). For an extensive overview of the state of the art, the interested reader is directed to the article of Näf et al. [5]. For several reasons, the Laevo exoskeleton (www.laevo.nl, version 2.5) was identified as the most promising and suitable one to possibly reduce the workload on the lower back and support workers with physical limitations. Being a passive system, it absorbs a portion of the effective weight (of a load and/or the worker's body) by help of a chest pad and diverts the resulting forces through spring elements around the lower back and via tow leg pads into the thighs. Pneumatic gas springs are storing the kinetic energy of the worker's motions in order to sustain the forward-bending posture or support erecting his body while lifting a load. The system can be adjusted to the body size of the workers and to the working conditions, especially the working height as well as the level of support required. Whilst the efficacy (muscles activity, endurance and biomechanics) of the Laevo exoskeleton was evaluated in physiological laboratory studies [6, 7], no subjective evaluation of the system was provided, yet.

Therefore, the aim of the present field study was to evaluate the ergonomic benefit of the Laevo exoskeleton subjectively by assessing the lower back relief in two different working situations, static bending postures and dynamic repositioning of loads. Besides its efficacy with regard to workload relief, it was investigated how user acceptance is influenced by the exoskeleton's discomfort and usability, to determine success factors for its implementation.

2 METHODS

2.1 Selection of Participants and Workplaces

Since the main objective of the field study was to obtain a general understanding about the ergonomic effects of the Laevo exoskeleton only male workers were selected as participants. To cut down the risk of confounding, the amount of independent variables was reduced, since increased discomfort in the chest region due the specific anatomical conditions of female workers was expected. For the same reason, workers with lumbar pain history were excluded from participation. In total, 30 male test persons participated in the field study in Audi's logistics department, press shop and final car assembly at both German production facilities. Their average age was 29.2 (SD = ±10.6) years, mean body mass was 76 kg (SD = ±9) kg and mean body height 175.3 cm (SD = ±6.5) cm. The test persons had a work experience at their respective workplaces of at least 3 months, in average 56.7 (SD = ±78.9) months.

In order to find out about the beneficial effects of the Laevo exoskeleton regarding the workload relief, work tasks with static postures (18 subjects) as well as dynamic repositioning operations (12 subjects) have been assessed in a four-week field study.

Five workplaces with high workload on the lower back were selected at the assembly line and in the press shop, where the workers are working statically bent forward, assembling car parts or disassembling press tools. At the workplace "Ground screw connection footwell" (WP #1) at the assembly line of the Audi A4, five workers were obtained as voluntary participants. At the assembly line of the Audi A8, the Laevo exoskeleton was tested at the workplace "Trunk insulation" (WP #2) and "Installation cable harness" (WP #3) by two workers each. According to the job rotation regime, the workers wore the passive exoskeleton at the assembly lines on a daily base for about two hours. The cycle times vary between 1.4 and 5.4 minutes. In the press shop, nine workers were acquired from two different teams, the "maintenance team (WP #4) and press set up operators (WP #5), which used the Laevo exoskeletons only temporarily for physically demanding work tasks, when statically bent forward disassembling the press tools.

With the aim of evaluating the support of the exoskeleton in dynamic workload situations of manual material handling, workplaces in Audi's logistics department have been selected for the field study which are characterized by strong flexion angles of the back. Two different types of workplaces were used in the CKD packaging department and one in series logistics for the Audi A8. In the CKD packaging department (WP #6, WP #7), eight test persons used the Laevo exoskeleton for four weeks to lift car components out of a cargo carrier and to pack them into boxes for shipping. At those workplaces, the loads to be handled varied in size and weight between 3 to 20 kg. The picking workplace in series logistics (WP #8) is characterized by weights of up 15 kg, which need to be manipulated, when picking brake discs and calipers for the Audi A8. At this workplace, four workers used the Laevo exoskeleton for an entire week each.

To illustrate the working conditions of the two different working situations described, figure 1 exemplarily shows the workplaces at the assembly line (left), and the CKD packaging logistics (right).

Figure 1 - Workplaces assessed in the field study

2.2 Methodology

With the aim to conduct a subjective evaluation of the Laevo exoskeleton, a standardized survey tool was developed. The validity of which was ensured by adopting valid scales and items of existing questionnaires. Additionally, the field study was medically attended by Audi's occupational practitioners. For the purpose of participant selection, prior to the test the potential test persons were checked for medical contraindications. During and after the test, occupational medical examinations were conducted, firstly to assess occurring ailments caused by the use of the Laevo and secondly to rate the workload relief. To do so, work-related ailments were assessed by help of the Body Part Discomfort Scale [8] combined with a 7-point Likert scale prior to the test. The same scale was used during and after the test in order to monitor the symptoms and to find out about load redistribution as well as discomfort perceived. Wilcoxon signed-rank test was used to compare the paired samples of the assessments before and after the tests.

Furthermore, the perceived usability of the Laevo exoskeleton, when putting it on and off and when using it whilst performing the work task, was investigated based on the two items of UMUX lite

(Usability Metric for User Experience) [9]. Based on the Technology Acceptance Model (TAM2) [10], the behavioral intention to use the Laevo exoskeleton was rated by the test persons on a 7-point Likert scale, too. The statistical relationship between workload relief respectively discomfort and the intention of use and between usability and intention of use was assessed by calculating Spearman's rank correlation coefficient in order to describe its impact on the user acceptance.

In addition, the test persons were encouraged to participate in the further development of the Laevo by giving feedback about their positive or negative experiences by help of open-ended questions.

3 RESULTS

As shown in figure 2, for use of the exoskeleton in static working situations, it has been proven that the workload of the lower back is reduced significantly from 2.5 to 1.5 ($z=2.19$, $p=0.014$), with a large effect size ($r=0.5$). Besides that, also ailments caused by awkward wrist postures showed an improvement ($z=2.11$, $W=1.5 \leq W_{crit}=3$, $r=0.5$) when wearing the Laevo (as the workers do not have to support their posture with the arms anymore). At the same time, the study participants criticized significantly increasing discomfort in the chest area caused by pressure sensation and friction ($\tilde{x}=3$, $z=2.04$, $p=0.021$, $r=0.5$). Furthermore, the participants complained about increasing discomfort of the knees (hyperextended due to force insertion into thighs) ($r=0.4$) and thighs ($\tilde{x}=2.5$; $r=0.1$), not significantly though. These aspects are suggesting a load redistribution caused by both the chest pad and thigh pads. The usability was rated high by all workers for donning and doffing of the exoskeleton ($\tilde{x}=6$) as well as for using it when performing the task ($\tilde{x}=6$), but the usability decreased significantly ($z=2.34$, $p=0.01$, $r=0.5$) with progressing usage duration. At the same time, the intention of use dropped significantly ($z=2.71$, $p= 0.003$, $r=0.6$), although it was rated very high in the beginning of the test ($\tilde{x}=6$).

Figure 3 shows only slightly decreasing symptoms (with median scores from 2.0 to 1.0) of the lower back in dynamic working situations, and with that no significant workload relief through the use of the

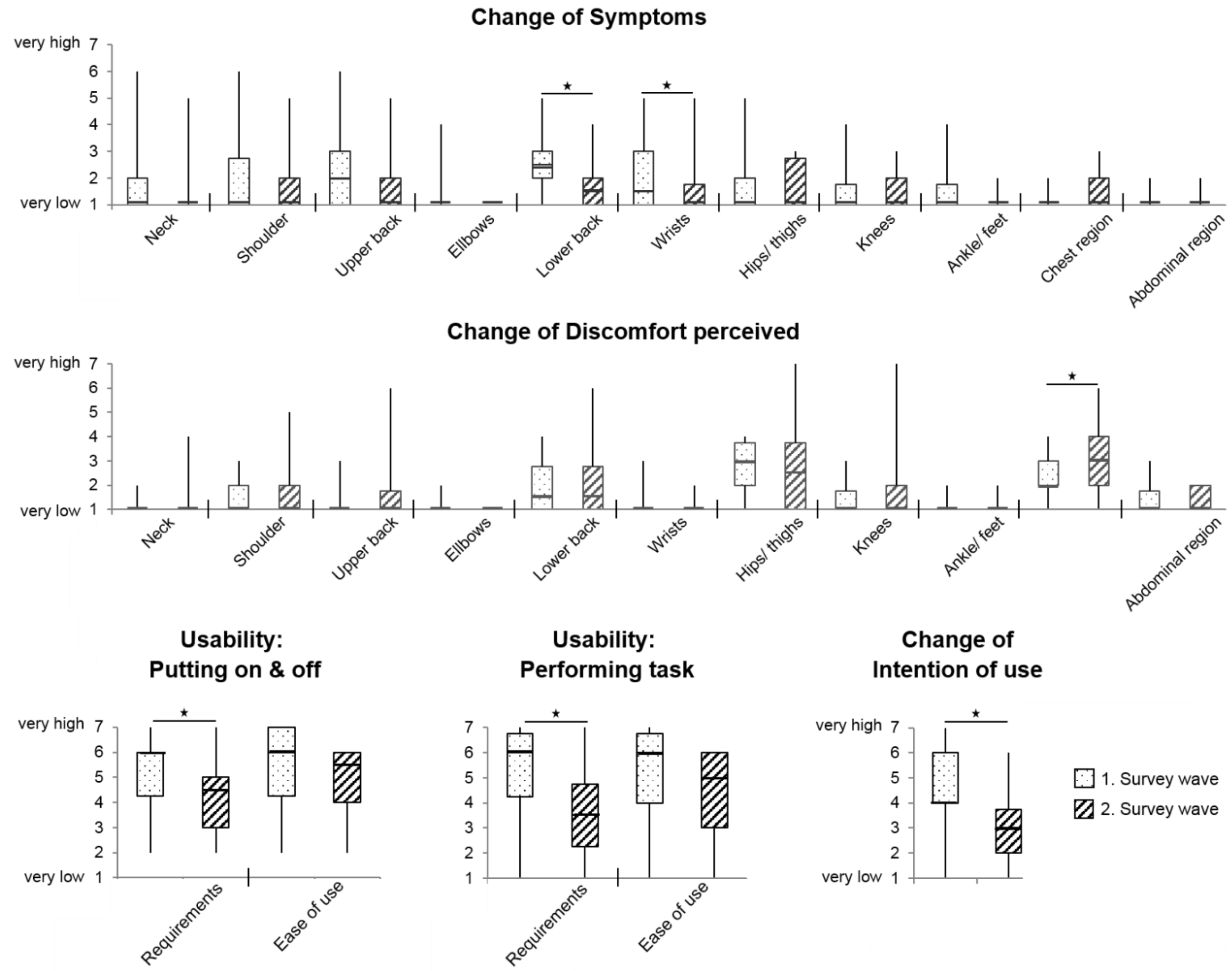


Figure 2 - Discomfort, usability perceived and intention to use of the test persons at static workplaces, changes between the beginning and the end of the test (N=18; *significant; α = 10%)

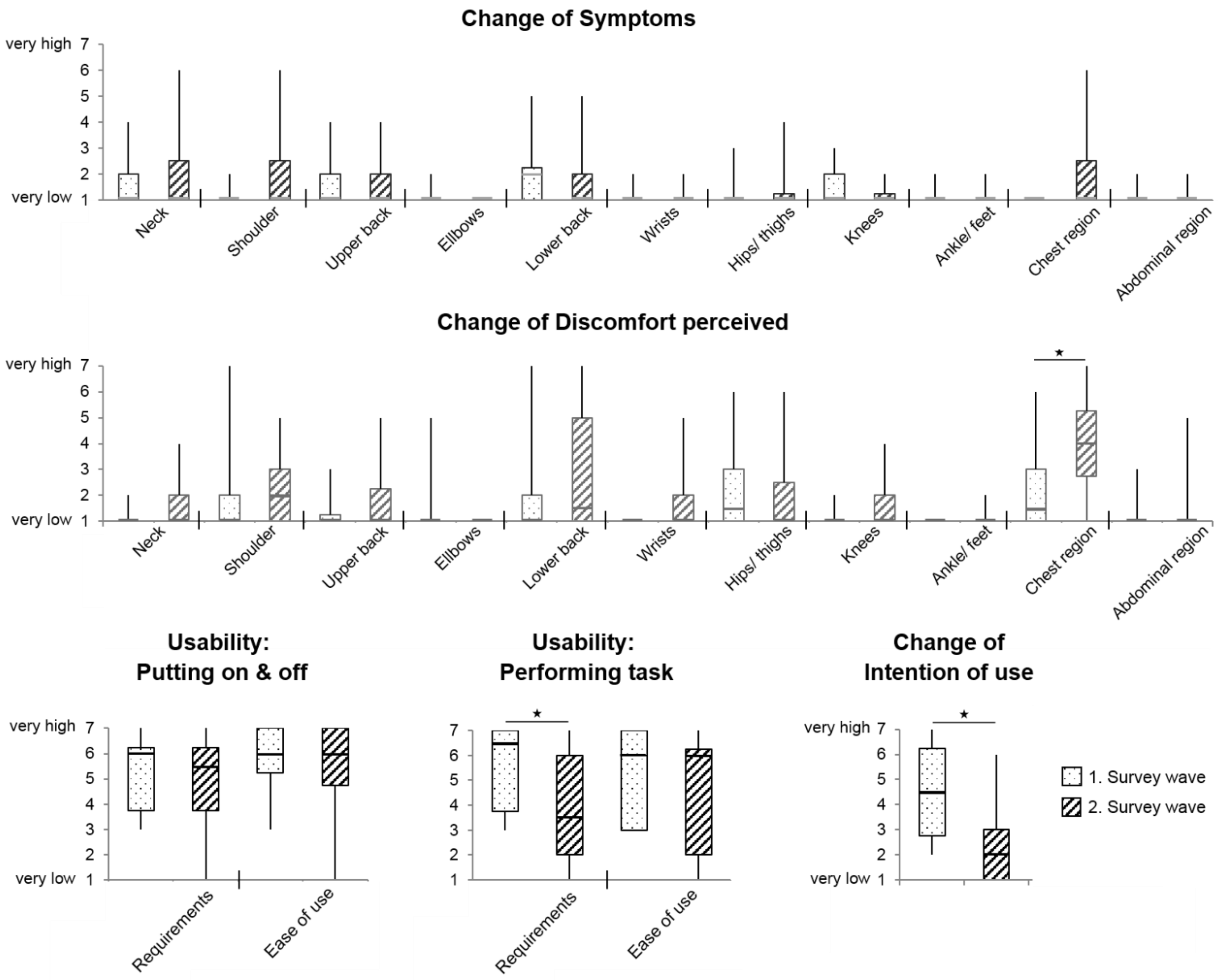


Figure 3 - Discomfort, usability perceived and intention to use of the test persons at dynamic workplaces, changes between the beginning and the end of the test (N=12; *significant; α = 10%)

Laevo. The discomfort perception in the chest region increased significantly (z=1.94, p=0.026) with an effect size of r=0.4 as well as in the shoulders (due to the straps) and knees (hyperextended) with a medium effect size of r=0.2. The usability was rated high by all the workers for donning and doffing the exoskeleton (x̃=6) and remained quite stable during the usage duration, in contrast to the static working situations, where the workers put them on and off more often due to the job rotation regime. When using the Laevo exoskeleton whilst performing the task, the usability decreased significantly (z=2.52, W=0 ≤ Wcrit=3) from 6.5 to 3.5 with a large effect size (r=0.5). The intention of use dropped significantly too (z=2.55, p= 0.005, r=0.5), although rated very high in the beginning of the test (x̃=5).

4 DISCUSSION

The results of the present field study show positive perception of the workers regarding the lower back support when using Laevo's exoskeleton. The results clearly indicate that the lower back support perceived is significantly higher in working situations with static postures, compared to manual material handling. However, it needs to be mentioned that the reduction of symptoms was relatively small, and work-related symptoms reported were very low in general.

Several significant correlations were found between workload relief, discomfort and usability, on the one hand, and the reported intention of use, on the other hand. Spearman's rank correlation coefficient was calculated and interpreted based on the effect size guidelines provided by Gignac & Szodorai [11], who recommend to consider correlations of 0.10, 0.20, and 0.30 as small, typical, and relatively large. There was a strong positive correlation (ρ=0.303) between perceived workload relief when using the exoskeleton and the intention of use (ρ=0.303) indicating that the perceived usefulness of the exoskeleton positively determines user acceptance. It was also proven that the discomfort perceived, mostly in the thoracic region, has a strongly negative impact on the acceptance (ρ=-0.385). Hence, even a minimal level of discomfort might hinder user's acceptance, although the back load relief is quite promising with regard to field implementation. Adverse outcomes predominantly caused by discomfort through the chest pad in the thoracic region and the leg pads on the thighs as well as the knees are barriers to its user acceptance. The hyperextension of the knees caused by the use of the exoskeleton might even result in long-term health risks, which should be further investigated. The strong positive association between rated usability and intention of use (ρ=0.389) points out the positive influence of the usability on user acceptance. Due to the fact that usability decreased significantly during usage of the exoskeleton, mainly because it is supposed to distract the execution of auxiliary tasks, the user acceptance also decreased drastically. By help of this statistical evaluation, it was shown, how perceived usefulness, usability and discomfort are determining Laevo's user acceptance.

According to the Technology Acceptance Model [8], user acceptance is mainly affected by the perceived usefulness as well as the perceived ease of use. The user judges the usefulness of a system based on the relative advantage [12], here by contrasting the workload relief perceived by using the exoskeleton with the discomfort experienced. The ease of use is predominantly determined by the usability of putting the exoskeleton on and off as well as of its use during main and auxiliary tasks. This underlines that apart from the efficacy of the exoskeleton, the subjective perception of the system determines the success of its implementation.

In general, limitations to this study are firstly related to the small amount of workers involved, and secondly to the representativeness of the sample, since female workers were reasonably excluded from participation. It is therefore necessary to review the findings of this study with an extended, more representative sample. Furthermore, workers with back pain history, who have been excluded from the study, should be involved too, since the Laevo could be potentially useful for them, after the ergonomic benefit for back load relief has been proven. Concomitant occupational health examinations during the test period should be conducted in order to gauge reactions and they development over the duration of the test run.

Furthermore, only the subjective perception of workers has been assessed in the study, whereas objectively measured physiological parameters could have helped to explain the findings more precisely. Hence, in additional work-physiology studies, not only the lower back support should be investigated by help of physiological measurements, but also the pressure distribution through the chest pad needs to be assessed in order to improve its constructional concept. By help of a longitudinal field study especially the long-term effect of the use of the Laevo exoskeleton on the knees should be investigated in order to eliminate doubts about health risk for the knees caused by their hyperextension. These aspects could form the basis to decide about the implementation of the Laevo exoskeleton in industrial applications in the sense of evidence-based ergonomics.

In addition to the subjective evaluation of the Laevo exoskeleton, the outcome of the study also proves that the methodological concept and the empirical tools are qualified as standardized procedure to validate exoskeletons subjectively. The developed questionnaire can also be used irrespectively of the specific system tested and provides the basis to compare the findings of future field studies on exoskeletons. The approach could even be used to predict the impact of technical improvements to exoskeletons on the user acceptance, based on ratings of the systems' usability and ergonomic efficacy. In order to utilize this potential, it would be important to validate the methodological approach of this study not only with an extended sample, but also with different exoskeletons.

In conclusion, the findings regarding the ergonomic efficacy of the Laevo exoskeleton correspond to the results of the work physiology studies mentioned above [6, 7], so do the findings about the discomfort in the chest region and the knees. Beyond that, the results of the study at hand were specified in real working situations providing a broader overview of possible application cases and giving implications on the user acceptance of the Laevo exoskeleton. The field study clearly helps to highlight potential for the further development of the system's constructional concept. Having completed this field study, it can be summarized that occupational exoskeletons have the potential to considerably reduce the underlying workload factors associated with the development of work-related musculoskeletal injuries. This could benefit both well-being of the worker and productivity of the work processes in order to ensure competitive manufacturing systems. Anyhow, the technological realization of the support principles determines both usability and discomfort of exoskeletons and is decisive for success or failure of their implementation in industrial applications, regardless their efficacy.

5 REFERENCES

[1] De Looze, M. P., Bosch, T., Krause, F., Stadler, K. S., O'Sullivan, L. W., 2016, Exoskeletons for industrial application and their potential effects on physical work load, Ergonomics, 59 (5), 671-681.

[2] Punnett, L., Fine, L. J., Keyserling, W. M., Herrin, G. D., Chan, D. B., 1991, Back disorders and nonneutral trunk postures of automobile assembly workers, Scandinavian Journal of Work, Environment & Health, 17 (5), 337-346.

[3] Keil, M., Hensel, R., Spanner-Ulmer, B, 2010, Fähigkeitsgerechte Prozessmodellbausteine zur Generierung altersdifferenzierter

Beanspruchungsprofile, Zeitschrift für Arbeitswissenschaft, 64 (3), 205-215.

[4] Anam K., Al-Jumaily, A.A., 2012, Active exoskeleton control systems: state of the art. Procedia Engineering, (41), 988-994.

[5] Näf, M.B., Koopman, A.S., Baltrusch, S., Rodriguez-Guerrero, C., Vanderborght, B., Lefeber, D. (2018) Passive Back Support Exoskeleton Improves Range of Motion Using Flexible Beams. Frontiers in Robotics and AI, 5.

[6] Baltrusch, S. J., van Dieën, J. H., van Bennekom, C. A. M., Houdijk, H., The effect of a passive trunk exoskeleton on functional performance in healthy individuals, Applied Ergonomics, 72 (10), 94-106.

[7] Bosch, T., van Eck, J., Knitel, K., de Looze, M., 2016, The effects of a passive exoskeleton on muscle activity, discomfort and endurance time in forward bending work, Applied Ergonomics, 54 (5), 212–217.

[8] Corlett, E. N., Bishop, R. P., 1976, A technique for measuring postural discomfort, Ergonomics, 19 (2), 175-182.

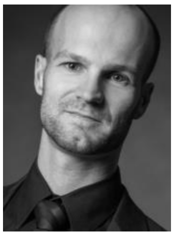
[9] Lewis, J. R., Utesch, B. S., Maher, D. E., 2013, UMUX-LITE: when there's no time for the SUS. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (2099-2102). ACM, New York.

[10] Davis, F. D., 1989, Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology, Computer and Information Systems Graduate School of Business Administration, Michigan.

[11] Gignac, G.E., Szodorai, E.T., 2016, Effect size guidelines for individual differences researchers. Personality and Individual Differences (102), 74–78.

[12] Rogers, E., 2003, Diffusion of Innovations, Free Press, New York.

6 BIOGRAPHY



Ralph Hensel obtained his PhD degree in Work Science (Industrial Engineering) from the Chemnitz University of Technology. Since 2016 he is Senior Expert for Ergonomics and Industrial Engineering at Audi's Industrial Engineering Methods Department.

Mathias Keil holds a PhD degree in Work Science (Industrial Engineering) from Chemnitz University of Technology. He is currently the Head of the Industrial Engineering Methods Department at AUDI AG in Ingolstadt.

Development of a Structured Approach for Reactive Disruption Management in Supply Chain Networks - an Integrated Perspective

S. Helming[1], C. Haverkamp[1], B. Haefner[1], G. Lanza[1]
[1]wbk Institute of Production Science, Karlsruhe Institute of Technology, Germany

Abstract
Increasing complexity and dynamics in globally distributed supply chain networks makes companies vulnerable to disruptions and does not only require good and robust planning, but also rapid troubleshooting as a reaction to unforeseen events. However, companies often neither possess a targeted approach for the systematic identification and communication of disruptions nor for the appropriate reaction to a disruption and the minimization of its consequences while considering the entire supply chain. Hence, this paper proposes a framework for reactive disruption management which allows for a multi-criteria evaluation and logical selection of measures in case of a disruption.

Keywords
Disruption Management, Integration, Supply Chain Network

1 INTRODUCTION

As a result of the advancing globalisation, value creation is increasingly organized in distributed supply chain networks (SCNs) today. However, this does not only intensify the interdependencies among the partners, but also raises uncertainty and complexity. The increasing unpredictability related with this hampers an efficient planning and operation of global SCNs, leading to a higher vulnerability to disturbances. According to a study by [1], 99% of the 261 surveyed participants stated that they had suffered from a disturbance in the supply chain (SC) within the last five years. Furthermore, almost 60% of them assessed the resulting disruption as serious problem [1]. Consequently, reactive SCs being able to quickly select and perform counter measures will have a competitive advantage [2]. Alarmingly, however, companies often still suffer a deficit of the systematic evaluation and selection of such measures [3]. In accordance with a study by [4], 80% of the respondents reported that they were not or only insufficiently equipped with plans for action for disturbances frequently leading to SC disruptions. The study furthermore reveals that a comprehensive analysis and monitoring of potential risks for the minimization of SC disruptions is mostly missing up to now [4]. Additionally, disturbances occur at different points in the SCN (production, logistical hub, supplier) [1], resulting in the fact that measures for mitigating the resulting disruption also have to be taken in different areas of responsibility (e.g. production, logistics, or procurement). Consequently, a stronger joint approach of the different areas of responsibility, and especially of production and logistics, could be seen as a means to consequently overcome SC disruptions [1]. While integrated models at a strategical or tactical planning level have been the subject of many studies, research on the more detailed, operational scheduling level has only just begun [5] and industrial companies often lack an integrated disruption management considering both production and logistics [2].

Summarizing the explanations above it can be deduced that there is an increased need for action within companies for the systematic selection and implementation of suitable measures as an ad-hoc reaction to an underlying disruption, especially in the era of big data. The stronger joint orientation of production and logistics might thereby be helpful to unleash hidden potentials [1]. Hence, this paper aims at introducing a framework for an intelligent reactive disruption management process (DMP) being carried out by a central control system (CCS) integrating the formerly decentralized logistics and production systems [6]. The overall objective thereby lies in the reduction of delays caused by disruptions, whereby disruptions and potential measures are explicitly considered both from the production as well as from the logistics perspective.

In accordance with previous disruption management approaches such as the ones proposed by [7] and [8], our approach particularly includes the five steps of (i) reporting a disruption, (ii) assessing its possible impact, (iii) deriving the principally applicable scope of action taking multiple criteria into consideration, (iv) conducting a benefit analysis to select the best possible measure out of the applicable ones, and (v) implementing this measure by updating the plan. Thus, the presented approach allows for a structured filtering process which, by a logical step by step exclusion of non-applicable measures and an evaluation of the remaining ones, selects the best possible measure using various criteria. The framework is exemplified for an aircraft manufacturer.

The remainder is organized as follows: Chapter 2 introduces the considered SCN and outlines fundamentals on disruptions and the DMP. Besides, relevant literature on disruption management is summarized. Based on the overall procedure for (re)-planning, Chapter 3 illustrates the different steps of the novel DMP and exemplifies them for an aircraft manufacturer. Chapter 4 concludes with an outlook.

2 FUNDAMENTALS AND STATE OF THE ART

As outlined above, this chapter first of all gives an overview of the regarded SCN and the fundamentals being essential for the further understanding.

In line with the definition proposed by [9], the term “supply chain network” is thereby used as a synonym for the term “production network”. With respect to its definition, a SCN consists of a number of value-adding external or internal suppliers as well as manufacturers [9]. Similar as in [9], the SCN in this work refers to a focal enterprise which is served by several suppliers delivering raw materials or intermediate products (see Figure 1). Contrary to the understanding in [9], however, distribution centres and final customers are not considered in this paper.

As shown in Figure 1, the activities within the SCN can be described by means of multiple interlinked process units capable to perform different processes (singular processes at the supplier plants are not included). A process is thereby characterized by the output it produces, its planned start and end time and by the inputs it requires. Inputs thereby consist of three different groups: materials (transported goods or input materials provided by other processes), human resources, and technical resources (vehicles, machines and other equipment such as sensors or tools that are needed for transportation/production). This understanding builds upon the differentiation between potential and repeating factors in [10]. To allow for a more precise characterization of a disruption in the course of this paper (see Chapter 3.2.1), however, potential factors are further broken down into human and technical resources.

As stated in Chapter 1, such SCNs, as a result of the increasing complexity and dynamics, are prone to disturbances. Literature thereby contains a large number of definitions of disturbances. In [10], e.g.,

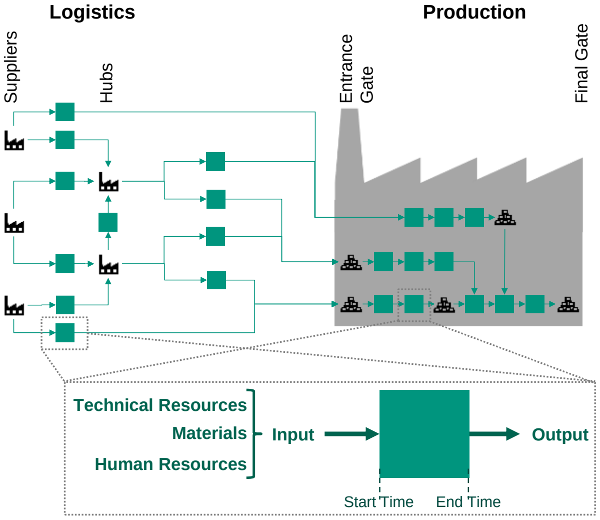


Figure 1 - Considered SCN

disturbances are considered as unexpectedly and unintentionally arising events whose impacts lead to deviations between actual and planned values [10]. Similarly, [11] defines disturbances as unexpected events that lead to interruptions or delays in the completion of a certain task, causing significant deviations between actual and planned values. To allow for a common understanding, disturbances in this work represent unexpectedly occurring events which, by affecting the availability of at least one of the three input factors, lead to disruptions of the normal process flow.

Such disruptions require appropriate measures as a reaction to the disruption. In literature, this is referred to as reactive disruption management. In contrast to such reactive strategies, prevention strategies focus on the proactive avoidance of disturbances [12]. However, they are not considered within the scope of this paper but can be understood as an extension of the approach presented below.

Similar to the definitions of disturbances, literature on reactive disruption management is manifold. [12], e.g., present a concept for an event-based early warning system which allows for a smooth production by identifying critical developments as early as possible and hence extending the time to react. Using a discrete-event simulation, the authors show that the impacts of a disruption can be reduced by a timely identification of deviations and the derivation of suitable measures [12]. They show that an early detection of disruptions improves the adherence to delivery dates and reduces the number of delayed products [12]. In [13], the effects of disruptions on production processes are simulatively assessed in a decentrally controlled job shop production. In the event of a disruption, a disruption reaction rule is used to examine which order can be carried out despite the disruption [13]. Nonexecutable jobs are either stopped or not even started [13]. [7] develop a concept for an integrated disruption management in transportation networks especially considering variable disruption durations. The authors focus on the determination and evaluation of a certain measure as well as on the adjustment of the affected plan [7]. When selecting a measure, the overall objective lies in ensuring a high planning stability by minimizing the number of reschedulings compared to the original plan [7]. A simulation-based case study showed that the evaluation of all suitable measures is reasonable to allow for an objective selection of countermeasures [7]. [3] propose an approach to evaluate predefined measures using six criteria that sort different measures according to their precision of fit and their usefulness. Besides duration, benefit and costs of a measure, the number of available human resources being able to perform the measure, the standardization level of the measure and the organizational escalation level (implementation level of a measure, e.g. machine, plant, SCN) are taken into account. Weighting factors then support the feasibility evaluation of a certain measure [3].

Similar to the approaches presented above, this paper proposes a framework for a reactive DMP. However, while previous approaches only focused on disruption management in either production (e.g. machine breakdown, wrong assembly etc.) or logistics (breakdown of transportation vehicle, traffic jam etc.), this paper combines both perspectives and integrates not only production, but also logistics disruptions and measures in the DMP. Especially in times of increasing interdependencies within SCNs, such an integrated perspective becomes inevitable as it widens the scope for action. The CCS thereby serves as prerequisite for a central and holistic decision-making and for the communication to the decentralized production and logistics units. Besides, it allows for a timely recognition of disruptions, a fact that has also been identified as decisive factor in [12]. Furthermore, this work is, to the best of the authors' knowledge, the first one that *first of all* matches currently *applicable* measures to an occurring disruption (applicability check) before the actual evaluation of the applicable measures (benefit analysis) is carried out. Therefore, the criteria proposed in [7] and [3] are adopted and converted to application and benefit criteria (see Chapter 3.2.4).

3 REACTIVE DMP

In this chapter, the complete procedure of the novel DMP process will be traversed and exemplified for an aircraft manufacturer. However, before actually explaining the novel DMP in detail, the following sub-chapter describes how the CCS generally plans and controls the SCN and how it reacts to a disruption.

3.1 General operation of the CCS

Based on inputs by the management and higher planning levels (e.g. production program), the CCS generates a highly detailed plan for a certain planning period (e.g. ca. seven months in the aviation industry), which specifically outlines respective process sequences as well as the required inputs, planned start and end times and the expected output for each process. Hence, changes in process characteristics (inputs, planned times, output) can only be realized by updating the plan. In contrast to former systems (Figure 2A), where – due to high complexities and the lack of ubiquitous data availability – production and logistics planning has been conducted independently without supporting coordinated decision making ([9]), the CCS pursues an integrated planning of both production and logistics. Unlike plan generation, plan execution is realized by decentralized logistics and production control systems that operate the process units according to the plan (Figure 2B). As stated by [12], an instant reporting of detected disruptions to the CCS is vital to allow for an immediate reaction. When receiving a disruption report (e.g. from a transportation unit), the CCS starts the DMP (Figure 2C): As implied before, the impact of the disruption is first assessed. If the disruption is impermissible, i.e. if it is expected to lead to a delay at the final gate, the algorithm searches all measures of a company for currently applicable measures being suitable for the reported disruption. This is based on the assumption that each company possesses a catalogue of measures which (in whatever manner) contains all actions focussing on the reduction or elimination of disruption impacts. Whether this catalogue is complete or not is not questioned here since this work focuses on selecting suitable measures and not on identifying them. Logistics measures, in such a catalogue, could e.g. be an additional transport, a change of cargo, or a change of the transport time. For production, a change in the production sequence, a reallocation of resources, or the use of extra capacities could be imaginable.

If the search for applicable measures is completed and if several measures are applicable, a multi-criteria benefit analysis selects the best measure out of the applicable ones. This measure is then implemented by updating the plan.

3.2 DMP

Based upon the first brief introduction of the DMP in the previous sub-chapter, the single steps will be explained in more detail in the following.

3.2.1 Reporting of disruptions

Immediately after its detection, a disruption is reported to the CCS in a standardized way. Here, three characteristics are defined that support a structured characterization: Disruptions are assumed to only affect a single process unit in the first place (e.g. station where the vertical stabilizer (VS) should be assembled). This process unit is responsible for reporting the disruption to the CCS and can therefore be regarded as its **local origin**. Naturally, the disruption of one single process unit can lead to disruptions of subsequent processes (stations after assembly of the VS). All of the disruptions, however, have the same **causal origin** lying in at least one of the input groups. Hence, the reason for a delay in the assembly of the VS, e.g., may either be a defective VS (material), a worker's mistake (human resource), or the failure of an important tool such as a mounting aid (technical resource). The causal origin can only be determined at the local origin. In addition to the local and the causal origin, disruptions can also be characterized by the **predicted process delay (PPD)** they cause. The PPD is thereby defined as the span between the start time and the predicted end time of the disruption. The latter is mostly an estimate based on a worker's experience, historical data (e.g. MTTR), or other calculations (e.g. navigation systems). In the example it equals the estimated addtional unplanned time for the assembly of the VS.

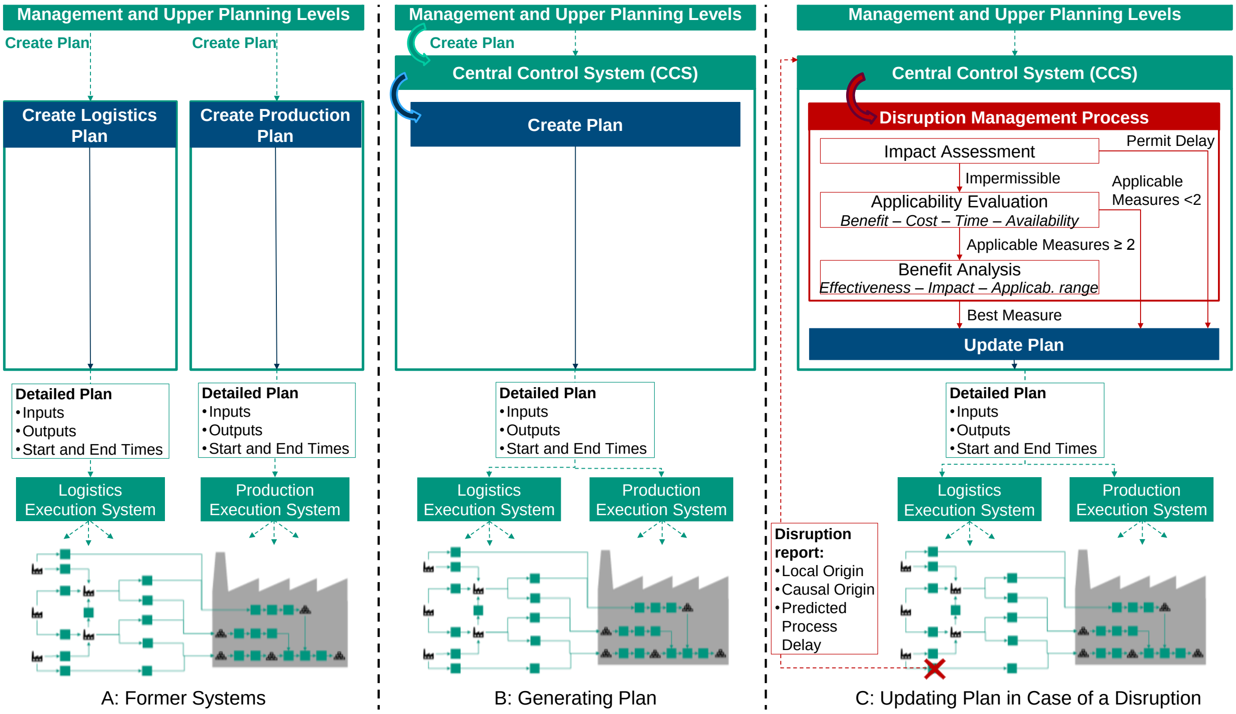


Figure 2 - General operation of the CCS.

3.2.2 *Impact Assessment*

Due to the instant reporting of disruptions and the existence of real-time data (which has also been considered a valid assumption in [8]), the CCS is assumed to always know the system's current state, e.g. the current process flexibilities (CPF) and the actual and planned values of all processes. The CPF is thereby defined as the maximum time a unit can be disrupted without causing disruptions of subsequent processes and is highlighted in green in Figure 3.

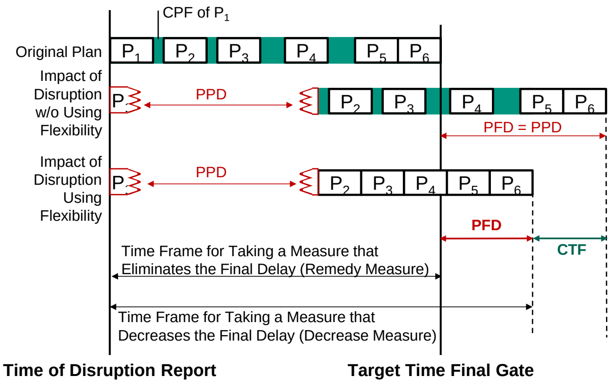


Figure 3 - Relationship between CPF and PFD.

Hence, if there is a material buffer between process P_n and P_{n+1} containing x units, the CPF of P_n is the product of x and the process time of P_{n+1} for x:

$$CPF = (number\ of\ units\ between\ P_n\ and\ P_{n+1}) \cdot (process\ time\ of\ P_{n+1}\ for\ a\ unit) \quad (1)$$

However, if e.g. no VS are buffered, the CPF of the VS assembly station equals 0. The sum of all CPFs from the disturbed process to the final gate is the current total flexibility (CTF):

$$CTF = \sum_{P_{disrupted}}^{P_{final}} CPF \quad (2)$$

In Figure 3, it refers to the CPFs of processes P_1-P_6. To evaluate the impact of a disruption, the predicted final delay (PFD) is calculated by the CCS as

$$PFD = PPD - CTF \quad (3)$$

In case of a positive PFD, the disruption is classified as impermissible and a measure from the catalogue needs to be taken for preventing or at least reducing the delay at the final gate. However, as buffer levels might change over time, the time span that can elapse before a disruption becomes impermissible is dynamic and depends on CPFs and the local origin of the disruption (the closer the process is to the final gate, the less CPFs can be added for the CTF).

3.2.3 *Scope of action*

If a disruption has been categorized as impermissible during the impact assessment, currently applicable measures (also being referred to as scope for action) have to be identified. Based on existing contributions in literature, especially four of the six criteria applied in [3] could thereby be identified as relevant: benefit [14], cost [15], availability of resources [7], and time [14]. The properties of these application criteria thereby depend on the disruption characteristics as well as on the time of the disruption report. While the previously mentioned four criteria can, as subsequently explained, be used to exclude inapplicable measures, the two remaining criteria in [3] (standardization level and organizational escalation level of a measure) shall be considered in the evaluation phase. For being applicable, a measure should first of all be **beneficial**, i.e. it should help to eliminate the impact of a disruption. Whether a measure can be classified as beneficial thereby specifically depends on the causal and local origin. The measure “extra transport of missing aircraft material from supplier to aircraft assembly”, e.g., won’t help to eliminate a disruption that is not caused by a missing VS but by a lack of human or technical resources in the assembly. Second, even if a measure is generally beneficial, there might not be enough **time** for implementing it. Therefore, all measures that require more time for their realization than the timeframe for implementing a decrease measure allows need to be excluded to ensure that the measure will at least reduce the final delay. Here, this timeframe is exemplarily defined as the time span that elapses from the start of the disruption until the last process affected by the disruption can finally start in case no measure is taken (Figure 3). However, this definition has to be elaborated in more detail in future work and may have to be revised. Third, realizing a measure always requires human and/or technical resources. Hence, these resources need to be **available** or at least available early enough for enabling a realization of the measure within the timeframe for implementing a decrease measure. Fourth, apart from being available, resources also have to be **feasible**. In this case, feasibility refers to the fact that the sum of the costs that might occur through the application of the measure and the costs for the remaining predicted delay are not allowed to exceed the costs for the predicted delay in case no measure is taken. Figure 4 condenses these findings. As shown, a measure is only applicable if all four criteria are fulfilled at once.

If no measure is found to be applicable, the only option is to permit the disruption and endure the delay. For only one applicable measure, the next step (benefit analysis) can be skipped and the measure is implemented. If several measures are applicable, the best one has to be determined by means of the subsequently described benefit analysis.

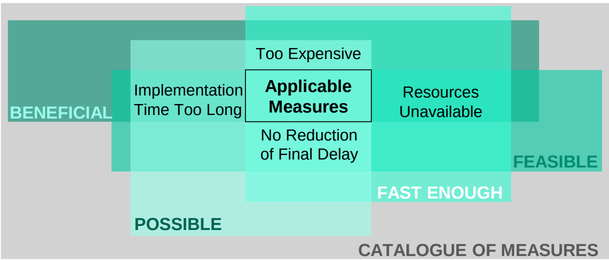


Figure 4 - Scope of action.

3.2.4 Benefit Analysis

Building upon [3] and [7], five criteria are considered meaningful to find the best applicable measure in this paper: remaining delay after application, relative costs of and number of processes affected by the measure, applicability range and standardization level of the measure. In contrast to application criteria (Chapter 3.2.3), benefit criteria do not take binary manifestations (fulfilled/ not fulfilled), but rather cover multiple values, allowing for a benefit analysis. While cost and benefit of a measure have already been considered in a binary way during the applicability check, a more detailed check is necessary for the evaluation of the measure’s benefit. The remaining delay after the application of a certain measure allows to draw conclusions about the **effectiveness** of the measure. As being beneficial is a prerequisite for being applicable, all applicable measures will reduce the delay caused by a disruption. However, only some measures will account for remedy measures completely eliminating the delay before affecting the final process (Figure 3). (For the aircraft manufacturer with a delayed VS assembly, overtime could, e.g., be sufficient to make up for the delay). They can be regarded as 100% beneficial and can be implemented within the remedy timeframe. All other measures are less beneficial so that they only decrease the delay at the final gate (e.g. moving the assembly of the VS to another downstream station can only reduce the PFD) (Figure 3). In a similar way, the costs of a measure can be examined. Each applicable measure is feasible, but the costs of the measures inside the scope of action will be different. Therefore, the **relative costs** being defined as the ratio of cost of the measure and the cost of the predicted final delay in case no measure is taken form the second criterion. As keeping extent and complexity of replanning small is vital for accepting and implementing a measure, both [3] (organizational escalation level) and [7] (planning stability) define a criterion to measure this impact. Hence, following the approach in [7], the number of processes affected by a measure is also used to evaluate its **impact** in this paper. Affected thereby refers to changes in inputs, planned times, or the output. If the assembly of the VS is moved to another downstream station, e.g., this affects the downstream station and all stations in between. Contrary to existing approaches, the matching of measures and a disruption here directly depends on the characteristics of the current disruption (e.g. the longer the PFD, the more expensive the disruption will become and the more measures will hence considered feasible). Since the end time of a disruption is dynamic and might deviate from the prediction, one also needs to consider the **applicability range** of a measure. It refers to the span within which the disruption duration may deviate without causing the measure to lose one of its four applicability criteria (Chapter 3.2.3). The measure of shifting the VS assembly, e.g., might lose its feasibility if a mistake in assembly can be rectified faster than expected. Finally, not only the disruption characteristics but also the measure itself is prone to uncertainties. The implementation time might take longer than expected, information about the availability of resources might be wrong or costs might increase. Based on [3], the **standardization level** of the measure is therefore integrated into the benefit analysis and it is assumed that the more often one kind of measure is applied, the more experience the company possesses for its implementation and the more standardized it is. A higher level of standardization in turn is assumed to ensure a high level of reliability in terms of applicability and evaluation of the measure. Table 1 outlines an exemplary benefit analysis for these criteria. As it is shown, the benefit of a measure can, in an initial and simplified step, be calculated as the sum of the products of the respective weighting factors *wf* and the achieved scores *s*. More complex methods might, however, be applied prospectively.

3.2.5 Implementing a measure - update the plan

As a last step of the DMP, the chosen measure is implemented and the plan updated (inputs, planned times, and/or outputs). Here, the DMP is assumed to be so fast that freezing the actual state of the SCN during run time is possible, meaning that no other disruption can occur before the process is finished and the plan is updated. By adapting the plan, a disruption becomes part of the plan and is therefore no longer a disruption in the proper sense. Moreover, even if there are no applicable measures for an impermissible disruption, the plan is adapted by adjusting the planned times so that there is no delay any more. By doing so, the current plan indirectly contains all disruptions and the measures that are or will be implemented. This way, they can be taken into account when a new disruption occurs.

Updating the plan for each impermissible disruption is the basis for ensuring a measure’s success and handling status updates of disruptions. If a measure causes problems during implementation or does not have the expected benefit, a disruption report will be created which starts the DPM once again. Similarly, status updates changing the predicted time of the original disruption can be handled like a completely new disruption. Thus, even positive disruptions (being too early) are possible as it might happen that a measure for speeding up processes is already taken when the local origin reports a shorter PFD than originally expected. To prevent the CCS from starting the DMP too often, it is important to define intervals for status updates.

4 CONCLUSION AND FUTURE WORK

Arisen from an increased need for action within companies for a systematic evaluation and selection of appropriate measures, this paper has proposed a framework for a structured DMP performed by a CCS which combines previously decentralized production and logistics systems. Besides a scheme that allows for a structured identification of disruptions by means of three characteristics (predicted process delay, local and causal origin), the DMP also offers a filtering process selecting suitable measures based on their applicability, effectiveness, relative costs, impact, standardization level, and applicability range.

Moreover, although the logic for selecting a measure has been provided within the scope of this paper, the concrete matching between a certain disruption and the appropriate measure(s) will be presented (and applied in the aviation industry) in future work. The classification of disruptions according to their local and causal origin and the PPD thereby form an essential and promising basis for this step.

Besides, the existence of catalogues of measures has been assumed. Therefore, future work, will not only take care of the development of such an overview but also of how to handle the selection of combinations of measures. Up to now, each action that changes either input, planned times, or outputs is regarded as a measure and combinations of measures can hence be regarded as measures again. However, this might lead to an infinite number of (combinations of) measures in the catalogue, making it impossible for the CCS to select a measure in an appropriate time. A simple listing of measures in a catalogue like it was assumed to be present in all companies will then not be sufficient. Therefore, a classification framework for measures needs to be developed that simplifies the applicability evaluation and the benefit analysis. The logic behind the simplification might, e.g., use the fact that a combination of measures is only applicable if each of the single measures is applicable. Hence, as all of the single applicable measures are beneficial and reduce the delay, the effectiveness of a combination of applicable measures cannot be worse than the effectiveness of a single one.

5 ACKNOWLEDGEMENT

We extend our sincere thanks to the German Federal Ministry of Education and Research (BMBF) for supporting this research project 01IS17094 “Research on modelling and software technology for the flexible and integrated production and logistics planning in dynamic networks”.

Criteria		***Score s*** ***Very good (s = 3)***	***Good (s =2)***	***Poor (s =1)***	***Very poor (s =0)***	***Weighting factor wf***
Effectiveness s_{eff}	Reduction of final delay	100%	≥60%	≥30%	<30%	wf_{eff}
Relative costs s_{rc}	Cost of measure/cost of PFD	<10%	≥30%	≥70%	≥90%	wf_{rc}
Impact s_{imp}	Number of processes affected	only disrupted process	<50%	≥50%	all processes	wf_{imp} } Σ=1
Applicability range s_{vr}	Size of applicability range	≥6σ	≥4σ	≥2σ	<2σ	wf_{vr}
Standardization level s_{sl}	Experience level	standard measure	well-known measure	seldomly applied	never applied before	w_s

$$benefit\ of\ measure = s_{eff} * wf_{eff} + s_{rc} * wf_{rc} + s_{imp} * wf_{imp} + s_{vr} * wf_{vr} + s_{sl} * wf_{sl}$$

Table 1 - Exemplary benefit analysis (σ represents the variance of the PPD).

6 REFERENCES

[1] Bendul, J., Brüning, M., 2017, Collaborative Supply Chain Risk Management: New measures for managing severe supply chain disruptions.

[2] Schenk, M., Stich, V., 2014, Managing Supply Chain Disturbances – Review and Synthesis of Existing Contributions, in Grabot B, Vallespir B, Abdelaziz Bouras SG, Kiritsis D, (Eds.). Advances in Production Management Systems. Innovative and Knowledge-Based Production Management in a Global-Local World, Springer, pp 262–269.

[3] Stich, V., Schroeter, M., Jordan, F., Wenger, L., Blum, M., 2014, Assessment of Counter-Measures for Disturbance Management in Manufacturing Environments, in Grabot B, Vallespir B, Abdelaziz Bouras SG, Kiritsis D, (Eds.). Advances in Production Management Systems. Innovative and Knowledge-Based Production Management in a Global-Local World, Springer, pp 449–456.

[4] riskmethods GmbH, Bundesverband Materialwirtschaft, Einkauf und Logistik e.V. (BME), (Eds.), 2017, Supply Chain Risk Management: Herausforderungen und Status Quo, München.

[5] Chen, Z.-L., 2010, Integrated Production and Outbound Distribution Scheduling: Review and Extensions, Operations Research, 58/1:130–148.

[6] Rinza, T., Schenk, A., Mandelartz, J., Dombrowski, U., Schürer, S., da Silva Matos, I., Schulze, S., 2010, Standortvorteil Deutschland: Erfolgsfaktor integrierte Produktion & Logistik.

[7] Dreßler, D., Beißert, U., 2013, Entwicklung eines integrierten Ansatzes für das Störungsmanagement in komplexen Transportnetzen unter Berücksichtigung variabler Störungsdauern, in Dangelmaier W, Laroque, Christoph, Klaas, Alexander Klaas, (Eds.). Simulation in Produktion und Logistik: Entscheidungsunterstützung von der Planung bis zur Steuerung, W.V. Westfalia Druck GmbH, pp 157–166.

[8] Schuh, G., Stich, V., Hocken, C., Schenk, M., 2015, Design of an Integrated Model for the Real-Time Disturbance Management in Transportation Supply Networks, in Umeda S, (Ed.). Advances in production management systems: Innovative production management towards sustainable growth; IFIP WG 5.7 international conference, APMS 2015, Tokyo, Japan, September 7-9, 2015; proceedings, Springer, pp 144–151.

[9] Váncza, J., 2016, Production Networks, in Laperrière L, Reinhart G., (Eds.). CIRP Encyclopedia of Production Engineering, Springer.

[10] Patig, S., Thorhauer, S., 2002, Ein Planungsansatz zum Umgang mit Störungen bei der Produktion: Die flexible Produktionsfeinplanung mithilfe von Planungsschritten, Wirtschaftsinformatik, 44/4:355–366.

[11] Verband für Arbeitsstudien und Betriebsorganisation e.V. (REFA):, 1991, Methodenlehre der Betriebsorganisation – Planung und Steuerung - Teil 3, 1st Edition, Hanser.

[12] Genc, E., Duffie, N., Reinhart, G., 2014, Event-based Supply Chain Early Warning System for an Adaptive Production Control, Procedia CIRP, 19:39–44.

[13] Vojdani, N., Erichsen, B.E., 2016, Untersuchung der Auswirkungen von Störungen auf produktionslogistische Prozesse, Logistics Journal: Proceedings, 10:1–6.

[14] Fischer, J.-H., Pfeiffer, D., Hellingrath, B., Scavarda, L.F., Martins, R.A., 2014, Robust Parameter Setting of Supply Chain Flexibility Measures Using Distributed Evolutionary Computing, Procedia CIRP, 19:75–80.

[15] Lloret, J., Garcia-Sabater, J.P., Marin-Garcia, J.A., 2009, Cooperative Supply Chain Re-scheduling: The Case of an Engine Supply Chain, in Luo Y, (Ed.). International Conference on Cooperative Design, Visualization and Engineering, Springer, 376–383, 2009.

7 BIOGRAPHY

Sina Helming holds a MSc. degree in Industrial Engineering from Karlsruhe Institute of Technology (KIT). Currently, she is a Research Associate at the Institute of Production Science (wbk) at KIT focusing on Global Production Strategies.

Clio Haverkamp holds a BSc. degree in Industrial Engineering from Karlsruhe Institute of Technology (KIT). She is currently working on her master thesis at the Institute of Production Science (wbk) at KIT.

Dr.-Ing. Benjamin Häfner holds a MSc. degree in Mechanical Engineering from Georgia Institute of Technology and a Dipl.-Wi.-Ing. degree in Business Engineering at Karlsruhe Institute of Technology (KIT). Currently, he is Chief Engineer at the Institute of Production Science (wbk) at KIT.

Prof. Dr.-Ing. Gisela Lanza holds the Professorship "Production System and Quality Management" at the Karlsruhe Institute of Technology (KIT) and is Head of the Institute of Production Science (wbk) at KIT.

Development of a Method for Communication-Oriented Supplier Evaluation

D. Herberger[1], P. Nyhuis[1]
[1] Institute of Production Systems and Logistics, Leibniz University Hanover, Germany

Abstract
In times of globalisation, ever-shorter product life cycles and highly volatile markets, companies are increasingly joining forces to form production networks. As a result, the performance of the entire network becomes an essential competitive factor that needs to be constantly improved. It is not uncommon for a company to form a network with several thousand suppliers. As long-term partnerships are usually sought, the selection of suppliers already has a direct impact on the success of the cooperation. Nevertheless, the relationship with existing suppliers must also be constantly developed in order to achieve continuous improvement and sustainable success. The selection and the development of Suppliers are preceded by the supplier evaluation, which forms the basis for the measures and decisions taken. The evaluation, selection and development of suppliers is carried out by supplier management as part of the supply chain management. There are numerous methods and approaches for the evaluation of suppliers, which concentrate on different criteria, often related to direct cost and performance data. However, what the existing approaches have in common is that communication receives insufficient attention. The communication, as essential interfaces between the cooperation partners, has a substantial influence on the costs and the performance and consequently on the efficiency of the entire network. Consequently, a model for the communication-oriented evaluation of supplier relationships is needed in order to continuously reduce the risk of cost-increasing inefficiencies. This paper provides an overview of the existing research gap and presents the procedure for developing a method for communication-oriented supplier evaluation.

Keywords
Supplier management, supplier evaluation, production network, communication

1 INTRODUCTION

Increasing demands on industrial production, such as increasing competitive pressure and shorter product and technology life cycles, are more and more frequently met by companies by merging into production networks. Their steadily growing importance is motivated by access to resources and new markets, concentration on core competencies and an increase in production capacities. Due to the resulting relevance of the network partners, the efficiency of the supplier structure is increasingly becoming an essential competitive factor. In order to ensure quality standards and logistics performance as well as compliance with cost specifications, long-term partnerships are preferred. [1]

As a result, a far-reaching decision is made as early as the supplier selection stage, which has a significant influence on the subsequent performance of the cooperation. The evaluation, selection and development of the suppliers of a company are summarised by the so-called supplier management. This in turn is part of supply chain management, which considers the supply chain between suppliers, manufacturers and customers as well as their respective production, warehouse, logistics and sales units. The focus is on planning and improving the flow of money, materials and information (refer to **Figure 1**). [2]

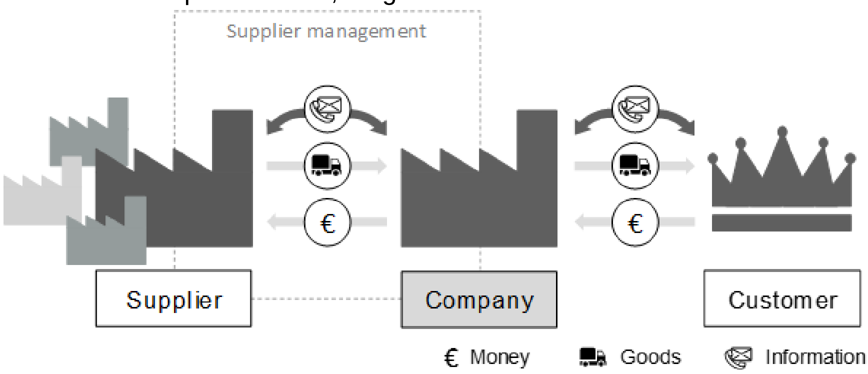


Figure 1 - Schematic representation of a supply chain including cash, material and information flows

The cooperation in the supply chain as well as the efficiency of the entire network is significantly influenced by the communication between the actors [3–6]. The number of suppliers of large companies is often three to four digits, which entails a considerable coordinative effort. These suppliers are usually looked after by a fraction of the buyers. It is not uncommon for a single buyer to be responsible for 100 or more suppliers. If the essential communication processes show inefficiencies, this can result in far-reaching consequences in terms of costs and performance. Small and medium-sized enterprises (SME) are less affected by this problem due to the lower number of suppliers. A more serious challenge for SME results from the usually small purchase quantities. Missing economies of scale lead among other things to a relatively high portion of logistics and communication costs. If, for example, an on-site coordination with a supplier is necessary, the resulting costs can quickly equalise the targeted savings potential. [7, 8]
Due to the economic potential and the competitive necessity, it is therefore necessary to optimise processes along the entire supply chain instead of only within one factory. To this end, the actors of a network must be coordinated in order to enable a goal-oriented production of goods and services through joint and efficient action. [9–11] In order to develop the full potential, the evaluation and selection of new suppliers must be considered under this focus in addition to the adaptation of existing communication relationships. In current approaches and procedures for supplier evaluation, however, communication is usually only insufficiently included. [10]

2 NEED FOR RESEARCH

The supplier evaluation, as part of the supplier management, can be subdivided for example into the evaluation of new suppliers, as basis for the decision-making during the selection, as well as into the evaluation of existing suppliers, as basis for the supplier development. [12–14]. In both cases, the objective of the supplier evaluation is to determine a supplier's performance capability in order to identify risks and opportunities at an early stage and avoid cost-effective performance deficits. [15]. The results of the supplier evaluation can then be used, for example, to preselect new suppliers or to identify the development potential of existing supplier relationships (refer to **Figure 2**).

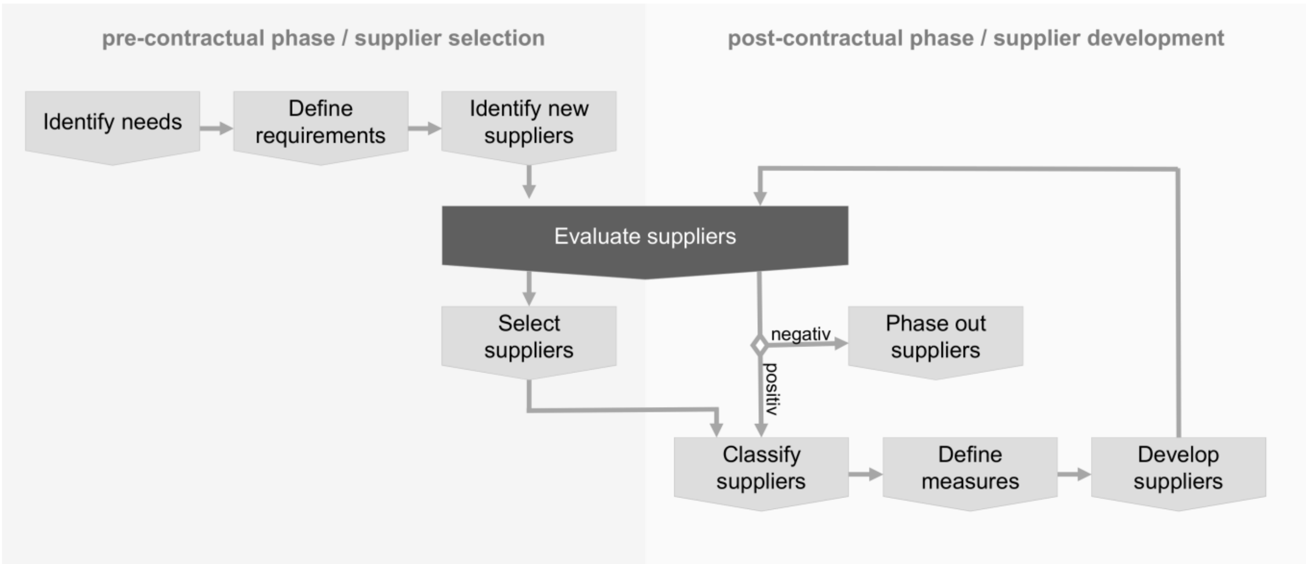


Figure 2 - Schematic representation of supplier selection and supplier development process

Existing approaches focus on criteria such as cost structures, logistics performance, product quality, market position, availability, production know-how, flexibility or innovation potential of suppliers. [8, 15, 16]
Communication is also a factor that is repeatedly mentioned as a relevant criterion, but often receives insufficient attention [17–19]. The relevant components of the criterion for supplier evaluation and their effects on the costs and performance of the supplier relationship have so far rarely been the subject of scientific research. The often very broadly defined term 'communication' refers, for example, to the supplier's reaction time, compliance with promises, conduct in negotiations or openness in discussions [19, 20]. The basic elements of communication (refer to **Figure *3***), which are necessary for the execution of the actual communication process and can significantly influence the costs and performance of communication processes, have not been taken into account in supplier evaluation so far [17].

In addition to the sender and one or more receivers, these elements consist of the communication content, the communication medium, the requirements for the communication process and the communication barriers. The communication content represents the core of the communication process. It describes the information that is transmitted in the form of a message. The communication medium is a technical or non-technical channel which is necessary for the transmission of the message. The framework conditions that must be observed during execution are described by the requirements for the communication process. The communication barriers are disturbance variables that impair the communication process and consequently lead to inefficiencies. [21] **Figure *3*** shows the elements of communication in a schematic context.

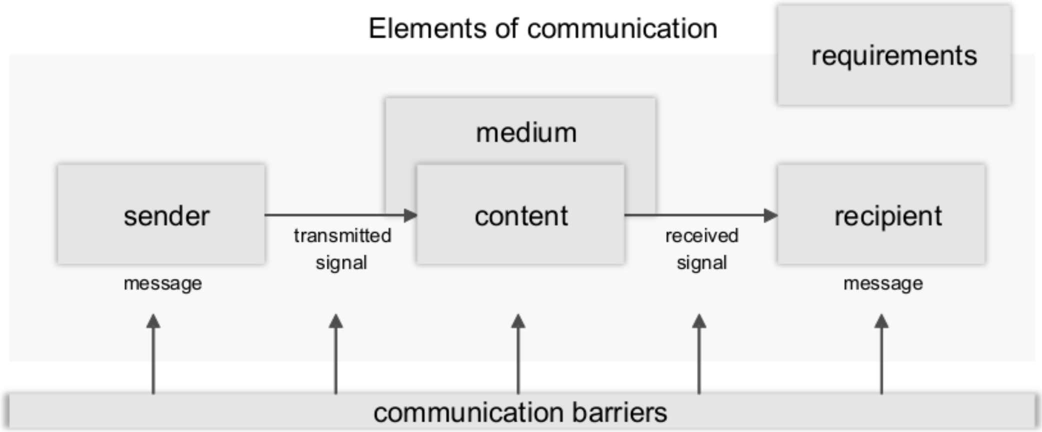


Figure 3 - Simplified representation of the elements of a general communication process (according to [22])

In particular, the communication barriers must be taken into account when evaluating suppliers, as they have a significant influence on the efficiency of a process and can therefore directly influence costs and performance. The hypothesis that significant savings potentials can be achieved by reducing barriers in communication processes with suppliers can among others be derived from the survey of an employee of a large medium-sized company. The surveyed expert works in the operative purchasing department and is responsible for approximately 85 suppliers. In addition to him, 14 other employees work in the same department. The hypothesis could be derived from the assessment that the expenditure for a communication process with a supplier could be reduced by an average of approx. 15-25%[1] if the communication barriers that occur were reduced and the processes meaningfully standardised. There are currently no standards for numerous communication processes in the company described. The processes are hindered by numerous and different communication barriers, but these have not yet been identified holistically nor have employees been sensitised to them.

Assuming that each of the 15 employees in the purchasing department works approximately 4.5 hours[1] of his daily working time with communication processes with suppliers and that the effort could be reduced by an average of 20%, this would result in a savings potential of 13.5 hours per day and almost 3,000 hours per year for all employees. However, a continuous improvement of the processes by a previous analysis and evaluation of the communication relations does not take place. This is due to the fact that the employees were not aware of the extent of the potentials and that there is no standardised procedure for evaluating communication relationships.

The hypothesis derived from industrial practice is also supported by numerous scientific contributions. PRAHINSKI AND BENTON, MILLER, PILNY AND ATOUBA as well as LARGE, among others, argue that the improvement of communication processes can lead to a significant increase of efficiency, but that there is currently a lack of methods to analyse and evaluate communication processes satisfactorily. [8, 23–25]

In order to close this research gap, a model is needed that enables existing and expected communication relationships with suppliers to be analysed and thus the supplier evaluation to be specified. This extension of the existing criteria and methods is intended to create a decision basis for supplier selection and development that is as goal-oriented as possible, thereby further reducing the risk of cost-effective inefficiencies. If existing deficits are not identified, the success and efficiency of the partnership suffer, which can be reflected, for example, in poor process and product quality, inadequate performance and high costs - and ultimately weakens the entire network [17, 26].

3 APPROACH TO DEVELOP A COMMUNICATION-ORIENTED SUPPLIER EVALUATION METHOD

For the development of a method for communication-oriented supplier evaluation, the first step was to identify and describe the components of the aforementioned elements of communication in detail.

[1] The values stated are estimates of the employees, which have not been audited. There is no claim to correctness. The values only serve to sensitise to potential savings.

In total 12 communication media, 14 requirements and 48 barriers relevant for supplier evaluation were identified by means of literature research, expert surveys and evaluations of past contract research projects. An extract of the identified components can be found in **Figure 4**.

Media	Requirements	Barriers
Face-to-face	Documentability	Time zone differences
Letter	Fast transmission of information	Lack of resources
Telephone	Accuracy	Missing standards
Video telephony	Instant feedback	Lack of access to communication media
Email	Time to think about response	Different languages
Fax	Confidentiality	Cultural differences
Test message	Safety	Knowledge protection
Instant messaging	Transmission of extensive communication contents	Missing framework contracts
ERP system	Transmission of nonverbal signals	Lack of possibility to integrate the EDP systems
Chat	Opportunity for teamwork	Missing or excessive media diversity
Cloud server	Transmission of complex information	Great geographical distance
…	Transmission of standard data	Supremacy of the supplier
	Unequivocal authorship	Poor communications infrastructure in the region
	Influencing the reception time	Poor transmission quality
	…	…

Figure 4 - Exemplary excerpt of the identified communication media, requirements and barriers

The lists shown represent a work status which is continuously checked, expanded and sharpened.

For the evaluation of communication processes, the focus must be on the communication barriers, which, as sources of disturbance, directly and indirectly influence the efficiency of communication processes. The relevant barriers for the supplier relationship can have different characters and therefore be classified according to different criteria. For example, they can be subdivided according to their sphere of influence into ‘assessable in supplier selection’ and ‘assessable in supplier development’ or according to their reference point into ‘employee-related’ and ‘process-related’. Employee-related communication barriers are, for example, factors such as lack of specialist knowledge, distrust or uncertainty [27]. Process-related barriers describe factors such as a lack of standard processes, a lack of media diversity, great geographical distances or constantly changing contact persons.

The communication barriers are operationalised in a next step and embedded in a procedure so that individual communication processes can be evaluated. All components of the different elements are then compared in Design Structure Matrices or Domain Mapping Matrices in order to identify possible interactions. The interactions should support the identification of typologies and potentials as well as the efficient design of communication processes after evaluation. Figure 5 schematically shows the extent to which the appearance or characteristics of a communication barrier can be influenced by a communication medium.

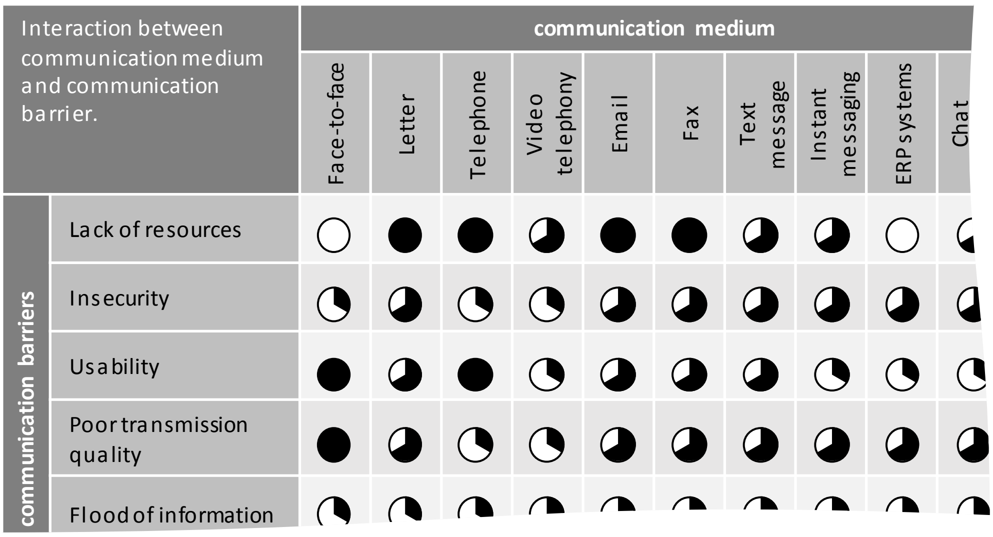


Figure 5 - Schematic representation of the interactions between communication barriers and communication media

Subsequently, a standardised and structured procedure for the communication-oriented supplier evaluation must be formulated in order to obtain a reproducible and comparable result. In doing so, a comprehensible and simple description of the instructions for action must be ensured so that non-experts can also carry out the procedure.

4 OUTLOOK

This article presents the need for research and the procedure for developing a method for communication-oriented supplier evaluation. Currently, the method for extending the existing procedures is being developed at the Institute of Production Systems and Logistics at Leibniz University Hanover.

The current identification of interactions as well as the elaboration of a standardised and structured evaluation procedure represent one of the last conceptual steps before a broad field study by means of case studies.

The case studies are carried out at partner companies and serve to evaluate the identified components of the elements, the evaluation method and the procedure for implementation. By means of questionnaires a broad data collection for the communication-oriented supplier evaluation is accomplished. The evaluation results are then discussed and assessed in workshops with participants from the partner companies. Then, measures to increase efficiency in selected communication processes, supported by the created matrices, are jointly developed and subsequently re-evaluated. These results are also evaluated in joint workshops in order to assess the effectiveness of the measures. To validate the results, several companies have already been found to actively support the project. The results of the case studies will then be used to adapt and further specify the developed method according to the findings. After completion, the method will be made available to companies for independent use.

Funding

This work was supported by the German Research Foundation [grant number NY 4/59-1].

5 REFERENCES

[1] MERTINS, Kai (Hrsg.); KRAUSE, Oliver (Hrsg.); SCHALLOCK, Burkhard (Hrsg.): *Global Production Management : IFIP WG5.7 International Conference on Advances in Production Management Systems September 6-10, 1999, Berlin, Germany*. Boston, MA : Springer, 1999 (IFIP - The International Federation for Information Processing 24)

[2] MENTZER J. T. ; DE WITT W. ; KEEBLER J. S. ; MIN S. ; NIX N. W. ; SMITH C. D. ; ZACHARIA Z. G.: *Defining Supply Chain Management*. In: *Journal of Business Logistics* (2001), Vol. 22, No. 2

[3] ALAVI, M. ; LEIDNER, D.: *Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues*. In: *MIS Quarterly* (2001), Vol. 25, S. 107–136. URL http://www.jstor.org/stable/3250961

[4] BOSE, Ranjit: *Knowledge management metrics*. In: *Industrial Management & Data Systems* (2004), Vol. 104 Issue: 6, S. 457–468. URL https://doi.org/10.1108/02635570410543771 – Überprüfungsdatum 2017-10-27

[5] BALLAND, Pierre-Alexandre ; BOSCHMA, Ron ; FRENKEN, Koen: *Proximity and innovation : From statics to dynamics*. In: *Regional studies* 49 (2015), Nr. 6, S. 907–920

[6] HANDLEY, Sean ; GRAY, John: *Managing quality in a heterogeneous contract manufacturing environment*. In: *Decision sciences : DS* 46 (2015), Nr. 6, S. 1011–1048

[7] LEHNEN, Marc: *Wettbewerbsstrategie und regionale Reichweite*. Wiesbaden : Deutscher Universitätsverlag, 2002

[8] LARGE, Rudolf: *Strategisches Beschaffungsmanagement : Eine praxisorientierte Einführung mit Fallstudien*. 4., vollständig überarbeitete Auflage. Wiesbaden : Gabler Verlag / GWV Fachverlage GmbH Wiesbaden, 2009

[9] THOMAS, Stefan: *Produktionsnetzwerksysteme – Ein Weg zu effizienten Produktionsnetzwerken*. St.Gallen, Universität St.Gallen. Dissertation. 2013

[10] MODI, Sachin B. ; MABERT, Vincent A.: *Supplier development: Improving supplier performance through knowledge transfer*. In: *Journal of Operations Management* (2007), Nr. 25, S. 42–64

[11] INKPEN, Andrew C.: *Learning Through Alliances : General Motors and NUMMI*. In: *California management review* (2005), Vol. 47, No. 4

[12] HELMOLD, Marc ; TERRY, Brian: *Lieferantenmanagement 2030* : *Wertschöpfung und Sicherung der Wettbewerbsfähigkeit in digitalen und globalen Märkten*. Wiesbaden : Springer Gabler, 2016

[13] DISSELKAMP, Markus ; SCHÜLLER, Rudolf: *Lieferantenrating : Instrumente, Kriterien, Checklisten*. Wiesbaden, s.l. : Gabler Verlag, 2004

[14] LARGE, Rudolf: *Interpersonelle Kommunikation und erfolgreiches Lieferantenmanagement : Eine Kausalanalyse des externen Kommunikationsverhaltens von Beschaffungsmanagern*. Wiesbaden :

Deutscher Universitätsverlag, 2003 (nbf Neue Betriebswirtschaftliche Forschung 321)

[15] ARNOLD, Bernhard: *Strategische Lieferantenintegration : Ein Modell zur Entscheidungsunterstützung für die Automobilindustrie und den Maschinenbau*. Gabler Edition Wissenschaft. Wiesbaden, s.l. : Deutscher Universitätsverlag, 2004 (Logistik-Management)

[16] IRLINGER, Wolfgang: *Kausalmodelle zur Lieferantenbewertung*. Zugl.: Passau, Univ., Diss., 2011. 1. Aufl. Wiesbaden : Gabler Verlag / Springer Fachmedien Wiesbaden GmbH Wiesbaden, 2012 (Gabler Research)

[17] LARGE, Rudolf O.: *Interpersonelle Kommunikation und erfolgreiches Lieferantenmanagement. Analyse des externen Kommunikationsverhaltens von Beschaffungsmanagern auf Basis eines Strukturgleichungsmodells*. In: *Zeitschrift für Betriebswirtschaft* 76 (2006), Nr. 10, S. 1005–1034

[18] KIVIMAKI, Mika ; LANSISALMI, Hannakaisa ; ELOVAINIO, Marko ; HEIKKILA, Armo ; LINDSTROM, Kari ; HARISALO, Risto ; SIPILA, Kari ; PUOLIMATKA, Leena: *Communication as a determinant of organizational innovation*. In: *R and D Management* 30 (2000), Nr. 1, S. 33–42

[19] JANKER, Christian G.: *Multivariate Lieferantenbewertung*. Wiesbaden : Springer Fachmedien, 2009 (Gabler Edition Wissenschaft)

[20] HOFBAUER, Günter ; GLAZUNOVA, Anna ; HECHT, Dirk: *Strategische Lieferantenauswahl*. Ingolstadt, 2015 (Arbeitsberichte - Working Papers 36)

[21] HERBERGER, David ; NYHUIS, Peter: *Gestaltung von Kommunikationsprozessen in Produktionsnetzwerken*. In: *ZWF Zeitschrift für wirtschaftlichen Fabrikbetrieb* 113 (2018), Nr. 5, S. 272–276

[22] HERBERGER, D. ; BUSSEMER, F. ; NYHUIS, P. ; NOENNIG, J. ; SCHMIEDGEN, P.: *Communication Interfaces in interorganizational production networks*. In: *Global Conference on Business and Finance Proceedings* (2018), Vol. 13, No. 1

[23] MILLER, Vernon D. ; POOLE, Marshall Scott ; SEIBOLD, David R. ; MYERS, Karen K. ; PARK, Hee Sun ; MONGE, Peter ; FULK, Janet ; FRANK, Lauren B. ; MARGOLIN, Drew B. ; SCHULTZ, Courtney M. ; SHEN, Cuihua ; WEBER, Matthew ; LEE, Seungyoon ; SHUMATE, Michelle: *Advancing Research in Organizational Communication Through Quantitative Methodology*. In: *Management Communication Quarterly* 25 (2011), Nr. 1, S. 4–58

[24] PRAHINSKI, Carol ; BENTON, W.C: *Supplier evaluations : Communication strategies to improve supplier performance*. In: *Journal of Operations Management* 22 (2004), Nr. 1, S. 39–62

[25] PILNY, Andrew ; ATOUBA, Yannick: *Modeling Valued Organizational Communication Networks Using Exponential Random Graph Models*. In: *Management Communication Quarterly* 32 (2018), Nr. 2, S. 250–264

[26] RAUTER, Romana: *Interorganisationaler Wissenstransfer : Zusammenarbeit zwischen Forschungseinrichtungen und KMU*. Zugl.: Graz, Karl-Franzens-Univ., Diss., 2011. Wiesbaden : Springer Gabler, 2013 (Springer Gabler Research)

[27] SCHMIEDGEN, Peter: *Barrieren im interorganisationalen Wissensaustausch : Ordnungsrahmen, Analysemethoden und empirische Studie*. Zugl.: Dresden, Techn. Univ., Diss., 2015. Dresden : TUDpress, 2015

6 BIOGRAPHY

David Herberger (*1990) has been a research associate in the specialist factory planning group at the Institute of Production Systems and Logistics (IFA) at Leibniz Universität Hannover since 2016. He previously studied mechanical engineering (B.SC, M.Sc) at the Karlsruhe Institute of Technology (KIT).

Peter Nyhuis (*1957) has been head of the Institute of Production Systems and Logistics (IFA) at Leibniz Universität Hannover since 2003. Prof. Dr.-Ing. habil. Peter Nyhuis was a member of the German Science Council (2014-2018) and has been Chairman of the Science Council of the AiF (GAG 3) since 2015. He is also a member of the German Academy of Technical Sciences acatech and the Scientific Society for Production Technology (WGP).

Characterization of Supply Chains in the Regeneration of Complex Capital Goods

T. Lucht[1], T. Kämpfer[2], P. Nyhuis[3]
[1,2,3] Institute of Production Systems and Logistics, Leibniz University Hannover, Germany.

Abstract
In a wide array of enterprises, complex capital goods (e.g., aircraft engines) represent the basis for delivering services or producing goods. For this reason, one of the most important goals in the regeneration of complex capital goods, in addition to cost savings, is to attain the highest possible adherence to agreed delivery dates and thus the optimization of the logistics performance. Due to the complexity and interdependent interactions, a large variety of configurations of regeneration supply chains are possible. This makes the design and configuration of the regeneration supply chain a central challenge for regeneration service providers. However, it is not possible to estimate the effects of changes in the configuration of regeneration supply chains and their parameterization on the logistical objectives without extensive effort, yet. To support individually designing and optimizing regeneration processes for complex capital goods an assessment model will be developed at the Institute of Production Systems and Logistics (IFA) in the context of the Collaborative Research Center (CRC) 871. Therefore the effects of different regeneration-specific supply chain configurations are to be made quantitatively describable. In order to enable the general applicability of the approach within different industries and to create a transparent overview of possible options for the characteristic features to be set, a number of regeneration supply chains of complex capital goods are analysed. Based on this analysis an approach for systematic characterization of possible regeneration supply chain configurations is presented.

Keywords
regeneration, supply chain configuration, complex capital goods, characterization model, MRO

1 INTRODUCTION

Complex capital goods like aircraft engines are characterized by a large number of different components with a multitude of interdependencies among them as well as with the operational and environmental conditions [1,2]. Due to their great value, these capital goods can only be amortized over a long service life. Regeneration service providers therefore offer maintenance, repair and overhaul (MRO) measures to regenerate these capital goods at the end of a service phase and thus extend their service life [3]. Besides being cost efficient, regeneration service providers need to meet their customers' demand for a high logistic performance [4]. Customers perceive the logistic performance through the delivery times offered by the regeneration service providers as well as their due date reliability [5]. If regenerated goods are not delivered to the customer punctually, some customers are not able to deliver the services they offer due to a lack of replacement goods [3]. Consequently, they lose sales and profits. On the other hand, the regeneration service provider faces contractual penalties and very likely a decreasing number of follow-up orders. For the regeneration process to be profitable for the service provider, they need to attain a high on time delivery and short throughput times, while keeping the logistic costs to a minimum. The configuration of the regeneration supply chain is defined to an extent by the capital good itself as well as the technologies required to regenerate it. From a production logistics perspective, there are degrees of freedom with regards to supplying capacities, methods for scheduling, processing strategies and control methods. Superior the number and arrangement of process and storage stages is a further important degree of freedom in designing supply chains in regeneration. How these degrees of freedom are configured directly impacts the logistic performance and costs of all the individual processes [7]. In view of this, as part of the third funded phase of Collaborative Research Centre (SFB) 871, "Regeneration of Complex Capital Goods", the Institute of Production Systems and Logistics (IFA) focuses on modelling, planning, controlling and assessing regeneration processes with regards of logistics. The aim is to optimize the logistic performance and to provide support in individually designing and improving regeneration processes executed by regeneration services providers. [7,8,9]

The multitude of process elements and features that regeneration supply chains consist of distinguishes them from conventional production processes. In addition, the regeneration good (e.g., aircraft engines, wind turbines) plays a decisive role in how the different process elements in the supply chain have to be configured. [8] To ensure the universal applicability of the assessment model to different regeneration supply chain configurations of different types of complex capital goods, it is first necessary to identify, analyze and characterize different possible configurations of supply chains in regeneration.

2 IDENTIFICATION AND SYSTEMATIC ANALYSIS OF SUPPLY CHAIN CONFIGU-RATIONS FOR THE REGENERATION OF COMPLEX CAPITAL GOODS

The generic process model by Eickemeyer et al. represents a process model for the regeneration of complex capital goods. Based on the SCOR model, the authors define the processes of planning, diagnosis, disassembly, inspection, repair, assembly and quality assurance as elementary processes of the regeneration. The described elementary processes and the resulting generic process model of regeneration are shown in Figure 1. [10]

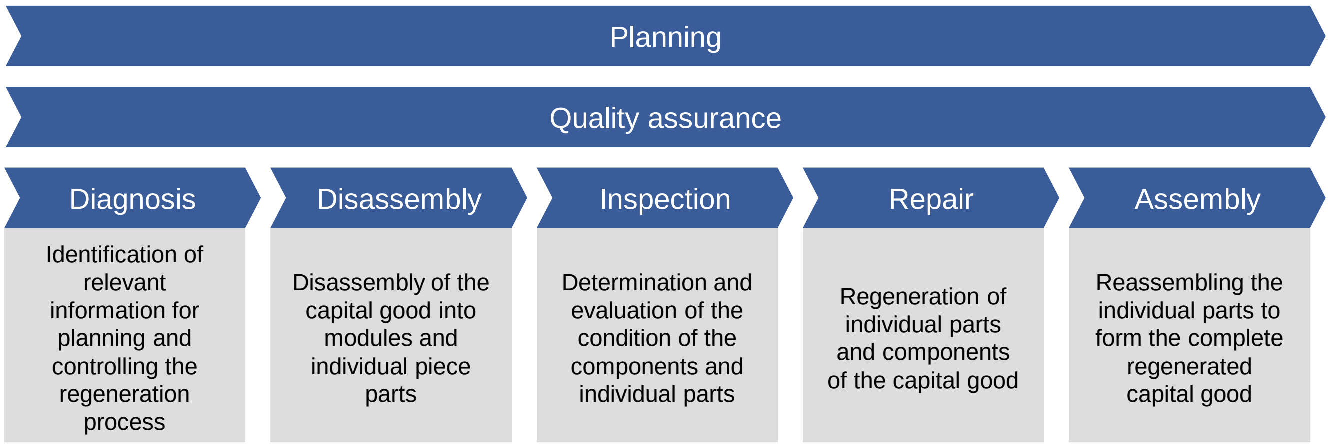


Figure 1 - Generic process model for the regeneration of complex capital goods (on the basis of [10])

It gives an overview of elementary processes that exist in many regeneration supply chains. However, differences in the composition of individual real regeneration supply chains out of the elementary processes as well as their exact arrangement and linkage are not considered by this model.

In order to develop a characterization model to systematically describe configurations of regeneration supply chains, regeneration supply chains established in industrial practice and described in the scientific literature were analysed in regards to identifying different configurations of regeneration supply chains and possible characteristics that shape the structure of the regeneration supply chain. The identified characteristics and characteristic values form the basis of the characterization model described in the following section. Six representative capital goods were selected for the analysis and the underlying regeneration supply chains were documented. This selection of representative capital goods was based on a set of criteria presented below [11, 12]:

Complex capital goods to be considered must have a long service life and at the same time lose functionality over time. At the end of a utilization phase, they must still have a remaining value, which makes a regeneration of the capital asset financially advantageous compared to buying a replacement. Despite an impairment of the functionality, the capital goods still has to offer a high remaining value. In addition, the capital goods must be standardised and the components of the potential regenerated material must be interchangeable. The product technology used must also be stable. Since regeneration service providers with their own location and fixed regeneration areas should be considered within the scope of the systematic analysis, only non-stationary goods that can be sent to the service provider were used for an analysis.

Based on the presented criteria, capital goods from various industries were selected for an analysis of the respective regeneration supply chains, which are aircrafts, aircraft engines, heavy and light rail vehicles, rail vehicle transformers and wind turbines.

3 CHARACTERIZATION MODEL OF SUPPLY CHAHAIN CONFIGURATIONS FOR THE REGENERATION OF COMPLEX CAPITAL GOODS

The identification of characteristics and characteristic values from scientific literature and expert discussions allowed the creation of a characterization model, which is shown in Figure 2. Its morphological structure allows a comprehensive overview of the identified characteristics and the observed characteristic values. Each row of the model represents one identified characteristic of the regeneration supply chains and lists identified characteristic values. These are coloured according to their number of occurrence. The more often the application of the respective characteristic value was identified during the analysis, the darker it is coloured in the characterization model. The configurations of the supply chains for the regeneration of rail vehicle transformers is exemplarily shown in red.

The characteristics taken into account can be divided into those that are determined by the capital goods themselves and those that result from their regeneration process.

3.1 Characteristics of the capital good

The capital good itself is described by the characteristics complexity and value. 'Complexity' in the context of capital goods describes the number of different individual parts in the capital goods, the number of subassemblies and the width and depth of the parts list structure [1]. While the degree of complexity of a capital good decisively determines the complexity of the regeneration processes and thus the process costs, the consideration of the capital good's value is particularly decisive due to the influence on the capital commitment costs. The analysis of the various supply chains revealed different levels of complexity and value. While capital goods such as rail vehicle transformers have a value of less than € 10 million and a low complexity, aircrafts have a value range of more than € 100 million and a very broad and deep parts list structure. Characteristics as the weight or volume of a complex capital good primarily influence handling and storage, but have a rather small influence on the basic configuration of the regeneration supply chain. Therefore they are not considered here.

3.2 Characteristics of the regeneration process

The selected primary target of regeneration planning has a decisive impact on the configuration of the regeneration supply chain. The targets 'due date reliability' and 'delivery time' are particularly noteworthy as shown in chapter 1. For the operator of the capital goods, the delivery time represents the time until the goods are ready to be put back into service. The due date reliability describes whether an order was completed within an agreed delivery date tolerance. [11] It has a considerable influence on the deployment planning of the capital goods. [13] One instrument in product regeneration to enable service providers to meet both lead time and due date reliability requirements is pooling. Pooling represents a process for compensating long lead times of repair measures [13]. Components that undergo repair and cannot be completed until their planned due date for reassembly can be replaced by identical components from a material pool. The repaired component, which was originally removed from the capital good, is later added to the material pool in place of the assembled component taken from the pool. [14]

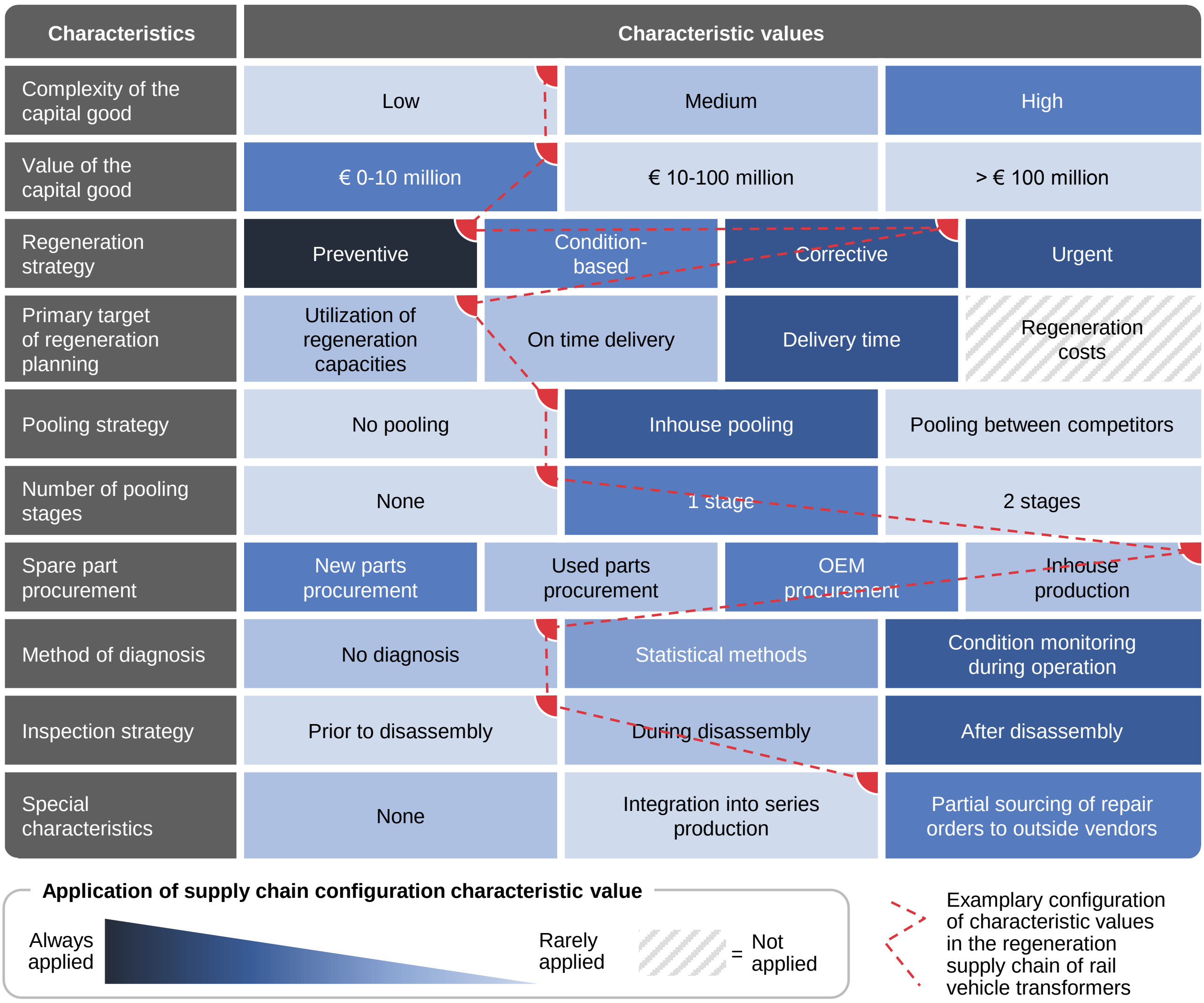


Figure 2 - Characterization of supply chain configurations for the regeneration of complex capital goods

Figure 3 exemplarily shows the use and positioning of stock and pooling stages within the analysed supply chains for regeneration.

While some regeneration service providers do not have a pool stock (e.g. service providers in the regeneration of rail vehicle transformers), others hold a second pool stage to compensate long lead times of repair operations (e.g. service providers in the regeneration of aircraft engines). While the second pool stage only stores material that is serviceable, because it has already been regenerated or it has been newly procured, the first pool stage stores material that still has to be brought into a serviceable state by applying repair measures. This unserviceable material comes from capital goods to be regenerated or was procured on a used parts market. Service providers in the regeneration of light rail vehicles that were analysed within this project did establish one pool stage. Their pool stock mainly includes newly procured electronic components and wearing parts. Pooling can also be used if components are found to be scrap, i.e. can no longer be repaired. If these components are not available in the pool stocks, spare parts must be procured. Regeneration service providers fall back on different spare part procurement strategies.

the aircraft engine, this number of usage units is measured in cycles, i.e. the number of take-offs and landings [11]. A special type of the preventive regeneration strategy is condition-based regeneration, which assesses the necessity of regeneration measures on the basis of condition data of capital goods and their subassemblies. Corrective regeneration is carried out in response to the failure of a component. Ad-hoc regeneration is defined as soon as the corrective regeneration is time-critical, i.e. must be carried out as quickly as possible. [15] The regeneration strategy is an important anchor for the planning of the regeneration of a capital good and therefore plays a central role in the characterization model. While the preventive, corrective and ad-hoc strategies are applied in a large number of supply chains, only some regeneration service providers apply the condition-based strategy. In the regeneration of aircraft engines, for example, various engine parameters are evaluated, while digital and analogue signals such as voltages and currents are evaluated in rail vehicles. [16, 17] The recorded parameters serve to estimate the probability of failure of individual components of the capital goods thus to estimate the expected shop load and demand for capacity. [17]

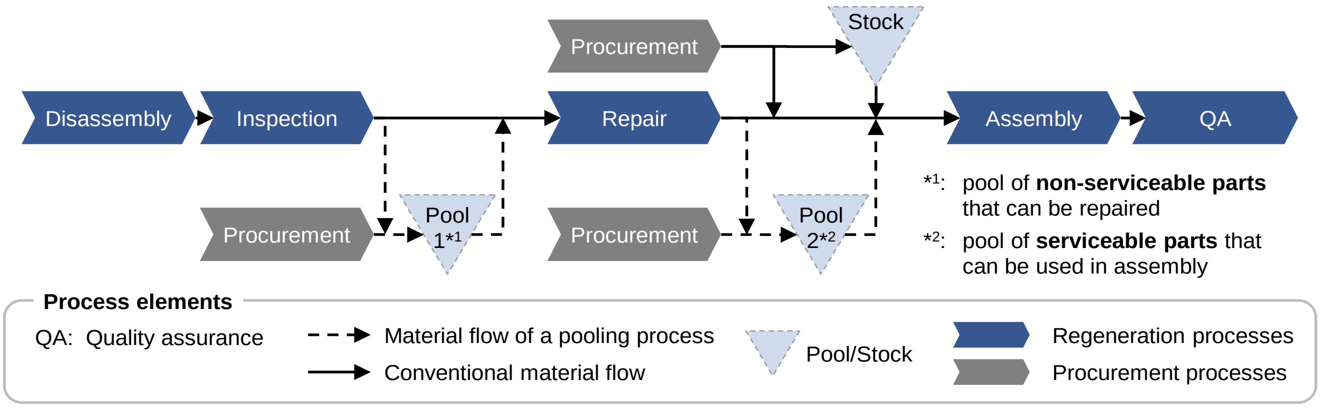


Figure 3 - Pooling stages in the supply chain for regeneration

A typical characteristic of regeneration supply chains is the use of relationships with the original equipment manufacturer (OEM) as a supplier of spare parts. If the assembly of used parts is permissible, a used spare parts market can be accessed; this is carried out, for example, in the regeneration of aircraft engines. There are regeneration supply chains (e.g. in the regeneration of rail vehicle transformers) which have integrated the production of spare parts into the conventional production of new parts.

The triggers that initiate the regeneration itself are described by the regeneration strategy. Identified regeneration strategies are the preventive, the condition-based, the corrective, as well as the ad-hoc regeneration strategy. Following the preventive strategy the execution of regenerative measures is based on defined time intervals or alternatively according to other criteria such as the achieving of a previously defined number of utilization units. With

To apply a condition-based regeneration strategy, a diagnosis needs to be done prior to the arrival at the regeneration service providers shop while the capital good to be regenerated is still in use. At this point, however, it is difficult to predict the condition of individual components even in preventive, planned regeneration. [18] The accuracy of information regarding the incoming workload is an important input variable for capacity and material planning in regeneration. This is underlined by the increased attention of this topic within scientific literature [3, 10, 19, 20]. The methods described can be divided into statistical and condition-based methods. Within the framework of statistical methods, for example, failure probabilities of certain components are calculated on the basis of historical values, while condition-based methods evaluate the need for regeneration by measuring different parameters of the capital goods during operation. In aircraft engines, for example, pressures and temperatures are recorded at different points inside the engine, as well as shaft speeds or fuel flow during operation. [17, 21]

While a diagnosis is used to determine the state of a capital good during operation, the inspection is used to assess the state of a capital good and its components while it is located at the facilities of the regeneration service provider. In the majority of the considered supply chains, the inspection is either preceded by or carried out during the disassembly process. However, the inspection strategy for the regeneration of rail vehicle transformers differs from the other supply chains considered here. Here the disassembly is preceded by the inspection. External damage to the capital goods can be detected by visual inspection, while the internal state of the transformer can be determined by electrical measurements and the analysis of an oil sample without previous disassembly activities.

Besides the characteristics that already have been named, some regeneration supply chains also showed special characteristics with regards to capacity planning. So in the regeneration of rail vehicle transformers the repair measures are implemented in the serial production line and thus competing for capacities with the new parts production. Due to the quality standards in the aviation industry regeneration service providers for aircrafts and aircraft engines have to outsource specific repair measures to external vendors, if they are not certified for those procedures. This also represents a strategy to overcome shortages in capacity.

3.3 Summary

With regard to the composition of the analysed regeneration supply chains, it was largely possible to determine similarities in the individual processes found in each case. Inspection, disassembly, repair, assembly and quality assurance exist as sub-processes in all of the considered supply chains. This confirms the general validity of the generic process model (see Figure 1). Diagnostic procedures and pooling processes, however, do not exist in all the analysed supply chains for regeneration. The regeneration strategy and the primary target of the regeneration planning are alike for all considered supply chains. It is regenerated both preventively and correctively. Usually, the planning aims at short throughput times, which at the same time mean shorter times until the goods can be put back to operation. Adherence to delivery dates was also confirmed as the primary target figure. Regeneration costs, on the other hand, were set more rarely as the primary target. Pooling is carried out in most regeneration supply chains and can therefore be mentioned as a central feature of a regeneration supply chain. The use of pooling coincides with the choice of the primary targets ‘throughput time’ and ‘on time delivery’, as pooling can support the achievement of these targets.

4 CONCLUSIONS

Within sub-project D1 of the Collaborative Research Centre (CRC) 871, “Regeneration of Complex Capital Goods”, the Institute of Production Systems and Logistics (IFA) is focused on modelling, planning, controlling and assessing the logistics of regeneration processes. The goal of this research is to increase the logistic performance of regeneration processes by developing an assessment model for quantitatively evaluating arbitrary and common supply chain configurations with regards to their attainable logistic performance and resulting logistic costs.

As a basis, a characterization model was presented in this paper that gives an overview of possible configurations of regeneration supply chains that have to be taken into account for developing the assessment model. To ensure the universal applicability of the assessment model, the regeneration supply chains of various capital goods were analysed, compared and examined with regard to individual or uniform characteristic configurations. The analysis showed broad similarities regarding existing sub-processes and their specific sequence in the considered supply chains among other similar characteristics like applied regeneration strategies and primary targets of regeneration planning. These similarities exist regardless of differences in the value or complexity of the capital good.

5 ACKNOWLEDGMENTS

The authors kindly thank the German Research Foundation (DFG) for the financial support to accomplish the research project D1 “Modelling Regeneration Supply Chains” within the Collaborative Research Centre (CRC) 871 – Regeneration of Complex Capital Goods.

6 REFERENCES

[1] Lohse, A., 2002, Auftragsmanagement von komplexen Produkten in agilen Unternehmensstrukturen, IFU Series. Volume 5. Shaker, Aachen.

[2] Herde, F., 2017, Revenue-Management-Ansatz für eine Annahmesteuerung kundenspezifischer Regenerationsaufträge komplexer Investitionsgüter, Springer, Wiesbaden.

[3] Eickemeyer, S.C., 2014, Kapazitätsplanung und -abstimmung für die Regeneration komplexer Investitionsgüter, Dissertation, Leibniz Universität Hannover.

[4] Eickemeyer, S.C., 2010, Kapazitätsplanung und -abstimmung bei unscharfen Belastungsinformationen.
ZWF - Zeitschrift für wirtschaftlichen Fabrikbetrieb, vol. 105, no. 4, pp. 323-327.

[5] Remémy, C., Staudacher, S., 2011, MRO: Organisation der Produktion sowie von

Produktionsplanung und -steuerung, wt - Werkstatttechnik online, vol. 101, no. 4, pp. 242-248.

[6] Nyhuis, P., Wiendahl, H.-P., 2012, Logistische Kennlinien. Grundlagen, Werkzeuge und Anwendungen, 3rd Ed. Springer, Berlin, Heidelberg.

[7] Berkholz, D., 2012, Grundmodell zur Kapazitäts- und Belastungsabstimmung eines Arbeitssystems in der Regeneration, Dissertation, Leibniz Universität Hannover.

[8] Lödding, H., 2016, Verfahren der Fertigungssteuerung. Grundlagen, Beschreibung, Konfiguration, 3rd Ed. Springer, Berlin, Heidelberg.

[9] Lucht, T., Schäfers, P., Nyhuis, P., 2018, Durchgängige modellbasierte Bewertung von Regenerationslieferketten, ZWF - Zeitschrift für wirtschaftlichen Fabrikbetrieb, vol. 113, no. 4, pp. 220-224.

[10] Eickemeyer, S. C., Doroudian, S., Schäfer, S., Nyhuis, P., 2011, Ein generisches Prozessmodell für die Regeneration komplexer Investitionsgüter, ZWF - Zeitschrift für wirtschaftlichen Fabrikbetrieb, vol. 106, no. 11, pp. 861-865.

[11] Herde, F., 2013, Rahmenbedingungen der industriellen Regeneration von zivilen Flugzeugtriebwerken, BoD – Books on Demand, Norderstedt.

[12] Lund, T., 1998, Remanufacturing: an American resource, Proceedings of the Fifth International Congress Environmentally Conscious Design and Manufacturing, Rochester Institute of Technology, Rochester, NY, USA.

[13] Kuprat, T., 2018, Modellgestütztes Ersatzteilmanagement in der Regeneration komplexer Investitionsgüter, Dissertation, Leibniz Universität Hannover.

[14] Georgiadis, A., 2017, Ersatzteildisposition und Reihenfolgebildung in der Produktregeneration, Berichte aus dem IFW, vol. 6.

[15] Kelly, A., 2007, Strategic maintenance planning, 1st Ed. reprinted, Elsevier, Amsterdam.

[16] Roberts, C., Chen, J., 2006, Railway condition monitoring, Condition monitoring and diagnostic engineering management, pp. 575-584.

[17] Bräunling, W. J. G., 2015, Flugzeug-triebwerke. Grundlagen, Aero-Thermo-dynamik, ideale und reale Kreisprozesse, thermische Turbomaschinen, Komponenten, Emissionen und Systeme, 4. Ed., Springer, Berlin.

[18] Dombrowski, U., Sendler, M., Schröer, N., Kaufmann, M., 2017, Kapazitätsplanung in Instandhaltungswerkstätten, ZWF - Zeitschrift für wirtschaftlichen Fabrikbetrieb, vol. 112, no. 11, pp. 753-758

[19] Eickemeyer, S. C., Borcherding, T., Schäfer, S., Nyhuis, P., 2013, Validation of data fusion as a method for forecasting the regeneration workload for complex capital goods, Production Engineering, vol. 7, no. 2-3, pp. 131-139.

[20] Kellenbrink, C., Herde, F., Eickemeyer, S. C., Kuprat, T., Nyhuis, P., 2014, Planning the Regeneration Processes of Complex Capital Goods, Procedia CIRP, vol. 24, pp. 140–145.

[21] Burnell, D., 1995, Engine Condition Monitoring - Engine System and Aircraft Operational Effects, Proceedings of the 18th Symposium Aircraft Integrated Monitoring Systems.

7 BIOGRAPHY

Torben Lucht, born in Germany in 1991, studied industrial engineering with the focus on production technology at RWTH Aachen University. Since 2018, he works as a research associate in the field of production management at the Institute of Production Systems and Logistics at the Leibniz University Hannover.

Tim Kämpfer, born in Germany in 1992, studied production engineering and logistics at Leibniz University Hannover with the focus on technical logistics and supply chain management.

Prof. Dr.-Ing. habil. Peter Nyhuis, born in 1957, studied mechanical engineering at Leibniz University Hannover and subsequently worked as a research assistant at IFA. After completing his doctorate in engineering, he received his habilitation before working as a manager in the field of supply chain management in the electronics and mechanical engineering industry. He is heading the IFA since 2003. In 2008 he became managing partner of the IPH - Institut für Integrierte Produktion Hannover gGmbH.

Antifragility as a Means to Utilize Uncertainty in Product Development Decisions

J. de Lange[1], E. Lutters[1], D. Kennon[2] C.S.L. Schutte[2]
[1]Dep. of Design, Production and Management, University of Twente, The Netherlands
[2]Dep. of Industrial Engineering, Stellenbosch University, South Africa

Abstract

Decisions are the common denominator of any product development cycle by fostering and consolidating the product definition, the requirement specification and the design rationale of these cycles. Such decisions are generally arduous as they are often based on incomplete information and always involve significant levels of uncertainty. This ubiquitous influence of uncertainty is characteristic of any product development cycle. Habitually, this influence is regarded as predominantly negative, so trying to mitigate uncertainties seem the obvious thing to do. However, uncertainty is not by definition detrimental or undesirable. In essence, it is hardly more than a common characteristic of any activity, decision or result in product development cycles. Primarily, uncertainty in product development should therefore be regarded neutrally. In this respect, the notion antifragility not only assumes such a neutral stance, it relates uncertainty to the response of a system – in this case an evolving product development cycle. By means of this theory, this paper aims at addressing and utilizing uncertainties with regard to the impact that decisions have on the product development cycle. The resulting solution principles allow to assess the potential impact of decisions and to determine and evaluate the desired behaviour of the product definition on various types of uncertainty. Simultaneously it allows for rigorous evaluation of both the quality and veracity of design decisions, aiming to make uncertainty instrumental in governing product development cycles.

Keywords
Antifragility, product development, decision making

1 INTRODUCTION

The core responsibilities of product development are under constant pressure and subjective to change. Important global development in, for instance, industry 4.0 and circular economy challenge the feasibility and viability of the product development strategies currently applied. This raises questions on how product development cycles can anticipate on any such changes. This paper does not propose one specific solution for the aforementioned developments. Rather, it focusses on dealing with the fundamental aspect of uncertainty when considering product development under changing circumstances. Uncertainty is a common characteristic of every product development cycle. Understanding, determining or estimating its impact on decision making is one of the most difficult, yet essential, tasks for a product developer. Habitually, the impact of uncertainty is regarded as being predominantly negative. Consequentially, trying to mitigate uncertainties seem the obvious thing to do. However, uncertainty is not by definition detrimental or undesirable. This paper researches the impact of uncertainty, explores the contribution to be made by defining antifragility and outlines a framework in which decision impact can be addressed explicitly. The aim is to make uncertainty instrumental and instigate continuous improvement of the product development cycle.

2 UNCERTAINTY IN PRODUCT DEVELOPMENT

2.1 Non-deterministic aspects

Product development of discrete physical artefacts as it is regarded in this paper is not an isolated discipline; rather, it is understood as the integrated collection of all preparatory work leading to the actual realisation and manufacturing of products. Consequently, the result of a product development cycle is not the physical product itself, but a formal definition of that product. Whereas product development might start from a clear stated purpose [1], the context and development cycle itself is essentially characterised by incomplete, indecisive and unreliable information in a development cycle encompassing an abundance of explorations, iterations, endeavours and (re)considerations. At the same time, the resulting product definition can only drive the actual manufacturing processes if it is transparent, unambiguous and complete. In the evolvement of that definition, designers continuously aim to translate a set of requirements, perceptions and inspirations into well-defined solutions, although such requirements often fail to be complete, concrete and objective. To bring product development cycles to a conclusive and successful result, a product development team needs to unequivocally convert the innumerable inputs, insights and influences into an objective and communicable product definition.

In product development, the evolving product definition is constantly weighed against the (also evolving) requirements specification and the initial stated purpose. This discrepancy is in fact one of the thriving factors of that product development cycle. Any evolving product definition might not meet important requirements, thus leading to new concepts and solution directions. However, it also works the other way around: an intuitive new solution might offer a completely new perspective on either the problem, thus nullifying some requirements and fostering the need to redefine these requirements.

Reaching an adequate compromise between requirements specification and product definition can be seen as a stopping criterion for the development cycle. At the same time, any inconsistencies will incite product developers to further the development cycle. In challenging the stability of the current state of development, for example the uncertainty about the product definition, the requirements specification and the balance between the two can be instrumental. As such, product development cycles are regarded as a system of evolving product definition and requirements and decision making.

Product developers have adopted to becoming experts in following two paths at the same time. They can follow a designated route - established by for example a design method or a stage-gate process - while they simultaneously have the ability to leave the beaten track by dealing with the uncertainty, ambiguousness and unpredictability that characterises the development cycle. Product developers continuously weigh the reliability of deliberate and reproducible considerations against the potential added value of daring ventures with many visible as well as implicit or hidden impacts. Striking a balance between reproducibility and potentiality is a non-deterministic act by definition. This act is the core origin of indefiniteness in product development and with this the inducement of uncertainty in all its appearances.

2.2 The impact of uncertainty

Uncertainty is present on all levels and in all aspects of product development. As mentioned, mitigating the potential effects is seen as the logical approach. Once a product is being manufactured, an upper threshold will be allowed for uncertainty to limit the risk on instantiated products. Illustrative for this upper threshold of uncertainty is the tolerancing aspect of technical drawings when transferred from design to a manufacturing department. Whereas the designer sees a dimension with a certain (functional) tolerance, the machine operator can merely attempt to manufacture the product with an accuracy that lies within the bandwidth specified by that tolerance. The relation between this accuracy and the tolerance is surrounded by different types of uncertainty, influencing producibility and functionality of the product while interpreting the information available. In avoiding the impacts of such uncertainties, the dominant approaches range from robust design and reliability design to tolerancing design [2]. However, this only partly covers the influence and symptoms of uncertainty in product development cycles.

A claim made in this research is that uncertainty in itself can have a positive and constructive role in product development. This positive influence stems from the fact that uncertainty in any real-world situation is ubiquitous and therefore actually contributes to the need for new or altered products. Moreover, in product development cycles, uncertainty is a precondition to be able to explore both the problem and the evolving solution in a concurrent and integrated manner. In other words, if any product definition would not contain uncertainty from the beginning, the entire endeavour of product development would be pointless. In fact, without uncertainty the rationale behind the entire design discipline would be lost. Therefore, uncertainty in product development is by no means detrimental or undesirable by definition. Uncertainty is hardly more than a characterisation of every aspect of the product development cycle and every piece of the corresponding product definition. Consequently, it makes sense to observe any product development activity or any part of the product definition as not being entirely fathomable and therefore including uncertainty by definition. The current dominant attitude and corresponding approaches towards uncertainty do not cater for a terminology that allows for a two-sided (negative and positive) perspective on uncertainty. Based on a situational characterisation of uncertainty, the next section introduces the concept of antifragility to fill this hiatus.

3 ANTIFRAGILITY AND UNCERTAINTY

The notion uncertainty is explored and defined before focussing on the potential effects of uncertainty on the product definition and the corresponding development cycle. It is essential to understand the way in which the entire product definition responds to uncertainty. In other words: it becomes important to define the response to a change in, or an increase of, uncertainty in development cycles.

3.1 Volatile, uncertain, complex and ambiguous

The notion uncertainty is contextualized in terms of the VUCA framework [1, 3]. Both the focus on the indefinite aspects of any situation and the inherent appreciation of lifelike decision processes make this framework an appropriate backbone for describing uncertainty in relation to product development. The acronym VUCA stands for Volatility, Uncertainty, Complexity and Ambiguity and is intended to offer a guide in how to assess the status and criticality of a situation:

- *Volatility* describes the variability of situation. A volatile situation is subjective to change in which timing and magnitude of events not necessarily relate to the significance of the impact.

- *Uncertainty* can be divided into ingrained and uncontrollable parts and manageable uncertainty. Manageable uncertainty relates to that portion of uncertainty an actor can purposefully reduce, providing that sufficient time and efforts would be available to do so.
- *Complexity stems from* the inability to disentangle any state into addressable sub-states. With this, complexity exceeds intricacy and complicatedness, because of the interdependent perspectives and levels of aggregation.
- *Ambiguity* is the state in which a situation, concept or decision outcome has more than one meaning. Consequently, ambiguity delineates circumstances in which it is – by definition – not knowable to determine which uncertainties are at play.

3.2 Fragile, robust, resilient and antifragile

In a product development cycle, the coherent set of requirements specification, evolving product definition and design decisions is the system that will show a response to any stressor, explicitly including any VUCA aspect of such a stressor. The way in which that system responds to such a stressor is characteristic for the chances of success of the development cycle. If the system is incapable of dealing with a stressor, it is said to be fragile; it 'breaks'; this renders fragility an unwanted state, unless it is included as a purposeful 'fuse' that protects from irresponsible impactful steps in the development cycle. Robust and resilient system responses are often seen as the opposite of fragility, in which the system can withstand or recover from stressors respectively. The concept antifragility challenges the contrast between fragility and robustness/resilience. This stems from observations that systems can naturally flourish in volatile, uncertain, complex and ambiguous situations [4], e.g. in option trading. This positive effect does not necessarily correspond with the defensive or restorative aspects found in robustness and resilience. As such, Taleb introduced antifragile as the inverse of fragile responses, in which the response on uncertainty is completely positive [4]. For systems and organisations, antifragility offers new concepts on how to respond to (un)expected events and how to learn from it. For instance, based on an antifragility assessment framework, an organisation can measure and identify improvements on e.g. de-centralisation and flexibility that are derived from the typical antifragile system response [5]. For product development cycles and the corresponding product definition it offers a vocabulary that covers all potential effects on uncertainty. With antifragility being the system response that, now more explicitly, explains a system's gain and/or improvement from disorder [4]. This, together with the previously understood responses to uncertainty in development cycles provide a set of negative, neutral and positive results. With that, the set of possible responses of stressors to the development cycle becomes:

- *Fragility*: a system response that fails if introduced to the various types of uncertainties.
- *Robustness*: a system response that is sufficiently shielded against known uncertainties
- *Resilience*: a system response that can absorb the effect of uncertainties and – in time – return to the original functioning.
- *Antifragility*: a system response that absorbs the effect of uncertainties and additionally gains in quality of functioning or even new functionality.

Initially, fragility, robustness / resilience and antifragility were seen as being on one dimension [4]. Although this yields a logical and straightforward way to assess overall system responses, this continuum only yields true at one level of aggregation. Considering that any system consists of many parts or subsystems, the validity of this one-dimensional representation is already obsolete. After all, it might be indisputable that fragile behaviour of a subsystem might contribute to the overall antifragile response of a system. Therefore, this research positions system (also to include sub-system) responses in a triangular shape, thus transcending the linearity of the one-dimensional approach (Figure 1). This triangle spans an area of system responses in which robustness is not necessarily at the centre between fragile and antifragile [6]. This allows to better distinguish between required responses on multiple levels of aggregation, thus aiding product developers; not only in their tasks to further the evolving product definition, but also in more precisely addressing uncertainties in the appropriate manner.

Antifragility contributes to addressing the influence of uncertainty in a more (than) neutral manner. Simultaneously, it challenges the current practices in product development. For instance, designing with strict boundaries might not necessarily lead to antifragile system responses [7]. When translated to an evolving product definition, the vocabulary allows to (purposefully) determine the manners in which a product definition might or should respond to different types of uncertainty. Antifragility provides for a

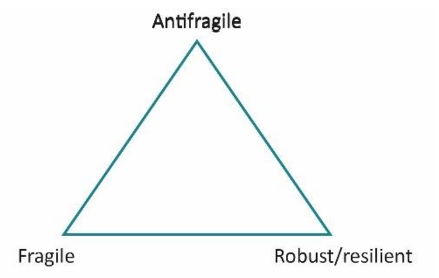


Figure 1 - Triangle of system responses (redrawn after [6])

definition of system responses to decision impact and changing circumstances and uncertainty as either negative, neutral or positive. This aligns with the possible impacts decisions have on the product development cycle and therefore does (more) justice to the overall potential impact of uncertainty in product development.

4 DECISION IMPACT MODEL

This section aims to combine all previously addressed subjects into core solution principles that enables decision impact assessment as being part of design support.

This research uses the decision impact model as a product development model [8]. Rather than relying on the process-based depictions of development cycles this model approaches product development as a discipline in which a diverse group of stakeholders conjointly develop, influence and adjust a product definition via a multitude of decisions. One of the main reasons for this differing point of view is that the modelling of product development as a series of sequential steps or phases does not necessarily do justice to the nondeterministic aspects of the development cycle. Given this decision-based approach, it is important to consider that:

- The product definition is at the core of any product development cycle. Considering this pivotal importance, centralizing its role in design support is a logical consequence. This product definition is expressed using various forms of implicit and explicit information. Consequently, the product definition is understood to be a collection of interconnected information entities.
- The process component of product development is defined as a network of decisions that have logical dependency but that are not necessarily sequentially organized in time. As such, there is a clear relation between decision making and the information content in a development cycle. Even more, if designing is understood in terms of creating (new) information content, rather than as a predefined course of actions, then decision making is the elementary activity in establishing that information content.

The decision impact model describes the relationships between decisions, the product definition and the product development environment (**Error! Reference source not found.**2), rendering it comparable to the product development system discussed in 3.2. It addresses decision making as an independent activity that meets the urge to influence the product definition, in other words, to transform the currently available information content. It does not put focus on the decision process itself, but on the effect that this decision will potentially have on the product definition. These decisions, as the common denominator of development cycles, is assumed to potentially affect the (evolvement of) the product definition, regardless of whether this decision is a minor change of a technical drawing made by one actor or strategically, a new solution direction involving multiple actors. Consequently, the coherent set of design decisions is seen as the scaffolding that enables a design team to initiate, substantiate and assess the evolvement of product definition against requirements specifications in its development context. In extrema: decision making is no longer seen as the activity that leaves behind a trail of information, it now becomes a tool to allow for purposeful evolvement of the information content.

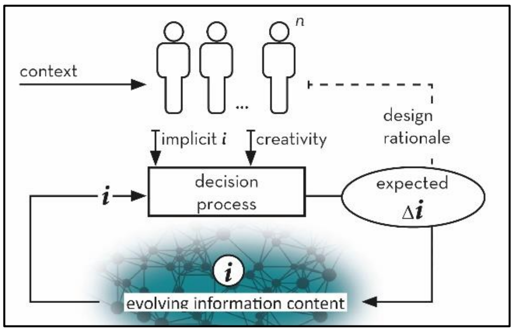


Figure 2 - Decision impact model

4.1 Decision as stressors on the product definition

As indicated in **Error! Reference source not found.**2, decisions and the product definition are defined in terms of the information content. Whereas the product definition is considered to be a collection of interconnected information sources (i), a decision is defined as the potential effect it is expected to have on that collection (Δi). The model stresses both the relation between decision making and the evolvement of the product definition as well as the importance to monitor the actual Δi once processed in the information content.

Based on the VUCA descriptions given in section 3.1, the various aspects of uncertainty can be related to information. In this, complexity is considered to be omnipresent, rudimentary and ingrained to such an extent that it will currently not change our ability to reach decisions. When expressed in information, the other aspects of uncertainty can be defined as decision characteristics [1]:

- Volatility addresses the *temporal validity* of a decision
- Manageable uncertainty addresses the *predictability* of the intended outcome of a decision
- Ambiguity addresses the *meaning* of a decision

Combined, these aspects address the veracity of the information entities that collectively embed a decision. While this covers the basic aspects of uncertainty and aids in making aspects of uncertainty explicit in information, it does not yet allow to assess its impact. With the concept of antifragility, the relation between decision making, requirement specification and the product definition can be defined in terms of stressors, system responses and uncertainty. Here, a decision can be defined as a (foreseen) stressor on the product development cycle. As such, decisions conjointly contribute to the overall (internal) stress of a product definition:

- *Decision impact* expresses the intended or expected response of the product definition on the stress caused by the decision.

The concept of antifragility and the corresponding full range of potential system responses enables us to explicitly address the responses of the product development cycle on such impacts. Combining this with the decision impact model, two basic principles for decision impact assessment are defined. Based on this core functionality, the intended design support functionality aiming at governing and improving development cycles under changing circumstances can be defined.

4.2 Intended design support functionality

By adding 'decision impact' as a decision characteristic, design decisions amalgamate in a network of stressors that influence the product definition. As a result, the status of a product development cycle can be represented by an evolving system with a certain amount of internal stress. Obviously, any product developer would benefit from being able to see the impact of reaching a certain decision with a certain Δi as a consequence. Moreover, if that product developer would be able to not only see the output of the decision, but also could predict the system response as a whole, it would allow him to be more effective and efficient in decision making, while explicitly considering the uncertainties related to the decision and the decision process. Moreover, it would allow product development teams to quickly and adequately adapt to changing circumstances. This functionality is independent of the methods, processes or working approaches that are applied, as it is directly linked to the evolving product definition itself. It enables a type of sensitivity analysis on the product development cycle itself that aids in determining appropriate measures for changing circumstances. In this, the vocabulary of system responses aims at making this impact instrumental in governing product development cycles. The next section outlines the design brief for this intended design support functionality and corresponding requirements.

5 DESIGN BRIEF

The solution principles as outlined in this paper are intended as an enabler for design support that allows product development teams to explicitly address and steer the (desired) behaviour of the product development cycle on various impacts and corresponding types of uncertainty. This core functionality should enable product development teams to prototype development cycles and better understand useful configurations in context to continuously improve product development cycles.

The design brief outlined in this section should grow into an evolving requirement specification that addresses the desired range of functionalities. Case studies addressing specific application domains and corresponding trends are considered to be vital in understanding these basic needs. Simultaneously the context and specific development environment involving the various stakeholders, type of industry and product, applied methods, tools/ techniques and the business context needs to be captured. These industry-specific scenarios are deemed indispensable in capturing the rationale of these specific functionalities. Moreover, it provides essential context and insights for developing specific solutions and a generic applicable approach.

Applicability for various industries in different product development domains addressing e.g. the beforementioned developments is essential. Therefore, the approach should not be imposed as a static, predefined method. It should rather be introduced as an inherent part of the common design support arsenal available for product development teams in daily practise. Consequently, developing this design support cannot be done without high involvement of industry. The aim here is to conjointly develop both the generic approach and tailored solutions for specific industry cases.

An information backbone is required that allows for the envisioned usage of information in determining decision impact. This backbone needs be able to capture the cycle as an interconnected system of information entities, independent of the many tools, standards and approaches used in managing this information. This network of information can be used as an explicit reflection of the current state in any product development cycle. The introduced decision characteristics on temporal validity, predictability and meaning aid in improving the veracity of this rendering. Moreover, this collection constitutes the basics for the required decision impact assessment.

Interfaces should be established that allow interaction with the information backbone aiming at simulating decision impact, tracing design rationale and monitoring decision making. For decision impact modelling focus is put on the integration of existing antifragility assessment frameworks in the decision impact model. The key challenge here is to find effective means to cater for a quick but adequate assessment appropriate for the dynamics and relative fast pace in decision making, allowing product developers to prototype and evaluate various scenarios. This impact assessment functionality needs to transcend any individually made impact assessment. Due to the overall non-deterministic character of product development cycles it is expected that not every decision taken will explicitly be available or definable. However, in using the information backbone as the starting point for every impact assessment and assuming that every consequences of any decision will (eventually) be, at least partially, processed in this information collection, the design support system should cater for monitoring decision making and tracing design rationale. In prototyping various configurations of product development cycles, it is essential to allow for smooth incorporation of (new) knowledge, tools and techniques. Simultaneously, the establishment of a testbed for the development of the envisioned design support itself is essential in test-driving potential solution principles. This allows for meaningful iteration and, more importantly, sets up a research environment in which the characteristics of antifragile behaviour can be made instrumental.

6 CONCLUDING REMARKS

Developments such as Industry 4.0, Circular Economy and Additive Manufacturing can have significant impact on the way products are currently developed. Together such developments constantly challenge the feasibility and viability of the chosen strategies and approaches for product development cycles. Consequently, it is difficult if not impossible for product development teams to control the effectiveness and efficiency of product development cycles while simultaneously developing sound product definitions. Decisions have to be made under time pressure, with limited resources, based on incomplete information and under high levels of uncertainty. As such, it involves a constant trade-off between doing the right things and doing the things the right way, without proper means to address the corresponding impact of such trade-offs beforehand.

The aim of the design support as outlined in this paper is to empower project teams to continuously improve their development cycle under such constantly changing circumstances. Essential in this design support is the impact assessment of a decision – not only on the product definition itself but on the product development trajectory as well. By means of the concept of antifragility and the decision impact model the corresponding behaviour of the product development system can be assessed. In utilizing the various types of uncertainty involved, the impact of decisions on the effectiveness and efficiency of the product develop cycle is made explicit. This core functionality act as an enabler to fast-track continuous improvement. It allows organizations to quickly and adequately prototype development cycles in appropriate timing and context. Moreover, it allows to better understand potential useful configurations of development cycles, capturing and keeping best practices while making other parts of the development cycle adaptable and flexible where required.

7 REFERENCES

[1] De Lange, J., E.J. Oude Luttikhuis, and E. Lutters, *Incertitudes in design decisions – impediments or instruments?*, in *COMA International Conference on Competitive Manufacturing*. 2016: Stellenbosch, South Africa.

[2] Morse, E., et al., *Tolerancing: Managing uncertainty from conceptual design to final product.* CIRP Annals, 2018. **67**(2): p. 695-717.

[3] Stiehm, J. Hicks, and N.W. Townsend, *The U.S. Army War College: Military Education in a Democracy.* The U.S. Army War College: Military Education in a Democracy, 2002: p. 6-6.

[4] Taleb, N.N., *Antifragile: things that gain from disorder*. 2012: Random House.

[5] Kennon, D., C.S.L. Schutte, and E. Lutters, *An alternative view to assessing antifragility in an organisation: A case study in a manufacturing SME.* CIRP Annals, 2015. **64**(1): p. 177-180.

[6] Kennon, D., *Towards an Antifragile South African SME*, in *Industrial Engineering*. 2017, Stellenbosch University, South Africa.

[7] Jones, K.H., *Engineering Antifragile Systems: A Change In Design Philosophy.* Procedia Computer Science, 2014. **32**: p. 870-875.

[8] De Lange, J., *Information as a workpiece: an architecture for content-driven design support*, in *Design, Production and Management*. 2018, University of Twente: Enschede.

8 BIOGRAPHY

Jos de Lange obtained his PhD. Degree in Industrial Design Engineering at the University of Twente. Currently, he is assistant professor in the Department of Design, Production and Management (faculty Engineering Technology) at that university.

Eric Lutters obtained his PhD. degree in Mechanical Engineering at the University of Twente, The Netherlands. At that university, he currently holds a position as an associate professor in the faculty Engineering Technology. He also is a professor extraordinary at the Industrial Engineering department at Stellenbosch University.

Denzil Kennon obtained is his PhD. Degree in Industrial Engineering at the Stellenbosch University. Currently he is Senior Lecturer Extraordinary at that same university and investment Associate at Greenpoint Capital.

Corné Schutte obtained is his PhD. Degree in Industrial Engineering at the Stellenbosch University. Currently he is professor at the same university and head of the department of Industrial Engineering.

Towards a Practical Implementation of Technology Management Frameworks – A Structured Review of Established Literature

J.C. Visser[1,2], K.H. von Leipzig[2], S.S. Grobbelaar[2,3]
[1]Alphawave Mobile Network Services PTY (LTD), Stellenbosch, South Africa
[2]Stellenbosch University, Department of Industrial Engineering
[3]DST-NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy (SciSTIP)

Abstract
Scientific knowledge is often difficult to apply in a practical context. The use of frameworks can assist with the valorisation of conceptual or abstract knowledge into real-world applications. At the same time many companies, especially software companies, are diversifying their technological knowledge and capabilities by developing increasingly complex technology portfolios aimed at different markets. Such activities raise the question: How can diversified companies ensure that their technological knowledge and capabilities are shared and applied efficiently? Technology management attempts to identify and utilise technological knowledge and capabilities within an organisation in order to leverage its competitive advantage. As such, technology management attempts to bridge the gap between technological potential and corporate strategy, thus balancing the pressures of technology push with market pull. In this paper the authors present a structured literature review to analyse the practical application of technology management frameworks. A sample of 50 journal articles were collected and evaluated. From the review, the potential benefits of implementing technology management frameworks were clear, but also revealed that few empirical studies exist of the implementation of these frameworks, especially with regard to SMEs in general and South Africa in particular. Through the literature review the following can be achieved: (1) ascertain the potential benefits of technology management frameworks, (2) deduce a process to map and analyse the current state of technology management in firms aiming to develop an accurate and actionable technology management framework, and (3) highlight the need for further research in the SME and South African contexts respectively.

Keywords
Technology Management; Management of Technology; Technology Management Frameworks; Technology Management Tools

1 INTRODUCTION

Scientific knowledge forms the core of technological innovation. The practical application of scientific knowledge is essential for any technological organisation to develop products that can adequately satisfy market pressures. As organisations become more complex, so too do their technology portfolios. The importance of technology management's role in linking technological capabilities with broader corporate and business strategies have been noted in various studies [1]–[7]. Technology management (TM) frameworks are useful mechanisms to link the different aspects of the successful implementation of technology into actionable and communicable business strategies.

A common question is how to develop and implement such a framework in a real-word and practical context? This paper will attempt to answer this question based on the available literature. A structured literature review was undertaken to determine how technology managed frameworks are developed from scientific principles and implemented in real-world situations. Fifty articles were included in the review. They comprised the 30 most cited articles on the Elsevier SCOPUS database regarding Technology Management Frameworks[1] and the 20 most relevant articles of recent years[2].

The individual articles included were examined and common features relating to TM between the articles were identified. These features included different aspects of TM, such as different themes, underlying principles, as well as common features of TM frameworks and tools. The context of each article was also established by specifically noting the

[1] The search terms used for the identification of relevant literature was *"Technology Management Framework"* and *"Management of Technology Framework"* respectively.

[2] The 20 most relevant articles published between 2015 and 2017 were included based on a review of their respective abstracts.

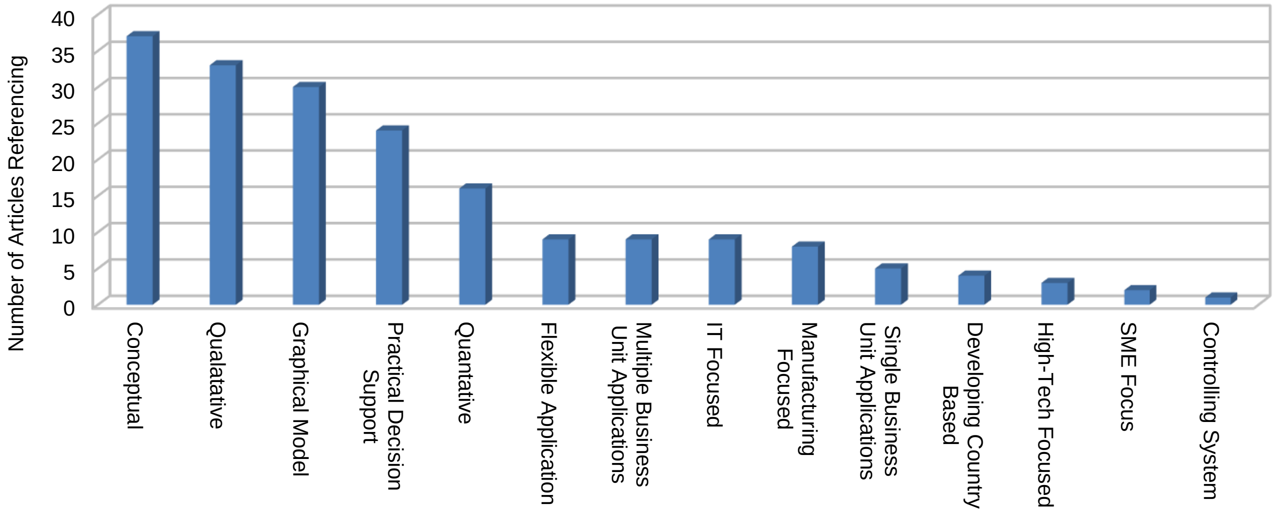


Figure 1 - Common TM framework features in articles under review

geographical context, the application level, and the industry in which the article is based. The dominant features were highlighted by plotting the occurrences of the various features on a matrix to clearly visualise trends in the literature. Due to space constraints and for the purpose of this article, the leading trends identified in this matrix are summarised in different graphs.

From the onset, identifying a single all-encompassing framework for technology management seemed an elusive goal. Various authors state that there is a need for a unified structure and framework in the technology management field with most of the resulting frameworks being only conceptual and generally unproven in real-word situations [2], [3], [6]–[8].

2 TECHNOLOGY MANAGEMENT FRAMEWORKS IN THE LITERATURE

When analysing the different frameworks covered by the 50 articles evaluated for this study, the most striking characteristic is the lack of a unified framework. Only one framework was discussed by more than one study, the Strategic Alignment Model (SAM). Although some of the principles of SAM are transferable to TM as a whole, it is primarily developed and implemented in the Information and Communications Technology (ICT) sector [9], [10]. In addition to the lack of a universal or unified framework to facilitate TM, various studies also call for methods to practically implement the principles of TM in a real world situation [3], [5]–[8], [11]. These two findings may indicate that TM practitioners tend to derive their own framework or approach which is best suited for their specific application.

Instead of only establishing the common features and characteristics of the various frameworks, it is also valuable to rather determine on what principles the various frameworks were based. These common principles between the various studies and frameworks then form the scientific core of a specialised custom-made framework for the specific application for which it is intended.

The common features of TM frameworks, as noted during this review, are shown in Figure 1. The majority of the frameworks covered in the various studies are conceptual and tend to be qualitative in nature. That said, most have some kind of graphical output or model and more than half of the frameworks applied offer some type of decision support.

Thus, although there is no unified framework available for TM, the aims of these frameworks seem evident. TM frameworks are developed and implemented to assist with strategic decision-making, whilst also enabling efficient communication between the different role players in an organisation.

3 CONTEXTUAL SITUATION OF STUDIES

The first definite observation that can be made from the articles reviewed is that the implementation of TM takes place mainly at firm level, as opposed to project, corporate or industry level. Only eight of the articles included in this literature review apply it at a level other than the firm. When attempting to implement TM in an organisation the firm seems to be the optimal level to focus on.

TMs origins lie mostly in the large manufacturing firms of the USA. The articles included in this review shows how TM has been predominantly applied in the manufacturing sector, followed by the ICT sector. It seems as if the potential advantages of TM in other sectors have only become apparent fairly recently and thus fewer articles discuss the implementation of TM in other sectors. Since the late 1990’s this has changed and with the increasingly rapid evolution of technology in all sectors, TM has also become increasingly important for companies to effectively leverage their competitive advantage [3], [8].

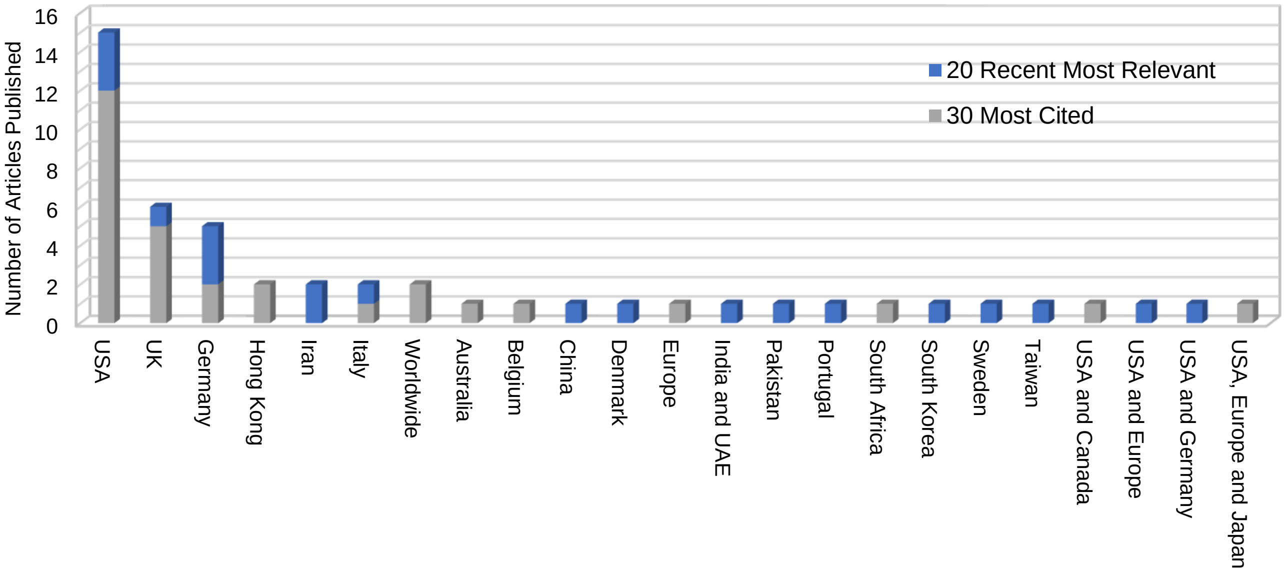


Figure 2 - Summary of articles reviewed based on geograhical context

TM was also viewed as important only for large firms with separate R&D departments. This view has also been changing over the past few years and new studies are being published detailing the importance and implementation of TM in firms of various sizes in different industries. The understanding and utilisation of the technological resources of a company is fast being seen as a core characteristic of companies that are able to innovate in dynamic economies and contexts [3], [8].

By noticing how TM has facilitated large corporations in developed economies to grow and flourish through turbulent economic times, many companies in developing economies are also trying to use TM to effectively leverage their unique technological capabilities [7]. TM seems especially important within economies where industrialisation is a priority.

It can clearly be seen in Figure 2, that the most articles included in this review were based on the context of the USA. As TM evolved out of necessity to counter the rapid advances of Japanese industry in the early 1980s, it is not surprising to see this trend. Following the USA is the UK and Germany. Articles based on the context of various different European countries and the UK contribute almost half of the articles included in this review. The skewness of this result is even more evident when analysing the articles included in the list of the top 30 most cited articles. Most of the articles regarding developing countries were published in the last three years. This seems to show that TM has mostly been studied and applied in industrialised countries, although there is some evidence that this has expanded to developing countries as well.

Indeed, the articles published regarding TM in Iran specifically discuss the importance of TM to develop a country's industrial capacity [5], [12]. Although not included in this study, this same sentiment is expressed by De Wet [13] in an article describing how TM should be used in developing countries trying to compete in technological markets. By incrementally developing their technological capabilities, companies in developing countries can increase their competitiveness on the global stage [12].

As with the focus shift of TM to the developing world, so too has TM been evolving out of its historical field of expertise. A large number of the studies included in the list of the 20 most relevant recent studies has focused increasingly on applying TM to various fields other than the traditional manufacturing and IT. There are a few studies which examine different diversified companies and attempt to show the relevance of TM in ensuring that they manage and exploit their diverse capabilities effectively in order to maintain their competitive advantage in various markets [5], [14], [15].

When viewing the results of the respective contextual situations of the articles included in the review it seems as if there is still scope for further research in the contexts where TM is newly being applied. In particular, the manner in which SMEs in a developing country, such as South Africa, is (or could be) implementing TM has the potential to grow the industrial capacity of the country. Firms with diversified capabilities would be especially valuable to research, based on the apparent success they have had in other developing countries. Only one study in the review was directly related to the South African context and did not examine the implementation of TM at firms as such.

4 LEADING PRINCIPLES OF TECHNOLOGY MANAGEMENT FRAMEWORKS

The leading theoretical principles referred to in this review are shown in Figure 3. As shown, the most common theoretical principle on which most studies based their frameworks and conclusions is the

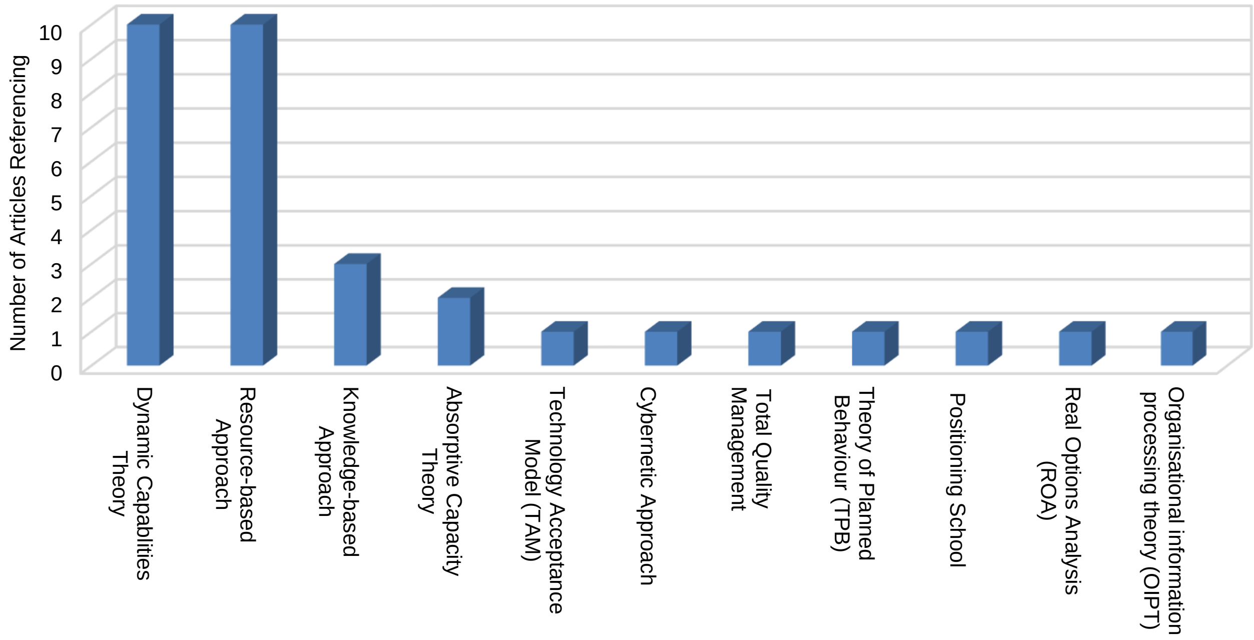


Figure 3 - Common TM principles in articles under review

Resource Based Approach [7], [8], [16], [17]. The next most common principle is the Dynamic Capabilities Theory [2], [3], [5], [6], [11], [17] as developed by Gregory [2] and is also based on the management of resources. The Dynamic Capabilities Theory seems particularly adept at managing the dynamic nature of technology and aspects of a firm. Another common approach is the Knowledge-based Approach [18]–[20], which is also an adaptation of the Resource-based Approach, which concerns itself with the management of knowledge in a firm instead of other resources [18].

These three approaches share the view that a firm should exploit its internal resources, capabilities and/or knowledge respectively to secure or maintain a competitive advantage. Especially when regarding the most recent articles included in the study, the resource-based approach and the related Dynamic Capabilities Theory has become the prevalent standard for TM frameworks. Twelve out of the twenty recent studies included are, at the very least, partially based on one of these three views. Six of the remaining eight articles do not specify the underlying TM principles. The final two articles refer to the Absorptive Capacity Theory and Total Quality Management. The Absorptive Capacity Theory is applied along with the Resource Based Approach in an article by Schuh and Kramer [8] already included in the previously mentioned twelve articles.

The Dynamic Capabilities Theory in particular expands on the concept of managing capabilities, but includes that the management of these capabilities should be dynamic in nature. Thus, a firm should be able to re-organise, re-align and expand their capabilities in a dynamic manner to effectively address rapidly changing external environments. This theory's premise is that effective management of technological capabilities can in fact be a core competence in itself [3]. This seems to be an ideal principle on which to base a framework to account for the unpredictable and dynamic state of technology and the related markets.

Two core objectives of TM were evident during the review of the literature, as shown in a summary of common TM themes noted during the review. The first is the importance of the alignment of technology and business strategy as one of the drivers of the successful leveraging of a firm's competitive advantage [1]–[4], [7]. The second objective, which is directly related to the former, is that TM facilitates communication within firms [1]–[3], [7]. In particular, TM attempts to bridge the communication gap between the business and technical elements in a firm. TM frameworks facilitate both these objectives, by firstly illustrating the linkages between business and technology objectives. Secondly, TM frameworks become efficient communication tools to share information and objectives at different levels and departments of firms [1], [3], [21].

The literature also highlights the overlaps between TM and other management forms. In particular, knowledge management and innovation management share many similarities with TM [3]. All three management types are concerned with the utilisation of resources available within a firm in order to obtain and sustain competitive advantages. The three different management forms are often used interchangeably in the literature and are being implemented simultaneously at an increased rate [3], [19].

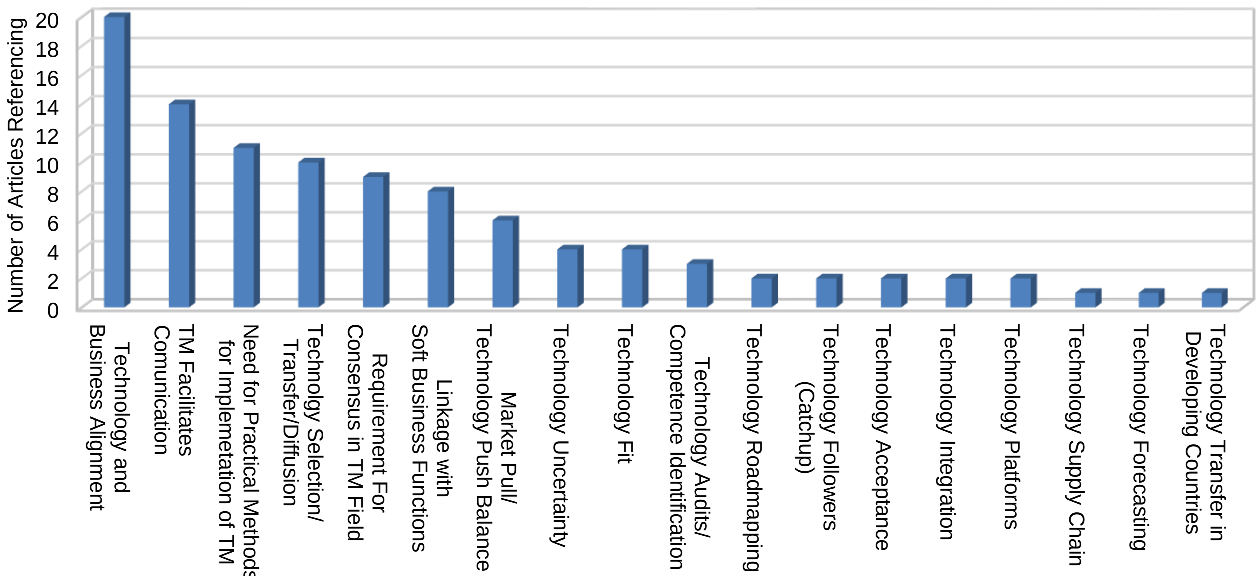


Figure 4 - Common TM themes in articles under review

5 CONSIDERATIONS FOR THE IMPLEMENTATION OF A TECHNOLOGY MANAGEMENT FRAMEWORK

With a large number of dimensions, the field of TM is one of complex interdependencies between many different role players and elements in any given firm [2]–[4]. This is well illustrated in Figure 4 and Figure 5, showing respectively the various TM themes noted during the review, as well as common organisational traits and other management aspects that influence the implementation of a TM framework in a firm.

To adequately manage the complexities of TM a balance must be struck between the two main drivers in any technological organisation. The one driver is market pull, which is the pressures and opportunities

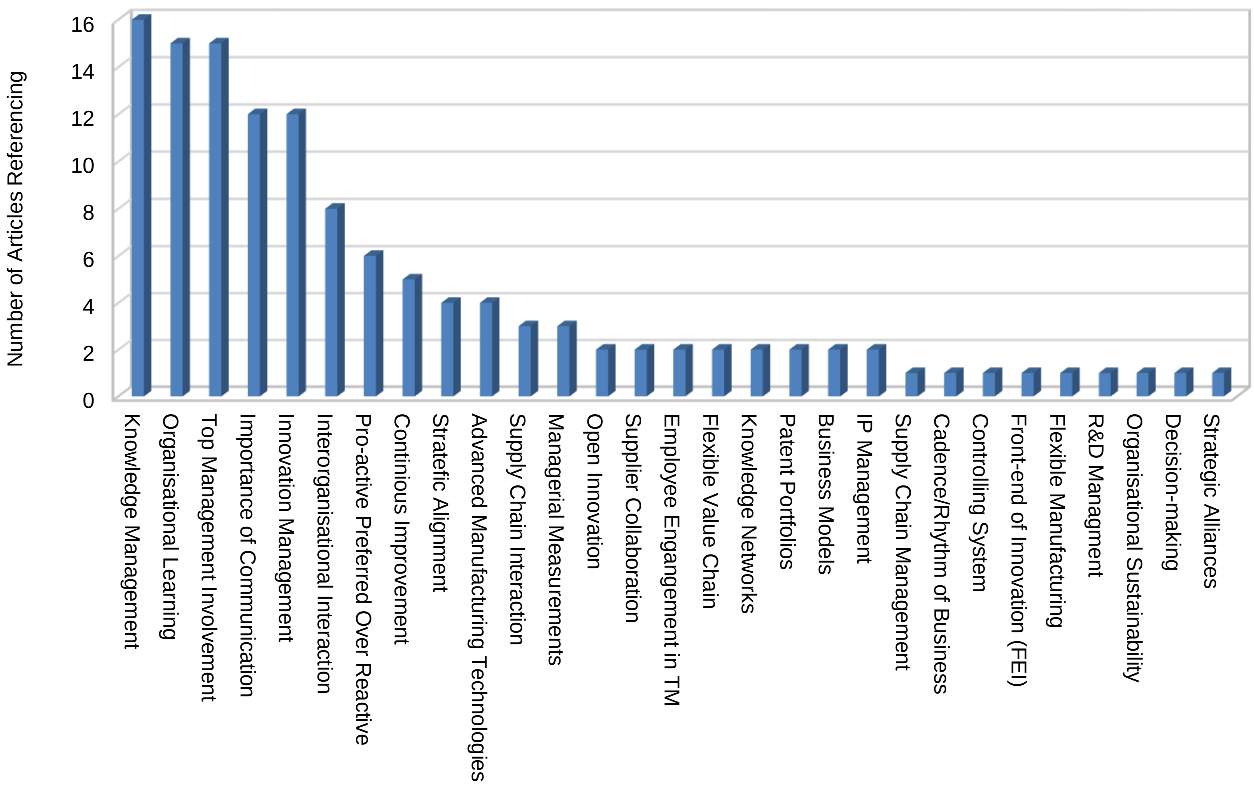


Figure 5 - Common organisational traits and other management themes in articles under review

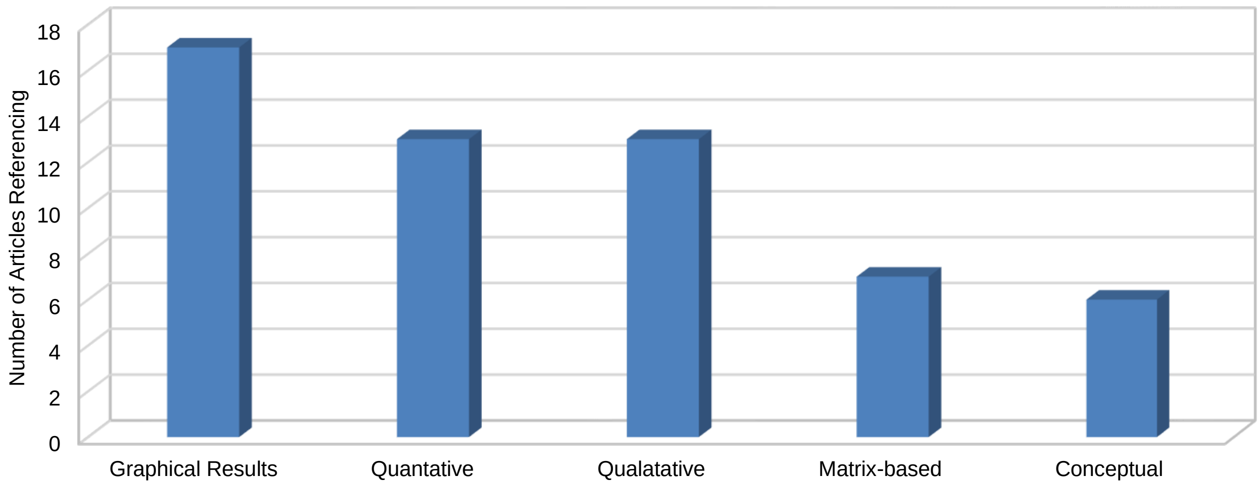


Figure 6 - Common TM tool features in articles under review

presented by different markets. The other driver is technology push, which is the potential of what can be created by the technological capabilities within a firm. A balance must be struck between what respective markets need and what a firm can potentially create. The optimum balance between the two drivers will lead to the most efficient leveraging of a firm's competitive advantage [3], [7], [21].

Finding this optimum balance is, however, no easy task and it is often difficult to combine all the different aspects related to technology into a manageable package. The first step to manage a firm's technology is to determine its current technological capacity. This is done through a process of technology auditing, wherein a firm determines which technological capabilities it possesses and how these capabilities are utilised to create products that can be supplied to relevant markets. Based on the results of these technology audits, a firm can determine the best strategies to follow to further expand and utilise their technological capabilities to their advantage [2], [3], [6].

To assist firms in identifying and managing the diverse applications and implications of TM, various tools have been developed and discussed in the literature. Similar to TM frameworks, it seems as if specialised tools are developed to achieve specific goals.[4]

As, again, there seems to be no unified set of tools to use for TM it is worthwhile to rather determine which features are most often associated with tools applied in the literature. These common features, as noted in the review, are summarised in Figure 6. As shown, the number of quantitative and qualitative tools are equally spread between the two. Some of these tools even combine quantitative and qualitative attributes into the same tool. Furthermore, most of the covered tools have some kind of graphical result, thus making the outcomes of these tools easily relatable to various role-players in TM [4].

Although a very small number of the tools covered were used more than once in any of the studies, there was one family of tools that did reappear often throughout the literature review. The matrix-based family of tools seems to be a popular set of tools and according to Phaal, Farrukh and Probert [4], these tools are easily relatable, flexible and adaptable. These tools can be used as qualitative, quantitative or both, depending on the application. As there are many different tools covered by the literature, the matrix family of tools seems especially appealing due to the fact that it can easily be adapted to the specific situation of a particular firm [4].

The tools used to develop and implement a TM framework require information from a variety of different role players in a firm. Structured interviews and workshops were popular ways to interact with subjects in the studies to obtain the required information. From the limited number of workshops, it does seem that interviews are the preferred choice, probably due to time constraints, which limit the practical arrangement of workshops. Workshops, especially recurring workshops, are time-consuming, thus interviews seem to be a more manageable alternative to engage with subjects.

6 CONCLUSIONS

The structured literature review of fifty different studies have revealed that although there is no single framework for TM, there are certain principles that have become the foundation of the field.

In particular, the resource-based approach and its derivatives (the Dynamic Capabilities Theory and the knowledge-based approach) seem to have become the prevalent standard for TM frameworks. In this approach a balance needs to be struck between the different drivers of TM. This balance can only be achieved if the there is sufficient alignment between the business and technology functions of a firm.

This balance requires efficient communication between the various role-players. TM, through various frameworks, tools and approaches, creates the context and provides the shared language between these two important functions. TM is then the link between the business strategy and technological potential of a firm. This link should be leveraged to ensure that a firm establishes or maintains a competitive advantage over its rivals.

Based on this review, it seems as if specialised frameworks should be developed for each respective context. This is based on the fact that there is still no single all-encompassing framework and that most of the articles included in this literature review created its own framework for its specific context. The Dynamic Capabilities Theory seems particularly adept at continually combining and aligning all the dynamic aspects of TM into one underlying principle. This makes it an ideal principle on which to base specialised frameworks for various different contexts.

The resulting TM framework should assist with strategic decision-making and support communication between the various stake-holders in the firm. Most frameworks have either a graphical output or model, which might assist the decision-making process and communication.

Various tools are available to assist with the development and/or implementation of a TM framework. Based on this review, however, the most common type of tools incorporated are the matrix-based tools. These tools are especially appealing due to their simplicity and adaptability to different contexts and situations.

The most practical way in which to obtain data to develop the required TM framework seems to be through workshops and interviews. Due to the time-intensiveness of workshops, interviews are often a more practical solution to obtain the required information from the different role-players.

Further research is still required to determine how a practical TM framework, as is described in this article, can be developed. In particular the manner in which information is obtained from different role-players in a real-world situation should be investigated. In addition, how this information would be developed into a workable and useful TM framework, which will add value to a specific firm.

Although there have been a few studies, as described above, there have only been limited studies relating to the practical implementation of TM frameworks in developing countries, diversified firms and SMEs in general.

As it seems worthwhile to research the practical implementation of TM frameworks in the South African context, especially with regard to diversified firms, a study of how a local SME implements TM is underway based on the literature review detailed in this paper. This implementation can be valuable to grow the industrial capacity of not just individual firms, but the country as a whole.

7 REFERENCES

[1] J. T. M. Van Der Zee and B. De Jong, "Alignment Is Not Enough: Integrating Business and Information Technology Management with the Balanced Business Scorecard," *J. Manag. Inf. Syst.*, vol. 16, no. 2, pp. 137–158, 1999.

[2] M. J. Gregory, "Technology management: a process approach," *Proc. Inst. Mech. Eng. Part B J. Eng. Manuf.*, vol. 209, no. 5, pp. 347–356, 1995.

[3] D. Cetindamar, R. Phaal, and D. Probert, "Understanding technology management as a dynamic capability: A framework for technology management activities," *Technovation*, vol. 29, no. 4, pp. 237–246, 2009.

[4] R. Phaal, C. J. P. Farrukh, and D. R. Probert, "Technology management tools: Concept, development and application," *Technovation*, vol. 26, no. 3, pp. 336–344, 2006.

[5] M. Arasti, M. Khaleghi, and J. Noori, "Corporate-level technology strategy and its linkage with corporate strategy in multi-business companies: IKCO case study," *Technol. Forecast. Soc. Change*, vol. 122, pp. 243–252, 2017.

[6] A. Ghezzi, R. Balocco, and A. Rangone, "A fuzzy framework assessing corporate resource management for the mobile content industry," *Technol. Forecast. Soc. Change*, vol. 96, pp. 153–172, 2015.

[7] Z. Husain, "Technology strategy framework for firms in growing economies," *J. Glob. Bus. Adv.*, vol. 9, no. 3, pp. 248–274, 2016.

[8] G. Schuh and L. Kramer, "Cybernetic Approach for Controlling Technology Management Activities," in *Procedia CIRP*, 2016, vol. 41, pp. 437–442.

[9] D. Avison, J. Jones, P. Powell, and D. Wilson, "Using and validating the strategic alignment model," *J. Strateg. Inf. Syst.*, vol. 13, no. 3, pp. 223–246, 2004.

[10] J. M. Burn and C. Szeto, "A comparison of the views of business and IT management on success factors for strategic alignment," *Inf. Manag.*, vol. 37, no. 4, pp. 197–216, 2000.

[11] T. Tambo, J. Bargholz, and L. Yde, “Evaluation of TOGAF as a Management of Technology Framework,” in *IAMOT 2016 - 25th International Association for Management of Technology Conference, Proceedings: Technology - Future Thinking*, 2016, pp. 833--849.

[12] R. Ansari, J. Soltanzadeh, and A. Tavassoli, “Technology selection between technology management and decision making: a case study from the Iranian automotive industry,” *Int. J. Automot. Technol. Manag.*, vol. 16, no. 4, pp. 365–388, 2016.

[13] G. De Wet, “Emerging from the technology colony: a view from the south,” *PICMET ’99 Portl. Int. Conf. Manag. Eng. Technol. Proc. Vol-1 B. Summ. (IEEE Cat. No.99CH36310)*, vol. 1, 1999.

[14] G. Schuh and S. Ryschka, “Strategy framework for technology platforms within the context of diversified companies,” in *IEEE International Conference on Industrial Engineering and Engineering Management*, 2016, vol. January, pp. 1514–1519.

[15] G. Schuh, S. Ryschka, A. Rollmann, and T. Wetterney, “Analyzing the integration of technology platform strategies into the situational context of diversified companies,” in *IAMOT 2016 - 25th International Association for Management of Technology Conference, Proceedings: Technology - Future Thinking*, 2016, pp. 44–62.

[16] D. M. Gann and A. J. Salter, “Innovation in project-based, service-enhanced firms: the construction of complex products and systems,” *Res. Policy*, vol. 29, no. 7–8, pp. 955–972, 2000.

[17] R. McAdam, U. Bititci, and B. Galbraith, “Technology alignment and business strategy: a performance measurement and Dynamic Capability perspective,” *Int. J. Prod. Res.*, vol. 55, no. 23, pp. 7168–7186, 2017.

[18] P. L. Drnevich and D. C. Croson, “Information technology and business-level strategy: Toward an integrated theoretical perspective,” *MIS Q.*, vol. 37, no. 2, pp. 483–509, 2013.

[19] L. Argote and M. Hora, “Organizational Learning and Management of Technology,” *Prod. Oper. Manag.*, vol. 26, no. 4, pp. 579–590, 2017.

[20] M. McKelvey, “Firms navigating through innovation spaces: a conceptualization of how firms search and perceive technological, market and productive opportunities globally,” *J. Evol. Econ.*, vol. 26, no. 4, pp. 785–802, 2016.

[21] A. Brem and K.-I. Voigt, “Integration of market pull and technology push in the corporate front end and innovation management-Insights from the German software industry,” *Technovation*, vol. 29, no. 5, pp. 351–367, 2009.

8 BIOGRAPHY

Christo Visser is employed as a project manager at Alphawave Mobile Network Services, Stellenbosch, South Africa. He is currently enrolled as a Master’s student in Engineering Management at the University of Stellenbosch, South Africa.

Konrad von Leipzig obtained both his B.Eng and M.Eng degrees in Industrial Engineering from Stellenbosch University. He later also completed a B.Comm degree. He has been a lecturer at the Department of Industrial Engineering at Stellenbosch University since 1987. He forms part of the team undertaking the initialisation and realisation of the Learning Factory in Stellenbosch. His research areas include process management, supply chain management, logistics and financial management.

Prof Sara Grobbelaar is employed as an Associate Professor at the Department of Industrial Engineering at Stellenbosch University. Her research areas include innovation systems, innovation ecosystems, innovation platforms, technology management and inclusive development.

Methods and Tools for SMEs to Support Digital Transformation in Production and Logistics

J. Bauer[1], V. Hummel[1]
[1]ESB Business School, Reutlingen University, Germany

Abstract

Changing requirements and qualification profiles of employees [1], increasingly complex digital systems up to artificial intelligence, missing standards for the seamless embedding of existing resources [2] and unpredictable return on investments [3] are just a few examples of the challenges of an SME in the age of digitalisation. In most cases there is a lack of suitable tools and methods to support companies in the digital transformation process in the value creation processes, but also of training and learning materials. A European research project (BITTMAS - Business Transformation towards Digitalisation and Smart systems, ERASMUS+, 2016-1-DE02-KA202-003437) with international partners from science, associations and industry has addressed this issue and developed various methods and instruments to support SMEs. Within the scope of a literature search, 16 suitable digitalisation concepts for production and logistics were identified. In the following, a learning platform with a literature database with multivariable sorting options according to branches and keywords of digitalisation, a video gallery with basic and advanced knowledge and a glossary were created in order to provide the user with consolidated and structured specialist knowledge. The 16 identifying concepts for transforming value-added processes in the context of digitalisation were transferred to a learning platform using developed learning paths in coaching and training to online course modules including test questions. A maturity model was developed and implemented in a self-assessment tool for the analysis to identify the potential of digitalisation in production and logistics in relation to the current technological digitalisation level of the company. As a result, the user receives one or more of the 16 potential digitalisation concepts suggested or the delta for the necessary, not yet available enabler technologies is presented as a spider diagram. For a successful implementation of the identified suitable digitalisation concepts in production and logistics, a further tool was developed to identify supplementary requirements for all company divisions and stakeholders in relation to the "digital transformation" in the form of a self-evaluation. This paper presents the methods and tools developed, the accompanying learning materials and the learning platform.

Keywords
digital transformation, enabler technologies for digitalisation, digitalisation potential

1 INTRODUCTION

On March 10th 2016, the Strategic Policy Forum on Digital Entrepreneurship stated very clearly - digital transformation is not an option – it is a must! [4] By 2030, global expenses of 15 trillion USD are forecasted in the area of industrial networking and digitalisation [5] [6] - eventually more than 80 % of the industrial companies in major industrial countries consider increased use of data along the value chain as a top priority and see opportunity for growth through both optimized processes as well as design and implementation of new business models [7].

Emerging technologies, the increasing availability of information and communication technology (ICT) and the simultaneous diffusion into industrial application, with the visionary final state of continuously and fully networked and virtualized entities inside and outside industrial complexes, along an holistic value chain within an "Internet of Things", commonly referred to as the "digital revolution", show potentials with regard to improved efficiency, raised productivity as well as additional profits from new business models [8].

However, to create value through the "Internet of Things" and seize the opportunity for a technological change the commitment of each enterprise to identify and quantify its individual potential is required [9]. The new approaches of digitalisation and smart systems are often barely validated and potential impact is mostly not proven. Transformation towards digitized, smart systems is not a one-off activity but rather an ongoing process and could be describe as a lifelong learning process for an enterprise.

In comparison to big companies with the financial strength to either hire consulting companies or to run own departments with a number of experts working on digitalisation and smart systems, especially SMEs do not have the same financial possibilities and therefore are facing a lack of expertise, overview, knowledge with regard to technologies, methods, approaches their mutual interaction as well as potential and risks [10] [11]. SMEs run the risk of waiting too long with the migration towards smart systems and their digitalisation, or pursue no holistic approach for operational improvements or possible impacts on their business model [12].

2 SPECIFICATIONS OF THE BITTMAS PROJECT

2.1 Target group & end-users

Europe 2020 as a strategy intends to focus the European Union (EU) and its member states on the important task of improving the EU's competitiveness. According to the strategy, its goal is to transform the EU into "a smart, sustainable and inclusive economy, delivering high levels of employment, productivity and social cohesion." At the heart of competitiveness is the level of productivity of an economy. According to the executive summary of "The Europe 2020 competitiveness report", the EU continues to under-perform in comparison to the United States and other advanced economies in terms of building a smart, innovation based, knowledge-driven economy [13]. More starkly, the EU is increasingly falling behind globally in building the digital infrastructure and innovative capacity that would allow its economies to unlock new sources of growth [14].

The European SME needs to explore the future potential of ICT, automation, sustainable and clean as well as human-centred work systems and processing technologies. According to Gunther Oettinger, EU Commissioner for Digital Economy and Society, within the High-level conference on the digital transformation of industries and enterprises on February 16th, 2016, European business leaders' areas like big data, the Internet of Things, cyber-physical systems and robotics offer great opportunities for industry. At a recent roundtable in Brussels on the digital transformation of Industry, ways to energize the digital transformation of Europe's industrial sector across Europe were discussed, to raise the potential for increasing flexibility, efficiency, productivity, competitiveness – all helping to create jobs and growth trans-nationally - especially within SMEs.

After all, Europe is a continent of SMEs — where nine out of ten companies are SMEs and two out of three jobs are in SMEs [15]. They have to be part of the digital journey; they are crucial to Europe´s growth and competitiveness [16]. Together, sectors (C) Manufacturing 9,33%; (H) Transport and Storage 5,09%; (J) Information and Communication 4,44% and (M) Professionals Science and Technical Administration 18,13% comprised 37% of the 22,3 million SME in 2016 [17]. Most of the 8,25 million SME are not yet aware of the implications of being not prepared for this business transformation in order to emphasize on improvements, growth and competitiveness [18] [19]. Increased sense of initiative and entrepreneurship of owners, managers and staff can only be expected if sensitization and tailored measures are applied within these companies.

2.2 Scientific objectives and project goals

For exactly this purpose, to improve the achievements in the relevant and high-level basics and transversal competence of "business transformation toward digitalisation and smart systems", BITTMAS is working on a solution to provide information, develop methods and solutions to gain knowledge and expertise that is required to assess existing technologies, processes and structures regarding improvement potentials and develop new business models, derive a road map with corresponding measures and implement an improvement process for trainers of VET organizations as well as staff and management of SMEs.

Major aim of BITTMAS is a user (VET trainers, staff and managers of SME) orientated approach guided by self-information, self-learning via web-based training modules, self-assessment of the maturity grade of the SME, self-identification of existing potentials, self-derivation of a road map and required measures as well by self-planning and implementation of the improvements. The BITTMAS Release Candidate consisting of the learning platform, coaching and training modules, a self-assessment to analyse digitalisation potentials and a self-assessment to consider further requirements focused on nine criteria based on the EFQM Model, improve and extend the range of high quality learning offerings for adults and strengthen the key competences in VET curricula. Digitalisation and smart systems do not only appear in the industrial and economic environment but also in the daily social life and the respectively required competencies can therefore be referred to as transversal.

For easing this business transformation and to keep pace with the rapid developments the BITTMAS supports training and self-driven processes for the required business transformation through employees and managers of SME by providing:

- a literature library, video gallery and glossary to inform and sensitize the users with regard to digitalisation and smart systems.
- a self-assessment to determine the potentials for process enhancements and new business solutions based on a tailored maturity model.
- online coaching and training modules to learn or to enhance knowledge regarding to 16 digitalisation concepts.
- a self-assessment to determine further supporting requirements for digital transformation.
- a free entry version of the developed BITTMAS Release Candidate.

2.3 Project consortium

The international consortium consists of an appropriate number of partners who contribute to the related work packages and support the successful development of BITTMAS according to their competences. One German University of applied Science (Reutlingen University, project leader), one University from Turkey (Sabanci University), two research institutes from Austria (Fraunhofer Austria Research GmbH, Vienna) or Romania (IPA SA, Craiova), one German software developer company (IBK - Management Solutions GmbH, Wiesbaden), a consulting company from Germany (IIC - International Industrial Consult AG, Frankfurt), an association of Electronics and Information Technology Industries from Spain (GAIA, Bilbao), a Chamber of Commerce from Ireland (Waterford Chamber of Commerce) and one University from Kosovo (University for Business and Technology, Pristina) as an associated partner, have taken up the challenge to identify digitalisation potentials and concepts, to structure them and to develop suitable tools for sensitization as well as to support the analysis of SME's own potentials for digitalisation.

3 BITTMAS METHODS AND TOOLS

3.1 Literature research and derivation

The implementation of the BITTMAS project started with an extensive literature research, on the one hand to collect, derive and define information and learning materials in digitalisation and smart systems on the other hand to sensitize each partner to the subject of digitalisation itself. The collection of documents reflected the state of the art in industry and the scientific discussions about Digitalisation and Industrie 4.0 concepts. The main aim was on studies, articles and standard references, use cases and best practice examples. The quantitative and qualitative analysis (including the use of the MAXQDA analysis tool from Verbi GmbH and expert interviews) of over 170 literature and 32 video sources, which the BITTMAS partners were able to identify, resulted in a ranking list of relevant topics for the BITTMAS target group. Based on the frequencies of the key words of the researched literature, the subsequent discussions within the BITTMAS consortium and the feedback of external experts, the following 16 core topics/concepts could be derived within the framework of digitalisation and smart systems: (1) Digital twin of the facility, (2) Digital twin of the assets, (3) Smart Factory, (4) Decentralised & Flexible value creation, (5) Digital Assistant systems, (6) Technical Assistant systems, (7) Integrated transparent value chain, (8) Autonomous intralogistics, (9) Predictive maintenance, (10) Data driven decision making, (11) Data-enabled resource optimisation, (12) Track and trace, (13) Smart Product, (14) Digital Human Resource Management, (15) Digital Marketing, (16) Digital Procurement.

3.2 Learning platform

Adequate Learning Management Systems were researched and resulted in 8 of 15 software programs, which were compared by functions and costs. Main focus was on an open source solution with a wide active community for continuous free support and sustainable development as well as a high level of user friendliness (e.g. navigation, upload/download of material, easy allocation of roles/permissions, different menu languages, etc.) and functionality (e.g. closed/public area, glossary module, search options, personal desktop). Regarding all requirements for BITTMAS project the comparison of the Learning Management Systems resulted in two favourites: ILIAS and Moodle. Since certain optional peripherals and support from third parties can cost money regarding the use of Moodle and there is already an expertise in using ILIAS at Reutlingen University (host of the BITTMAS platform), project partners decided to choose ILIAS as BITTMAS Learning Management System.

3.3 Information and training material

3.3.1 Literature library

To provide the BITTMAS users with topic specific up-to-date literature sources, researched and collected documents were categorised into the following three divisions and were implemented at the learning platform as a data table including filter option for each criteria:

- Application Guides: Technologies, Business models, Available resources (funding etc.), Good practice
- Studies: White papers / Standards, Strategic papers, Applied statements
- Sectors: Automotive, Logistics, Health care, Electronics, Agriculture, Maintenance, Know-ledge, Economy, IT, Finance

It was originally planned to provide free available documents as downloads, but copyright reasons prevented the implementation. As a consequence a web link, which refers to the original source of the respective publication, was added to the library.

3.3.2 Video galery

Additional to the identified documents in the different areas the partners decided to make use of other sources of information which are available and complement the SME specific approach. All partners were looking for videos which are useful to pass on basic or advanced knowledge concerning Industry 4.0. and Internet of Things aiming for an improved understanding of digitalisation processes. 32 videos were categorised in basic and advanced knowledge and were implemented in the BITTMAS learning platform.

3.3.3 Glossary

A glossary of terms with definitions in the frame of digitalisation and smart systems was agreed upon the partners as a by-product to contribute to a common understanding by developing definitions. The glossary includes so far 40 definitions of terms in the fields of Digitalisation and Industrie 4.0.

3.3.4 Coaching and training modules

Once the user has identified a suitable concept, he can acquire, refresh, expand and test knowledge about the concept with the help of the 16 coaching and training modules provided by the learning platform. In order to ensure an optimal and individual knowledge transfer to the potential BITTMAS users regarding to the 16 digitalisation concepts, the partners designed a tailored user-oriented learning path. Starting with an introduction including definitions and demonstration of the benefits or advantages and disadvantages, the user should gain a rough understanding of the terms and areas of application of the respective digitalisation concept. Furthermore, a meaningful use case and the corresponding respective enablers are illustrated, so that the user gets a first impression regarding to the scope, the effort, the chances and challenges of an exemplary implementation of the digitalisation concept. Additionally use cases, road maps, references and literature sources as well as a final test to review and monitor the learning progress round off the coaching and training modules.

4 SELF ASSESSMENT DIGIPOTAS

BITTMAS solution to determine potential for digitalisation and smart systems of SME is based on a Maturity Model that led to a self-assessment tool named “DigiPotAS” (= Digital Potential Assessment).

4.1.1 BITTMAS Enabler

The basic idea of the BITTMAS potential self-assessment is based on a comparison of the technological requirements (status quo) of a company with the "enabler" technologies still required to enable the application of innovative concepts of digitalisation, which subsequently leads to a recommendation for action. In order to determine the influencing technologies (enabler) of a digitalisation potential, a common definition of each individual digitalisation concept must exist. In this respect, BITTMAS consortium has developed a description for each concept, which indicates the necessary technologies for implementation.

1) Digital twin of the factory: The digital twin of the factory helps to plan, design and construct the factory building and infrastructure.
2) Digital twin of the production asset: A digital twin of one or more production assets is used for design, virtual start-up and ongoing operation.
3) Smart Factory: Smart Factory refers to a production environment that ideally organizes itself without human intervention. This includes production facilities and logistics systems. Core components are cyber-physical systems and intelligent networking based on the Internet of Things (IoT).
4) Decentralised & Flexible value creation: Use of flexible, modular production assets instead of traditional production lines.
5) Digital assistant systems: Visualisation and automation of factory processes, for example with mobile applications (apps) combined with virtual and augmented reality solutions like tablets or digital glasses.
6) Technical assistant systems: New technical assistance systems like exoskeleton suits can be used to aid workers on the assembly line or warehousing.
7) Integrated transparent value chain: Integrated planning and scheduling systems within the factory from machine level over Manufacturing Execution System (MES) to Enterprise Resource Planning (ERP) systems.
8) Autonomous intralogistics: Factory systems capable of operating and performing logistics activities without human intervention.
9) Predictive maintenance: Remote monitoring of dynamic condition of machines with help of sensor data and big data analytics to predict maintenance and repair situations.
10) Data driven process/quality optimisation: Big data analyses can help to detect patterns in production or quality data and provide insights to optimise processes or product quality.
11) Data-enabled resource optimisation: Data-enabled optimisation of energy and resource consumption through intelligent data analyses and controls, e.g. energy or pressurised air management in facilities based on actual demand and/or supply.
12) Track and trace: Location of products and raw material within the factory is tracked via sensors and integrated into a data platform connected to internal systems such as MES/SES or ERP systems.
13) Smart Product: Smart products are products that are capable to do computations, store data, communicate and interact with their environment.
14) Digital Human Resource Management: Human Resource Management is strongly influenced by increasing flexible production systems as well as by new services offered.
15) Digital Marketing: Improving sales by the use of new technologies and concepts for example in the area of Marketing Automation.
16) Digital Procurement: Use of Internet technology for facilitating operative procurement processes, such as ordering, as well as sourcing tasks, e. g. web-based supplier search or eAuctions.

After the 16 potential concepts were determined, it was necessary to define suitable enablers. An extensive research and coordination approach with all partners led to the following 10 enabler technologies as requirement for the application of the identified concepts: (1) Additive Manufacturing, (2) Big Data & Analytics, (3) Cloud / Edge computing, (4) Cyber-physical system (CPS), (5) Cyber Security, (6) Product & Machine data, (7) Internet of Things (IoT), (8) Horizontal / Vertical system integration, (9) External data, e.g. evironmental, economic data, (10) Augmented Reality / Virtual Reality.

Next step was to research/determine the correlation of the identified 10 enablers to the identified 16 digitalisation concepts.

4.1.2 BITTMAS Maturity Model

The BITTMAS Maturity Model serves to make a holistic statement about the degree of progress of the digitalisation process related to a company or an organization on the basis of a self-assessment in different categories. To determine the maturity levels and based on the 10 enablers, the project partners have developed 42 criteria, e.g. “level of implementation of operational / environmental data & sensors”, that can be assigned to the categories Enabling Technologies, Data & Information, Value Creation Process, Products & Services and Customers & Partners. The selected individual categories of the BITTMAS Maturity Model are derived from the Industry 4.0 Maturity Model by [20] and oriented on the one hand to the question of the existence of various technological and process-related prerequisites for penetrating digitalisation in the company and on the other hand to their comprehensive use in the areas of the company relevant to sales and thus success - namely value creation, product and service design and customer and supplier communication.

- Specifically, the Enabling Technology category examines the presence and coverage of basic technologies in the company, including identification and communication technologies, IT infrastructure and end devices.
- The category Data & Information queries the digitalisation status of company information and the use of data for company management, in particular corporate decision-making.
- The Value Creation Process examines the degree of implementation of value-added digitalisation applications, including automation, remote control and machine-to-machine communication.
- The Products & Services category deals with the use of digitalisation approaches by the product or services offered by the company.
- The Customers & Partners category essentially asks whether and to what extent the integration of external stakeholders takes place using digitalisation.

4.1.3 BITTMAS Maturity Levels and Evaluation

Each criterion, e.g. “the level of implementation of operational / environmental data & sensors”, has 4 possible criteria values, whereby the respective applicable one is determined by the evaluating company. Each of the four criteria values represents one of the four delinearized BITTMAS maturity levels (0-3). All questions, answers as well as the logic, defined by the correlation matrix and the maturity model, were transferred into a self-developed online tool, which was developed and optimized from the beginning for the usage with mobile devices (tablets, mobile phone) in addition to a conventional use of a desktop or notebook. An evaluation of all 42 criteria leads to the determination of the digitalisation maturity level of the organization in each of the individual digitalisation-relevant categories and also allows the determination of the overall maturity level, represented by spider web diagrams.

5 SELF ASSESSMENT DIGISURA

If the BITTMAS user decides to implement one of the concepts, he can use the additional BITTMAS self-assessment Digital Transformation Support Requirements Assessment (=DigiSuRA) to determine further supporting requirements for digital transformation. The assessment of specific support needs for the digitalisation transformation of the given organisation is a self-assessment designed along the Criteria and Part-Criteria (Enablers: Leadership, People, Strategy, Partnerships & Resources, Process, Products & Services, Results: People Results, Customer Results, Society Results, Business Results) of the European Excellence Model, as propagated by CAF and EFQM [21]. It uses a strongly reduced set of questions compared to the original Model and allows a fast overview of the organisations current position with a focus on those areas and activities that are important for or involved in the upcoming digital transformation projects. Identified areas with support requirements can be supplemented with support project/process suggestions and these then subsequently priorities for implementation and expanded into project assignments for realisation by the designed project.

6 SUMMARY AND OUTLOOK

The BITTMAS tools presented within this publication enables the user:

- to determine a suitable digitalisation concept by carrying out a self-assessment for potential analysis (BITTMAS self-assessment DigiPotAS).
- to generate knowledge on digitalisation, especially in relation to SMEs, via the BITTMAS learning platform and the literature library, video gallery and glossary provided with it (BITTMAS learning platform).
- to learn 16 digitalisation concepts via trainings and coaching modules with definitions, benefits, enabler description, example uses cases and review or test (BITTMAS coaching and training modules).

- to self-assess and determine further requirements for digital transformation at company or process level (BITTMAS self-assessment DigiSuRA).

The three-year project BITTMAS will be completed in August 2019. After the creation, implementation and fine-tuning of the various BITTMAS tools, further projects steps are focused on pilot user support as well as the validation and fine-tuning aiming the final BITTMAS release candidate and conducted by dissemination activities.

7 REFERENCES

[1] Hammermann, A.; Stettes, O., Qualifikations-bedarf und Qualifizierung: Anforderungen im Zeichen der Digitalisierung, 2016, IW policy paper, No. 3/2016, Institut der deutschen Wirtschaft, Köln

[2] Lag, S.; Brathagen With, S., Standardisation as enabler of digitalisation, 2017, position paper 2017, Group Technology & Research

[3] Geissbauer, R.; Schrauf, S.; Berttram, P.; Cheraghi, F.; Digital Factories 2020: Shaping the future of manufacturing, Pricewaterhouse Coopers GmbH, 2017, Page 6

[4] Strategic Policy Forum on Digital Entrepreneurship: Digital transformation is not an option – it is a must, 2016, https://ec.europa.eu, last accessed September 16th, 2018

[5] GTAI Germany Trade & Invest, Fact Sheet - The Digital Economy in Germany - Issue 2018/2019, 2018, supported by the Federal Ministry for Economic Affairs and Energy

[6] Peter C. Evans P. C.; Annunziata, M., "General Electric: Industrial Internet, Pushing the Boundaries of Minds and Machines, 2012

[7] General Electric Company, Industrial Internet Insights Report for 2015, 2014

[8] Becker W.; Ulrich P.; Botzkowski T.; Eurich S., Digitalisierung von Geschäftsmodellen, 2017, In: Schallmo D.; et.al., Digitale Transformation von Geschäftsmodellen, Springer Gabler, Wies-baden, 978-3-658-12387-1

[9] Deloitte AG, Study: Industry 4.0 – Challenges and solutions for digital transformation and use of exponential technologies, 2015, Deloitte AG, Zürich, Page 25

[10] Bley, K.; Leyh, C.; Schäffer, T., Digitization of German Enterprises in the Production Sector – Do they know how "digitized" they are?, Twenty-second Americas Conference on Information Systems, San Diego, 2016

[11] Bley, K.; Leyh, C.; Digitalisierung: Chance oder Risiko für den deutschen Mittelstand? – Eine Studie ausgewählter Unternehmen, 2016, HMD Praxis der Wirtschaftsinformatik

[12] Demary, V.; Engels, B.; Röhl, K.-H.; Rusche, C. , 2016, Digitalisierung und Mittelstand: Eine Metastudie, IW-Analysen, Köln

[13] World Economic Forum, The Europe 2020 Competitiveness Report, 2014 Edition

[14] European Round Tables Industrials, Benchmarking 2017 - European Competiveness and Industry, 2017

[15] Muller, P.; Julius, J.; Herr, D.; Annual report on European SMEs 2016/2017, 2017, European Commission, ISBN 978-92-79-74126-5

[16] Gros, D.; Roth, F., The Europe 2020 Strategy – can it maintain the EU`s competiveness in the world?, 2012, Centre for European Policy Studies, Brussels, Page 1

[17] Statista, Number of small and medium-sized enterprises the European Union in 2016, www. statista.com, last accessed March 16th, 2018

[18] McKinsey Global Institute, Digital Europe: Pushing the frontier, capturing the benefits, 2016, McKinsey & Company, Page 11

[19] Zimmermann, V., SMEs and digitalisation: The current position, recent developments and challenges, 2016, KfW Research, No. 138

[20] Schumacher, A.; Erol, S.; Sihn, W, A Maturity Model for Assessing Industry 4.0 Readiness and Maturity of Manufacturing Enterprises, 2016, Procedia CIRP. 52. 161-166

[21] EFQM, Das EFQM Excellence Modell 2013, www.efqm.de/efqm-modell.html, last accessed September 16th, 2018

8 BIOGRAPHY

Vera Hummel, Prof. Dr.-Ing., Dipl.-Ing., has been a professor at the ESB Business School, Reutlingen University since 2010. She is a founding member of the "Initiative of European Learning Factories" and the initiator of the "ESB Logistics Learning Factory" for education and industry training on the campus of Reutlingen University.

Jörg Bauer, Dipl.-Kfm., is a research associate at ESB Business School and part of the research, training and consultant group of ESB Logistics Learning Factory. His research focuses on learning management and assistance systems in the areas of production and logistics.

Methodology for an Integrative Manufacturing Change Management in Technology Planning for Medical Products

L. Hermann[1], J. Rey[1], T. Bergs[1]

[1]Laboratory for Machine Tools and Production Engineering (WZL) of RWTH Aachen University, Germany

Abstract

Growing population and demographic change lead to an increasing demand for medical products. In addition, ever shorter product life cycles and emerging legal requirements call for frequent product changes and process adaptions. The combination of these trends results in major challenges for technology planners in medical industry, due to the required cost and time consuming recertification of products and processes after conducted changes and adaptions. To address these challenges, a systematic methodology to derive, analyse and evaluate process adaption alternatives and their change propagation for manufacturing and inspection sequences in medical industry is presented. The methodology is validated using a case study from medical industry.

Keywords

Manufacturing Change Management, Technology Planning, Medical Technology

1 INTRODUCTION

Product changes and process adaptions pose a special challenge in medical industry, due to mandatory time and cost consuming recertifications [1]. Therefore, the identification, analysis and selection of process adaptions is of particular importance. The demand to solve the stated challenge arises from the strong growth of medical industry. This growth is exemplarily represented by the worldwide medical technology sales, which grew from $308 billion in 2009 to $403 billion in 2017 by 30 % [2]. Based on the complexity of the challenge and the strong demand to solve it, effective and efficient dealing with product changes and process adaptions in manufacturing of medical products is a topic of high relevance for today's research and industry.

An integrative methodology to identify, analyse and evaluate process adaption alternatives for manufacturing and inspection sequences (MIS) in medical industry is presented in the following. The core elements are the identification of adaption alternatives, the analysis of change propagation within the existing MIS and the economic evaluation of the identified adaption alternatives, to generate a basis for decision making. The integrative character of the methodology is derived from the combined study of manufacturing and inspection processes and their interdependencies. This is done to exploit cost saving potentials originating from the high ration of inspection costs on the overall production costs, which is another particularity of medical industry [3]. Finally, the methodology's applicability and usefulness is validated using a case study from medical industry. In the following, the expression change refers to products while adaption refers to processes if alterations are made.

2 STATE OF THE ART

2.1 Terms and scope of the methodology

The term Manufacturing Change Management (MCM) was addressed for the first time by ProSTEP iViP e. V. in 2014 [4]. MCM was derived from the more established Engineering Change Management (ECM), for which a variety of scientific publications already exists. The state of the art for ECM has been summarized by Ullah et al. [5]. The difference between ECM and MCM is the object of change. While ECM focuses on alterations made to a product, MCM focuses on alterations made to a factory [6]. MCM is a relatively new field of research that has not yet received much attention. However, the increasing number of scientific publications underlines the relevance of this field of research. In some publications, the terms ECM and MCM are not clearly delimited, which means that both subjects must be considered in order to examine the existing literature holistically [7]. This paper is based on the definition by Koch, who states "*Manufacturing Change Management refers to organizing and controlling the process of making alterations to a factory*" [6]. To give a concise but scientifically sufficient review on the existing literature, the following state of the art describes different types of approaches in ECM and MCM in general and their drawbacks to solve stated challenge. Subsequent, a more detailed description is given for approaches being closer to the scope of the presented methodology.

2.2 Existing models and methodologies

There is a variety of models and methodologies describing the superficial dealings with changes in production. Partially they do not only focus on manufacturing and inspection, but also on organizational and socio-technical aspects. One example for this is the Generic Change Process according to Jarratt et al. The Generic Change Process consists of six steps, starting from a change trigger and completing with a review of the conducted change process [8]. Another approach to handle changes in production was presented by Azab et al. The approach does not contain a stepwise procedure and draws attention to economic changes like fluctuating batch sizes instead of product changes [9]. Although these methodologies and approaches offer guidance coping with changes in production, their level of detail is too low to be used in medical industry. Adaption alternatives are not specified enough and therefore their impact on the product's characteristics is not part of the investigation, leading to a lack of information for the required certification of processes. More superficial models and methodologies lacking a high level of detail can be found in the prevailing literature review in [7].

In addition, more detailed approaches exist, focusing on the adaption of separate processes in production. Karl and Reinhart presented a methodology to reconfigure manufacturing resources in general, including a software supported algorithm to generate alternative reconfigurations [10]. Hoang et al. developed a similar approach to generate adaption options for machine tools, focusing on the interdependencies between products, processes and resources as well as on the impact analysis for adaptions [11]. Another approach was developed by Marks et al. Containing the generation of adaption options as well, the approach also consists of a producibility check for a requested change, which is based on a skill model of the considered production system [12]. Recent scientific models and methodologies to support technology planners in adapting separate manufacturing resources offer promising insights and are used in the presented methodology to adapt manufacturing and inspection sequences. Nonetheless, they cannot be used as a stand-alone solution for adapting MIS in medical industry due to the missing consideration of cross-process interactions. An investigation of single adaptions, without considering the adaptions' impact on the whole MIS does not enable a sufficient information acquisition on the affected product characteristics and their alterations to ensure the product's full functionality for certifications.

Beyond the presented approaches, there are further scientific publications dealing with MCM from different perspectives and levels of detail. The most relevant existing scientific publications in the field of MCM and the scope of the presented methodology are shown in figure 1. For their application to implement product changes through process adaptions in medical industry, these approaches partially lack in four fundamental requirements. Their degrees of fulfilment are depicted in figure 1 for each relevant approach. These requirements are a **sufficient level of detail** to support certifications by detecting all affected product characteristics and their alterations (**R1**), the **integrative design** of manufacturing and inspection (**R2**), the containment of a **holistic change impact and propagation analysis** (**R3**) as well as a fundamental **methodical procedure** (**R4**).

□□□□ Low fulfilment → High fulfilment ■■■■

Approach according to	Year	Ref.	R1	R2	R3	R4
Jacobs et al.	2018	[13]	■□□□	■□□□	□□□□	■■■□
Marks et al.	2018	[12]	■■■□	□□□□	■■□□	■■■□
Olmez et al.	2018	[14]	■■□□	■□□□	■■□□	■■■□
Bauer et al.	2017	[15]	■□□□	■□□□	■■□□	■■□□
Cichos	2017	[16]	■■□□	■■□□	■■□□	■■■□
Cichos et al.	2017	[17]	■■□□	■■□□	■□□□	■■■□
Hoang et al.	2017	[11]	■■■□	■□□□	■□□□	■■□□
Hoang et al.	2017	[18]	■■■□	■□□□	■□□□	■■□□
Koch	2017	[7]	■■□□	■□□□	■□□□	■■□□
Plehn	2017	[19]	■□□□	■□□□	■■■□	■■■□
Bruno	2016	[20]	■■□□	□□□□	■■□□	■□□□
Cichos et al.	2016	[21]	■■□□	□□□□	■■□□	■■■□
Cichos et al.	2016	[22]	■□□□	■□□□	□□□□	■■□□
Gernhardt et al.	2016	[23]	■□□□	■□□□	■□□□	■□□□
Koch et al.	2016	[7]	■■□□	□□□□	■□□□	□□□□
Koch et al.	2016	[24]	■□□□	□□□□	■□□□	■□□□
Cichos et al.	2015	[25]	■■□□	■□□□	■■□□	■■□□
Karl et al.	2015	[10]	■■□□	■□□□	■□□□	■■□□
Koch et al.	2014	[26]	■□□□	■□□□	■■□□	■□□□
Azab et al.	2013	[11]	■□□□	■□□□	□□□□	■■□□
Malak	2013	[27]	■■□□	■□□□	■■■□	■■■□

Figure 1 - Suitability of existing MCM approaches.

3 OBJECTIVE

Main objective of the conducted research is the development of a methodology to support technology planners in adapting manufacturing and inspection of medical products, so that required product changes can be implemented in a cost-optimized way. To reach this objective, the level of detail of the derived adaptions must be high enough to investigate their influence on the whole MIS and all functionally relevant characteristics of the product (R1 & R2). This level of detail and the analysis of impacts and propagations of adaptions is necessary to support a successful recertification of the MIS. Furthermore, the adaption of manufacturing and inspection has to be realized in an integrative manner to give respect to the high amount of inspection costs in medical technology and allow a cost optimized adaption (R3). Foundation for the conducted research was the development of an approach containing a systematic and methodical procedure to ensure the research findings are applicable for practical real world applications (R4). The methodology developed with respect to these objectives is presented in the following.

4 METHODOLOGY

The developed methodology is divided into five steps (cf. figure 2). In each step, models and methods are used that have been newly developed or adapted based on existing approaches. The steps as well as all used models and methods are described inside the following subsections. Due to the scope of this publication, the main purpose is to present the concept of the methodology holistically and not to give an in-depth description of every single step. Elaborated descriptions of all steps, models and methods will be content of future publications.

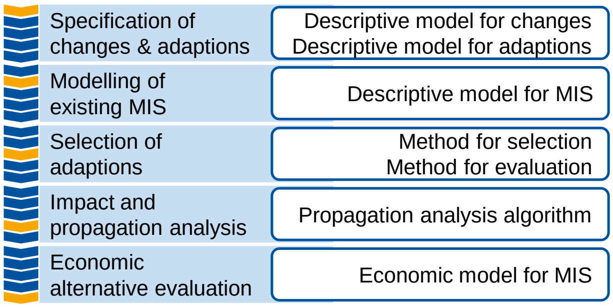


Figure 2 - Rough concept of the methodology.

4.1 Specification of changes and adaptions

Purpose of the first step is the specification of changes in terms of breaking down unprecise required product changes to quantifiable changes of characteristics. In addition, product changes and process adaptions that are prohibited are set. Therefore, a descriptive model for product changes and another one for process adaptions were developed. The descriptive model for product changes is structured hierarchically. In reference to KLOCKE ET AL., the first level is represented by the classes macro geometry, micro geometry and material properties [28]. These classes are comprised from subclasses. For example, the subclasses in the class macro geometry are size tolerance, shape tolerance, position tolerance and dimensions. These subclasses contain specific and quantifiable product characteristics like flatness or length. By assigning unprecise product changes to classes, then to subclasses and finally to characteristics, they are specified in a quantifiable manner, which is essential for further steps. For instance, an unprecise required product change like “smoother surface” is assigned to the class micro geometry, to the subclass roughness parameters of a surface and to the characteristic mean arithmetic height to become quantifiable. If a product change can be assigned to more than one class, subclass or characteristic, it is divided into separate product changes for further steps of the methodology. The functionality of the descriptive model for changes is depicted in figure 3.

Besides specifying product changes, it is possible to mark classes, subclasses or characteristics that are prohibited to change within the descriptive model for changes. These restrictions represent boundary conditions for the following steps. Restrictions are also set within the descriptive model for process adaptions. It is structured hierarchically as well, consisting of three classes, which contain several adaptions. The classes are control level, machine level and structure level. Within these levels, different adaptions like parameter change (control level), tool change (machine level) or process substitution (structure level) exist. Comparable to the descriptive model for changes, restrictions can be set hierarchically for processes (first level) for classes (second level) or single adaptions (third level).

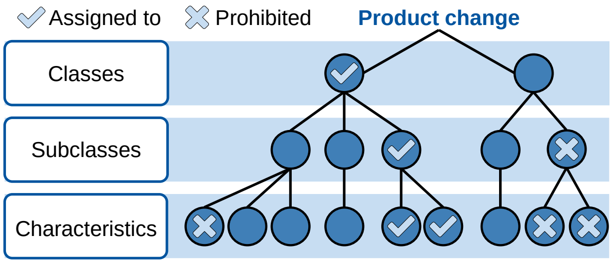


Figure 3 - Descriptive model for product changes.

4.2 Modeling of the existing MIS

To derive adaption alternatives and to investigate the impact and propagation of an adaption, the considered MIS and the existing interdependencies within it are modeled. To model complex systems and interdependencies, a Design Structure Matrix (DSM) was used in different approaches dealing with MCM (e.g. [11]). Since DSM has proven to be a suitable tool, it is adjusted and used for deriving adaption alternatives and analysing their impact and its propagation in the presented methodology as well.

To model a MIS through a DSM, every manufacturing process, inspection process and defined product characteristic is assigned to a column and a row of the DSM. Following, the interdependencies between all elements are entered into the matrix. If a row element influences a column element, the value 1 is entered into the corresponding field of the matrix. If there is no influence, the value 0 is entered. In addition, unintentional effects, like scratching surface while measuring its roughness by a tactile surface profilometer (TSP), have to be included as well by entering the value 1 into the field corresponding to the row “tactile surface profilometer” and the column “surface roughness”. An excerpt of a modeled MIS is shown in figure 4.

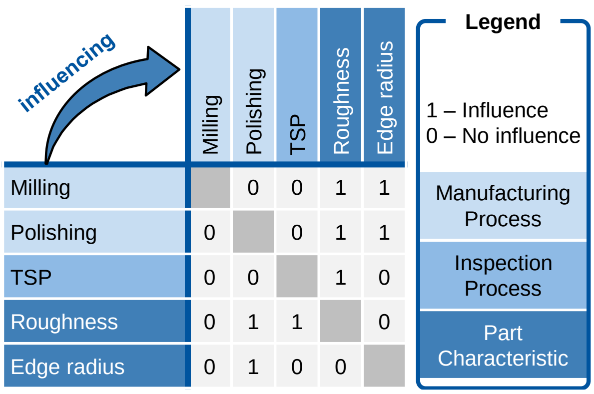


Figure 4 - MIS modelled through DSM.

Milling and polishing influence the characteristics roughness and edge radius, while tactile surface measurement only influences the roughness. The roughness influences the polishing process, because changing the roughness before the process leads to a changed roughness after the process, if no adaptions are made. The roughness also influences the tactile surface measurement due to possibly necessary adaptions of the inspection process after changing the surface roughness (e.g. using a different measuring tip). The given example is not universally valid for every MIS consisting of the depicted processes, but it illustrates how MIS are modeled through a DSM. The matrix's content is defined application-specific.

4.3 Selection of adaptions

In the third step of the methodology, the restrictions set in step one and the modeled MIS from step two are used to select adaptions, which are capable of implementing the specified product change. Therefore, all adaptions within the descriptive model are specified for concrete processes and examined regarding their technological capability to implement the required product change, as long as they are not subjection of a restriction.

To detect processes influencing the considered product characteristic, the MIS model's column of this characteristic is searched for ones in the first place. Referring to the example in figure 4, the surface roughness could be changed by adapting the polishing or the milling process. Following, the general adaptions from the descriptive model are specified for every identified process influencing the considered product characteristic. For the exemplary polishing process, this means general adaptions like parameter change or tool change are specified to increasing polishing time or using a different polish. These process specific adaptions are examined for their technological capability to change the characteristic expediently. For instance, adapting the feed rate of the milling process can be a possible way to reduce the surface roughness, while using a different lubricant coolant may not. The evaluation of the technological capability is always case specific for every MIS and has to be done by process experts and technology planners.

After identifying technologically feasible adaption alternatives, these alternatives are subject of a pre-evaluation. The pre-evaluation is an optional step of the methodology and is used to reduce the amount of adaption alternatives to investigate in further steps. A well suited method is the Analytic Hierarchy Process (AHP) according to SAATY [29]. Adaption alternatives are evaluated against multiple weighted criterions, hence a utility value for every alternative is calculated. By defining a threshold utility value, adaption alternatives are suspended for the following steps of the methodology.

4.4 Impact and propagation analysis

The impact and propagation analysis for adaptions marks the fourth step of the methodology. All previously selected adaption alternatives undergo a detailed investigation. This is done to get a holistic comprehension on the necessities to implement the required product change, without endangering the product's functionality or harming any restrictions. Besides planned changes of characteristics, unplanned changes caused by adaptions and the efforts to compensate them have to be considered.

For the impact and propagation analysis, an algorithm was developed, which investigates the consequences of an adaption based on the modeled MIS and expert knowledge (cf. figure 5). The algorithm consists of a complex if-else-interlacing, whereby the boolean-conditions for the if-statements depend on expert user input. Starting from the row of the considered adaption, the algorithm searches the MIS model's rows and columns for ones, to identify potential unintended characteristic changes and optional processes to counteract them. Since the value 1 within the model only describes a general impact, a user input was implemented into the algorithm, to determine if a characteristic changes through the specific considered adaption or not. If a change occurs, it is checked whether it can be accepted. If the change is not acceptable, another adaption is needed to compensate this change. For compensating adaptions, the algorithm is executed again. Due to the main purpose of this publication, to introduce the holistic concept and the functionality of the methodology, the algorithm is not explained in more detail. Further publications dealing only with the impact and propagation algorithm will follow.

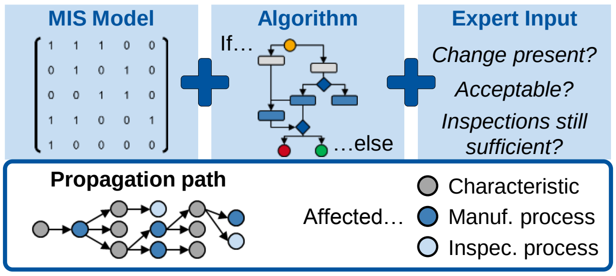


Figure 5 - Core elements of the analysis.

Outcome of the algorithm's application is a propagation path for every analysed adaption. The path describes every needed process adaption and characteristic change, to implement the initially required product change while ensuring the product's full functionality (cf. figure 5). This propagation path is an essential input for the economic evaluation in step five.

4.5 Economic alternative evaluation

The last step of the developed methodology is an economic evaluation of the analysed adaption alternatives. The evaluation serves as a basis for decision-making for choosing an alternative.

For the economic evaluation, not only the investment costs of adaption alternatives but also the changed production costs per product are subject of consideration. Within the presented methodology, the cost model according to KLOCKE ET AL. is used [3]. It takes both manufacturing and inspection costs into account and has a highly integrative character. Further, it is possible to extend the model by additional cost factors to adjust it to case specific applications.

The overall result of the methodology's application are generated, analysed and evaluated adaption alternatives to implement a required change of a medical product. Conclusively, technology planners are enabled to choose a cost optimized alternative while ensuring the product's full functionality and therefore supporting its recertification.

5 CASE STUDY

For the case study, the manufacturing of a heart assistant pump's lower case was investigated. The heart assistant pump is an implant to extend the natural functionality of a heart, to bridge the wait for a donor heart. As blood flows through the lower case, a high blood compatibility (hemocompatibility) is mandatory. Therefore, the required product change for the case study is smoothing the surface to improve the hemocompatibility.

In step one, the descriptive model was used to specify the required product change to an alteration of the product's characteristic arithmetic average roughness Ra from 0.1 µm to 0.08 µm. Further, changes of the macro geometry and adaptions to the structure and machine level were permitted. The considered MIS was modeled in step two. The MIS consisted of the manufacturing processes milling, drilling, lapping, polishing and inspections through a coordinate measuring machine, tactile surface profilometer, laser microscope as well as all relevant product characteristics. To reduce the roughness, the adaptions feed rate reduction for milling and machining time increase for polishing were selected in the third step. These adaptions were analysed for their impact and propagation in step four. For milling, no relevant unintended changes were found, whereas for adapting the polishing process, an unintended increase of an edge radius was identified, which had to be counteracted by adapting the NC-program of the milling process. In addition, the measurement cycle of the laser microscope had to be adapted, to measure the reduced Ra value with the same measurement accuracy as before. In step four, both alternatives were evaluated economically. The main cost driver was the increased manufacturing time for both alternatives. As a conclusive result, the adaption of the polishing process was found to be slightly cheaper than adapting the milling process. This difference in the cost-effectiveness represents the aimed basis for decision-making.

6 CONCLUSION

The presented methodology enables technology planners to identify, analyse and evaluate process adaptions to implement required product changes in a cost-efficient way and ensuring the product's full functionality. The first requirement for a high level of detail (R1) is realized through investigating product changes and process adaption on a product characteristic level. A holistic impact and change propagation analysis (R2) is put into place through the usage of the developed algorithm. A high level of integrativity between manufacturing and inspection (R3) is ensured by conducting manufacturing and inspection process equally within the impact and change propagation analysis as well as by using an integrative cost model for the economic evaluation. The last requirement for a methodical procedure (R4) is fulfilled through the stepwise and sequencing structure of the methodology. Conclusively, the case study based validation proofed the practical applicability of the methodology beyond artificial research conditions.

7 ACKNOWLEDGEMENTS

The authors would like to thank the German Research Foundation (Deutsche Forschungs-gemeinschaft - DFG) for the support of the depicted research within the project KL500/166-1 / SCHM1856/80-1 "Integration of technology and inspection planning for cost-optimized production processes of medical devices".

8 REFERENCES

[1] ISO 13485, 2016, International Organization for Standardization. Medical devices - Quality management systems - Requirements for regulatory purposes.

[2] Evaluate Ltd., 2017. Medtech World Report.

[3] Klocke, F. et al., 2017, Integrative technology and inspection planning - A case study in medical industry, Proceedings of the ASME 2017 12th International Manufacturing Science and Engineering Conference.

[4] ProSTEP ivip e. V., 2014, Manufacturing Change Management (Whitepaper): Management von Änderungen in der Produktion (Management of changes in production).

[5] Ullah, I., Tang, D., Yin, L., 2016, Engineering Product and Process Design Changes: A Literature Overview, Procedia CIRP, 56, pp. 25-33.

[6] Koch, J., 2017, Manufacturing Change Management - a Process Based Approach for the Management of Manufacturing Changes, Dissertation, Technische Universität München.

[7] Koch, J., Michels, N., Reinhart, G., 2016, Context model design for a process-oriented Manufacturing Change Management, Procedia CIRP, 41, pp. 33-38.

[8] Jarratt, T.A.W. et al., 2011, Engineering Change: An overview and perspective on the literature, Research Engineering Design, 22, pp. 103-124.

[9] Azab, A. et al., 2013, Mechanics of change: A framework to reconfigure manufacturing systems, CIRP Journal of Manufacturing Science and Technology, 6/2, pp. 110-119.

[10] Karl, F., Reinhart, G., 2015, Reconfigurations on manufacturing resources: Identification of needs and planning, Production Engineering, 9/3, p. 393-404.

[11] Hoang, X.-L. et al., 2017, Generation and impact analysis of adaptation options for automated manufacturing machines, 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), IEEE, pp. 1-8.

[12] Marks, P. et al., 2018, A systematic approach for supporting the adaptation process of discrete manufacturing machines, Research in Engineering Design, 24/3, pp. 1-21.

[13] Jacob, A. et al., 2018, Planning, Evaluation and Optimization of Product Design and Manufacturing Technology Chains for New Product and Production Technologies on the Example of Additive Manufacturing, Procedia CIRP, 70, pp. 108-113.

[14] Olmez, H. et al., 2018, Modelling Change within an Integrated Approach to Manufacturing System Design, Proceedings of the International Design Conference – Design 2018, pp. 611-622.

[15] Bauer, H., Schoonmann, A., Reinhart, G., 2017, Approach for model-based change impact analysis in factory systems, 2017 IEEE International Systems Engineering Symposium (ISSE), pp. 1-7.

[16] Cichos, D., 2017, Physikalische Modellierung technischer Änderungen in der Produktion, Dissertation, Technische Universität Kaiserslautern.

[17] Cichos, D., Aurich, J.C., 2017, Integration of a Physical Model into the Realization of Engineering Changes in Manufacturing Systems, AMM, 869, pp. 159-166.

[18] Hoang, X.-L. et al., 2017, Modeling of interdependencies between products, processes and resources to support the evolution of mechatronic systems, IFAC-PapersOnLine, 50, pp. 4348-4353.

[19] Plehn, C., 2017, A Method for Analyzing the Impact of Changes and their Propagation in Manufacturing Systems, Dissertation, Technische Universität München.

[20] Bruno, G., 2016, A Support System to Manage Product and Process Changes in Manufacturing, IFAC-PapersOnLine, 49/12, pp. 1080-1085.

[21] Cichos, D., Aurich, J.C., 2016, Support of Engineering Changes in Manufacturing Systems by Production Planning and Control Methods, Procedia CIRP, 41, pp. 165-170.

[22] Cichos, D., Kölsch, P., Aurich, J.C., 2016, Concept for Planning and Controlling of Multiple, Parallel Engineering Changes in Manufacturing Systems, AMR, 1140, pp. 497-504.

[23] Gernhardt, B. et al., 2016, Knowledge-Based Production Planning Within the Reference Planning Process Supporting Manufacturing Change Management, Materials; Biomanufacturing; Properties, Applications and Systems; Sustainable Manufacturing, ASME, pp. 1-10.

[24] Koch, J., Gritsch, A., Reinhart, G., 2016, Process design for the management of changes in manufacturing: Toward a Manufacturing Change Management process, CIRP Journal of Manufacturing Science and Technology, 14, pp. 10-19.

[25] Cichos, D., Aurich, J. C., 2015, Planning and Controlling of Multiple, Parallel Engineering Changes in Manufacturing Systems, Procedia CIRP, 33, pp. 81-86.

[26] Koch, J. et al., 2014, Structural Modeling of Extended Manufacturing Systems: An Approach to Support Changeability by Reconfiguration Planning, Procedia CIRP, 17, pp. 142-147.

[27] Malak, R.C., 2013, Methode zur softwarebasierten Planung technischer Änderungen in der Produktion, Dissertation, Technische Universität Kaiserslautern.

[28] Klocke, F., Brinksmeier, E., Weinert, K., 2005, Capability Profile of Hard Cutting and Grinding Processes, CIRP Annals, 54/2, pp. 22-45.

[29] Saaty, T.L., 1990, How to make a decision: The Analytic Hierarchy Process, European Journal of Operational Research, pp. 9-26.

9 BIOGRAPHY

Lennard Hermann, M. Sc. studied production engineering at the RWTH Aachen University. Since 2018 he works as research assistant in the workgroup Technology Planning at the Laboratory for Machine Tools and Production Engineering (WZL) of RWTH Aachen University.

Jan Rey, M. Sc. studied production engineering at the RWTH Aachen University. Since 2015 he works as a research assistant and took over the lead of the workgroup Technology Planning at the Laboratory for Machine Tools and Production Engineering (WZL) of RWTH Aachen University in 2018.

Prof. Dr.-Ing Thomas Bergs studied mechanical engineering at the University of Duisburg GH and design engineering at the RWTH Aachen University. In 2018 he was called to the RWTH Aachen as Professor of the Chair of Manufacturing Technology at the Laboratory for Machine Tools and Production Engineering (WZL).

Evaluation of the Applicability of Real-Time Data for Key Performance Indicators and Logistic Models for an Enhanced Manufacturing Operations Measurement

M. Hübner[1], P. Nyhuis[1]
[1]Institute of Production Systems and Logistics, Leibniz Universität Hannover, Germany

Abstract
In order to keep pace with growing demands in terms of short product life cycles, more diversity of variants and a growing competitiveness on world markets, companies must be able to recognise potentials and problems in their manufacturing processes as well as in their logistic processes as fast as possible. To achieve a high manufacturing and logistic performance and high customer satisfaction, the implementation of key performance indicators (KPIs) and Logistic Models serves, as an important clue for past, present and future developments in terms of strategic and operational target compliance. Through applying KPIs and Logistic Models, the relevant process data is compressed, summarised and used to reduce the analytical work to identify deficiencies and improve the decision making process. The availability of real-time data and its processing capabilities offer opportunities for an enhanced application of Logistic Models and KPIs. The contemplated hierarchy levels in manufacturing companies are the executive manager, the shop floor manager and the worker on the shop floor. Relevant KPIs and Logistic Models are chosen to fit the specific needs of these different roles and are assessed regarding their real-time capabilities. This paper evaluates real-time capable performance measurement systems and can be used to choose KPIs and Logistic Models for an enhanced manufacturing operations management.

Keywords
real-time data management; KPI; Logistic Models

1 INTRODUCTION

Change is a constant companion in the economic environment. Often seen as a critical factor it can also create new opportunities; to stand out from ones competitors. The ability to react or to foresee sudden changes in the production offers new possibilities to upscale the performance of a company. Logistic Models are used in this context to show the status or the development of production processes. These Logistic Models need to represent the important aspects of the considered company to improve the success of the company, which is done by using KPIs. [1]

Performance is the key to economic success. Performance measurement anyhow is failing in many organisations worldwide. Either the measures are not linked to the critical success factors of the organisation or their reviews are not constantly maintained. Hence, KPIs need to be emplaced, that constantly represent the critical success factors of the evaluated organisation. Only what is measured can be improved. Measurement furthermore allows people to recognise progress and focuses the attention on important issues. It improves objectivity and understanding of the decision making process. The processing and analysis of KPIs in real-time enables the organisation to react as fast as possible to incidents and to adjust the production program accordingly, with minimal impact on the customer. Therefore, companies need a system, which integrates Logistic Models and KPIs with real-time capability, for production planning and control systems regarding the individual operator characteristics. [2]

2 THEORETICAL FUNDAMENTALS

The following subsections declare the important vocabulary regarding KPIs and real-time data collection. In addition, the operator characteristics and hierarchy levels in a company are introduced. Furthermore, Modern Logistic Models, which can add to Key Data Systems and operate in a real-time environment are presented and elucidated by figures.

2.1 Logistic Objectives

A fair balance between the four logistic objectives Throughput Time, Delivery Reliability, Stock and Capacity Utilisation is necessary in a dynamic economic production environment [3]. Furthermore, flexibility and transparency are getting more important and can also be treated as target criteria. To describe logistic objectives, KPIs and Logistic Models can be chosen and related to these objectives to create a Key Data System.

2.2 Real Time

The concept of real-time describes the capability of a system to react almost at the same time an incident happens, so a response can take place before the system gets affected by the incident. Therefore, the information must be recorded at the Point of Creation

(POC) and immediately used at the Point of Action (POA), where decisions are made and implanted. The collected real-time data will be individually defined for each company and secured in a central database [4]. In this case, real-time means to sense changes in the production and react before the production system suffers from these changes. Therefore, four different data renewal intervals have been defined:

- Real-time data collecting,
- Daily data collecting,
- Monthly data collecting,
- Yearly data collecting.

Not all data is collected in real-time, because real-time data collecting is in many cases too expensive, in comparison to its potential benefits [5]. The potential benefits lie more precise production planning, founded on actual production data, and in the possibility of fast reactions to incidents, based on previous collected data, before or almost at the same time they occur. A more precise planning allows an optimised utilisation of resources and fast reactions decrease potential productivity losses caused by breakdowns [6].

2.3 Operator Characteristics

Hierarchy levels in companies are based on their individual needs and tasks [7]. In order to relate every KPI and Logistic Model to a hierarchy level, three different operator characteristics have to be identified. These three operator characteristics: worker at the shop floor, the shop floor manager and the executive management are divided by their planning and acting horizon, which decreases from the top to the bottom in the company hierarchy. [8]

- The executive management acts and plans towards a long-term future and defines the desired targets. The executive manager also splits the main target into smaller targets for each shop floor manager.
- The shop floor manager acts and plans towards a mid-term future and reports actual target compliance to the executive manager. The targets he receives from the executive manager are being transformed into short-term targets for the worker at the shop floor.
- The worker at the shop floor only acts towards a short-term future and reports actual targets to the shop floor manager.

2.4 Key Performance Indicators and Key Data Systems

A Key Performance Indicator is a quantified, compressed information about a specific situation which includes a large influence on the production process. It can be used for production planning and controlling, to reduce the data volume and to facilitate the decision making processes. The application of KPIs also creates transparency because every decision based on these KPIs is based on actual production data [8, 9]. To allow a comprised survey of the status in a company, different KPIs can be combined to a Key Data System [9, 10]. A Key Data System based on KPIs, which are related to logistic objectives, enables the company to track one individual logistic objective without disregarding the other.

2.5 Logistic Models

Several papers by the Institute of Production Systems and Logistics present Logistic Models to show interrelations between logistic objectives. Regarding a real-time environment the following Logistic Models need to be considered. These three models allow the user to calculate KPIs and the effect of changes in the production system concerning the logistic objectives based on material flow data.

- The Throughput Element (Figure 1) defines the Throughput Time and its components, which are the inter-operation time and the operation time. Subsections of the inter-operation time are the post-processing waiting time, the transport time and the pre-process waiting time. The operation time can be divided into the set up time and the processing time. Target times can be defined for each section of the throughput element and can be constantly compared to the time each step actually needed in order to evaluate performance and to find areas for improvement. [1]

- The Throughput Diagram (Figure 2) is a graphic representation of the Funnel Model and it shows the system performance. Therefore it assumes that every capacity unit in the production can be comprehensively described through the input, stock and output by comparing the planned and finished work hours to the Shop Calendar Days (SCD). It enables the user to quickly determine the input rate, the output rate, the stock and the reach of a production unit or workstation and gives an overall view of its capacity utilisation. Furthermore, the adherence to delivery dates can be integrated with an application of target and actual output rate. [1, 11]

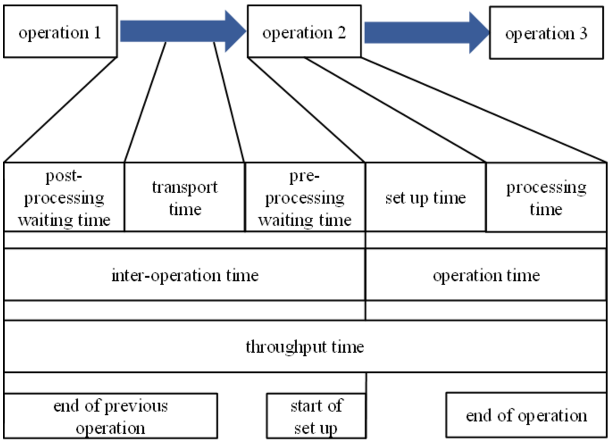


Figure 1 - Throughput Element

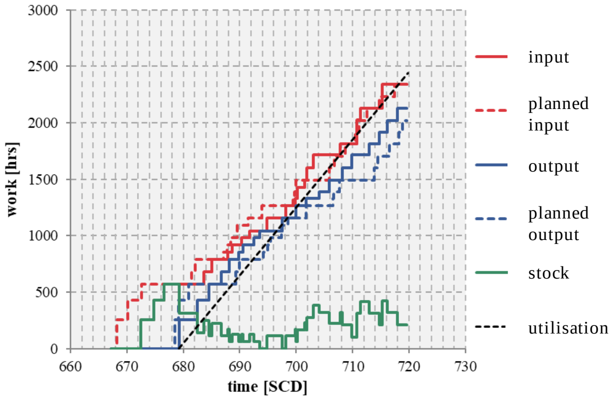


Figure 2 - Throughput Diagram

The Schedule Variance Histogram (Figure 3) is based on the Throughput Diagram. It shows the distribution of all orders which have been put through a certain workstation over a defined period of time, regarding their deviation from the specific due date. The bar chart visualises the relative frequency of the orders, while the line in the diagram shows the accumulated frequency. Hence, deficiencies regarding delivery reliability can be easily identified and countermeasures can be initiated before the backlog gets too big and other workstations or the delivery due date to the costumer is affected. [1]

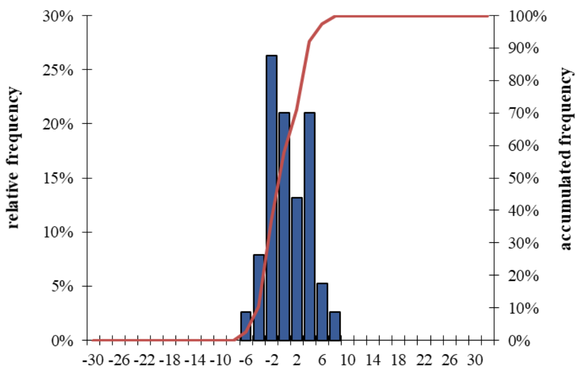


Figure 3 - Schedule Variance Histogram

3 ENHANCED OPERATIONS MANAGEMENT BY REAL TIME KEY PERFORMANCE INDICATORS AND LOGISTIC MODELS

In the following sections, relevant KPIs are selected and combined to create Key Data Systems to measure the defined logistic objectives of a production operation (c.f. 2.1). Afterwards the real-time capabilities of the identified KPIs and their Key Data Systems are evaluated and set into relation to specific operator levels. In addition, Data collection intervals for the selected KPI are proposed. In combination with Logistic Models it is possible to define an enhanced operations management, which is defined by KPIs and Models on recurring intervals on every hierarchy level in the company.

3.1 Creation of Key Data Systems by selection of Key Performance Indicators for the logistic objectives

Different Key Data Systems have been created with the intention to describe and analyse the status of the production and to show hidden potentials. To describe the stock as a logistic objective, the following six KPIs have been chosen: Stock, Turnover Rate, Stock Safety Level, Stock Safety Level Reliability, Store Range and Material Availability. This combination of different KPIs concerning the stock c a logistic objective focused on low stocks without risking a break in material flow. It is shown in Table 1. [12, 13]

Logistic objective	KPI	Formula
Stock	Stock	$= stock_{t-1} + inflow_t - outflow_t$
	Turnover Rate	$= \frac{outflow_t}{stock_t}$
	Stock Safety Level	$= replenishment\ time * consumption$
	Stock Safety Level Reliability	$= \frac{Days\ within\ Stock\ Safety\ Level}{Days}$
	Store Range	$= \frac{stock}{consumption}$
	Material Availability	$= \frac{\sum production\ orders\ in\ schedule}{\sum production\ orders}$

Table 1 - Key Data System: Stock

The logistic objective capacity utilisation is described by the six KPIs: Capacity Utilisation, System Availability, Efficiency Rate, Quality Ratio, Overall Equipment Efficiency (OEE), which includes the system availability, the efficiency rate and the quality ratio, and Total Effective Equipment Performance

(TEEP) which extends the OEE by the actually use capacity regarding the theoretical capacity. This Key Data Systems represents the idea of Total Productive Maintenance (TPM) by Nakajima and an additional view on unused capacities like the possibility of additional shifts. This Key Data System is shown in Table 2. [13–15]

To create a Key Data System for analysing the logistic objective Throughput Time, three KPIs have been chosen: Throughput Time, Throughput Time Deviation and Due Execution Time. With these KPIs the value adding processes can be described and put into relation to the total Throughput Time. In addition, potentials for optimisation can be identified and evaluated [1, 16, 17]. This Key Data system is shown in Table 3.

Logistic objective	KPI	Formula
Capacity Utilisation	Capacity Utilisation	$= \frac{actual\ output}{theoretical\ output}$
	(a) System Availability	$= \frac{work\ time}{operation\ time}$
	(b) Efficiency Rate	$= \frac{target\ prouction\ time}{operation\ time} *$
	(c) Quality Ratio	$= \frac{\sum good\ parts}{\sum parts}$
	Overall Equipment Efficiency (OEE)	$= a * b * c$
	Total Effective Equipment Performance (TEEP)	$= OEE * \frac{used\ capacity}{theoretical\ capacity}$

Table 2 - Key Data System: Capacity Utilisation

Logistic objective	KPI	Formula
Throughput Time	Throughput Time	$= order_{end} - order_{start}$
	Throughput Time Deviation	$= Throughput\ Time_{actual} - Throughput\ Time_{target}$
	Due Execution Time	$= \frac{Execution\ Time}{Throughput\ Time}$

Table 3 - Key Data System: Throughput Time

With the KPIs Due Date Reliability and Schedule Deviation the logistic objective Delivery Reliability can be analysed and optimised. These two KPIs measure the accuracy of the production planning regarding the delivery dates. The schedule deviation can also be used between two machines as well as to track the timeline of an order through the production, to identify time-critical processes [12]. The Key Data System delivery reliability is shown in Table 4.

Logistic objective	KPI	Formula
Delivery Reliability	Due Date Reliability	$= \frac{\sum deliveries\ in\ schedule}{\sum deliveries}$
	Schedule Deviation	$= start\ of\ production_{target} - start\ of\ production_{actual}$

Table 4 - Key Data System: Delivery Reliability

Transparency as logistic objective describes the ability to provide information about the status of a process or product [18]. This ability can be created by collecting and analysing data via KPIs and Logistic Models. There are also two KPIs which are indicators for transparency, the Data Renewal Interval and the order specific Work Progress. The small Data Renewal Interval means up-to-date production information by tracking the Work Progress. There is always a current state of progress for each order. The Key Data System transparency is shown in Table 5.

Logistic objective	KPI	Formula
Transparency	Data Renewal Interval	$= \frac{\sum data\ collections\ in\ period}{given\ period}$
	Work Progress	$= \frac{\sum finished\ production\ steps}{\sum production\ steps}$

Table 5 - Key Data System: Transparency

Flexibility describes the ability to react and adapt to occurrences in order to reduce possible negative impacts on production and logistic processes as fast as possible [19]. Therefore, the fallowing three KPIs have been chosen: Delivery Flexibility, Every Part Every Interval (EPEI) and the Variant Flexibility. These KPIs rate the flexibility based on retained capacities as well as the flexibility in delivery dates and variants [17, 20]. The Key Data System Flexibility is shown in Table 6.

Logistic objective	KPI	Formula
Flexibility	Delivery Flexibility	$= \frac{\sum granted\ special\ orders}{\sum special\ orders}$
	Every Part Every Interval (EPEI)	$= \left(\sum setup\ time\ per\ variant + \sum(production\ time * lot\ size)\ per\ variant\right) * \left(\frac{1}{available\ working\ time\ per\ day}\right)$
	Variant Flexibility	$= \sum economic\ fabricable\ variants$

Table 6 - Key Data System: Flexibility

3.2 Evaluation of real-time capabilities and proposal for relevant hierarchies of the presented Key Data Systems

The introduced KPIs vary in their ideal data renewal interval and affect different hierarchy levels in a company. Stock as KPI is real-time capable and affects all three management levels. The executive and the shop floor management define target stocks whereas on the worker level the actual stock level is reported. A tracking in real-time allows fast reactions in both directions of the hierarchy level. Trouble in the manufacturing process forces stock levels to rise. Rising stock levels are often an indicator for problems like low efficiency rate and increasing work time per part.
Another real-time capable KPI is the Store Range. A real-time tracking allows the shop floor manager to optimise the store stock level in the conflict between low stocks and high capacity utilisation. Turnover Rate, Stock Safety Level, Stock Safety Level Reliability and Material Reliability are also KPIs, which affect the shop floor manager. The Turnover Rate and the Stock Safety Level Reliability should be checked monthly, because they are used to define a minimum stock, which should not be changed daily. The Stock Safety Level defines the minimum stock in the store and can be updated monthly because it does not track the stock level itself. Also, the Material Availability is a KPI, which creates benefits, by analysing the data of a greater time period. A real-time collection would not create additional benefits for these last five KPIs because all of them need a larger database to unfold their potential. Table 7 summarises these results.

KPIs which are related to the logistic objective Capacity Utilisation create a benefit, if they are generated and tracked almost in real-time. This is based on the fact, that every loss in Capacity Utilisation directly results in a lower cost effectiveness. Therefore, the OEE (consisting of: System Availability, Efficiency Rate and Quality Ratio) measures the losses due to quality, failures and reduced output rates. The OEE can be used to rate the efficiency of the system and is a tool for the executive manager to define targets. The shop floor manager can use the OEE and its components to regulate the system whereas the worker uses the Efficiency Rate and Quality Ratio to optimise his work in real-time. The TEEP is an indicator for unused capacities because of unassigned shifts and low system loads. It can be used monthly to illustrate possibilities for higher system capacities. The TEEP is also related to the executive manager to facilitate strategic decisions regarding the capacity of the facility. Table 7 summarises the Data Renewal Interval and the affected hierarchy level for Capacity Utilisation KPIs.

The Throughput Time and the Throughput Time Deviation can be tracked in real-time. The variation of the Throughput Time can be an indicator for impending problems in the production line due to machine failure, missing parts or wrong assumptions in the production planning process. The executive manager can use the Throughput Time to determine key figures. The shop floor manager needs reliable data to fulfil those key figures. Using the Throughput Time Deviation it is possible to analyse the variation between the key figures and the actual Throughput Time values. The worker can also use the Throughput Time Variation to analyse and adapt his work speed. The Due Execution Time analyses the value generating part of the Throughput Time and can be used as a tool in a monthly interval or in case the Throughput Time indicates a problem. Thus, the Due Execution Time is a tool used by the shop floor manager. Table 7 shows the data renewal interval and the affected hierarchy levels of the Throughput Time KPIs.

Delivery Reliability has been described with Due Date Reliability and Schedule Deviation. Both KPIs can be tracked in real-time to guarantee a fast reaction if problems with targeted delivery dates occur. The Due Date Reliability concerns the shop floor and the executive manager. The shop floor manager must attain the target Delivery Reliability, which is provided by the executive manager. The Schedule Deviation can be used by the worker to adjust the work speed or work order sequences and by the shop floor manager to regulate imminent delays. Table 7 shows the Data Renewal Interval and the affected hierarchy levels.

Transparency is created by collecting and tracking data and visualising and documenting it frequently. Therefore, the KPI Data Renewal Interval can be used as Target ,defined by the executive manager, in a yearly interval as a standard for data collection, tracking and reviewing. The work progress can be tracked in real-time, which also creates transparency towards the customer. It is an important KPI for the shop floor manager. If it is combined with delivery dates and milestones, he can regulate the production process. Table 7 shows the KPIs and affected hierarchy levels, their data renewal intervals and their tractability in the Logistic Models.

The Flexibility is tracked with the KPIs Delivery Flexibility, EPEI (consisting of: the setup time, the production time and the lot size per variant in regard to the available working time per day) and the Variant Flexibility. The Delivery and Variant Flexibility are KPIs which are target values defined by the executive manager. Therefor the Delivery Flexibility must be updated monthly as sub-goals for the shop floor manager. Variant Flexibility is a meta goal, determined yearly or even longer. Companies not only have to change processes, but also adjust their goods, for example setting up a new production line, to ensure variant flexibility. The EPEI can be used as a tool by the shop floor manager to rate the Variant Flexibility in his production line. These Data Renewal Intervals, affected hierarchy levels and Logistic Models are shown in Table 7.

			Hierarchy Level			Logistic Models		
Logistic Objective	**KPI**	Data Collecting and Tracking	Woker	Shop floor Manager	Executive Manager	Throughput Element	Throughput Diagram	Shedule Variance Histogram
Stock	Stock	**R**	**X**	**X**	partly		**X**	
	Turnover Rate	**M**		**X**			**X**	
	Stock Safety Level	**M**		**X**			partly	
	Stock Safety Level Reliability	**D/M**		**X**				
	Store Range	**R**		**X**			**X**	
	Material Availability	**Y**		**X**			**X**	
Capacity Utilization	Capacity Utilization	**R**		**X**			**X**	**X**
	System Availability	**R**		**X**		**X**		
	Efficiency Rate	**R**	**X**	**X**			**X**	
	Quality Ratio	**R**	**X**	**X**				
	OEE	**R**		**X**	**X**		partly	
	TEEP	**M**			**X**		partly	
Throughput Time	Throughput Time	**R**		**X**	**X**	**X**	**X**	
	Throughput Time Deviation	**R**	**X**	**X**		**X**	**X**	
	Due Execution Time	**M**		**X**		**X**		
Delivery Reliability	Due Date Reliability	**R**		**X**	**X**		**X**	**X**
	Schedule Deviation	**R**	**X**	**X**			**X**	**X**
Transparency	Data Renewal Interval	**Y**			**X**			
	Work Progress	**R**		**X**			**X**	
Flexibility	Delivery Flexibility	**M**			**X**			
	EPEI	**M**		**X**		**X**		
	Variant Flexibility	**Y**			**X**			
Table	**R** = Real-time **W** = Weekly **M** = Monthly							

Table 7 - Evaluation of Data Renewal Interval and Hierarchy Levels for the presented Key Data Systems and Logistic Models

In addition to the listed KPIs, the Throughput Diagram is capable to track the overall performance, the input schedule deviation, the backlog regarding the input and the output.

3.3 Evaluation of real-time capabilities and proposal for relevant hierarchies for Logistic Models

Logistic Models are tools for the shop floor manager to track important KPIs based on data from the ERP-System. The Throughput Element, the Throughput Diagram and the Schedule Variance Histogram are capable of handling real-time data and can be used by the worker to predict imminent problems in the workflow. The incoming and exiting material flow is recorded at a work station with the Throughput Diagram, these actual values can be compared to the target values, and thereby a prediction for a possible delay can be made. Both, worker and shop floor manager can react and correct this delay. The Throughput Element can be used to define a target Throughput Time, which is done by the shop floor manager. For the worker, it is just an information about the target value. He automatically reports the actual Throughput Time back to upper hierarchy levels. The Schedule Variance Diagram builds on this information and visualises the deviation to the predefined due date in real-time, for the worker as well as the shop floor manager. The real-time capability and the hierarchy levels are shown in Table 8.

KPI	Real-time capability	Hierarchy Level		
		Worker	Shopfloor Manager	Executive Manager
Throughput Element	**yes**	**X**	**X**	**X**
Funnel model and Throughput Diagram	**yes**	**X**	**X**	
Schedule Variance Histogram	**yes**		**X**	

Table 8 - Logistic models: Real-time capability and Hierarchy Levels

4 SUMMARY AND OUTLOOK

In this paper the real-time capabilities of selected KPIs and Logistic Models are evaluated in order to assess the production target compliance regarding the logistic objectives: Throughput Time, Delivery Reliability, Stock, Capacity Utilisation Flexibility and Transparency. Overall it is concluded that real-time information can be relevant for most of the selected KPIs. The selected KPIs and Logistic Models can simplify the daily work of the related operator and generate value by reducing the analytical work and displaying important information about the production system. In addition the ceated Key Data Systems can be used to define and pursue the logistic objectives on each hierarchy level regarding their needs. It is notable that especially in the middle hierarchy levels of manufacturing companies a high demand of real time KPIs is expected, since Shop floor managers and Manufacturing Execution Systems (MES) are the key roles to assessing manufacturing performance and taking measures accordingly.

The presented results can be used to choose and calibrate KPIs and Logistic Models for an enhanced manufacturing operations management with defined review and data collection intervals.

Further research should address the assessment of economic benefits of real-time data collection. These benefits depend on individual companies characteristics, for example the layout of the production system and the degree of digitalisation. Some advantages that are reated through collection of real time data are hard to assess, e.g. the benefits of transparency cannot be measured easily and should be considered in future research.

5 REFERENCES

[1] Nyhuis P, Wiendahl H-P (2003) Logistische Kennlinien. Springer Berlin Heidelberg, Berlin, Heidelberg

[2] Parmenter D (2015) Key Performance Indicators: Developing, implementing, and using winning KPIs, Third edition. Wiley, Hoboken, New Jersey

[3] Nyhuis P (2007) Practical Applications of Logistic Operating Curves. CIRP Annals - Manufacturing Technology (56): 483–486

[4] Alt R, Österle H (2004) Real-time Business: Lösungen, Bausteine und Potenziale des Business Networking. Springer, Berlin, Heidelberg

[5] Irrenhauser T (2014) Bewertung der Wirtschaftlichkeit von RFID im Wertschöpfungsnetz. Zugl.: München, Techn. Univ., Diss., 2014. Utz, München

[6] Götz T, Safai S, Beer P (2006) Efficient supply chain management with SAP solutions for RFID, 1. ed., vol 14. SAP Press, Boston, Mass.

[7] Ebeling I, Vogelauer W, Kemm R (2012) Die systemisch-dynamische Organisation im Wandel: Vom fließenden Umgang mit Hierarchie und Netzwerk im Veränderungsprozess, 1. ed. Haupt, Bern

[8] Badawy M, El-Aziz A, Idress A et al. (2016) A survey on exploring key performance indicators. Future Computing and Informatics Journal (1-2): 47–52

[9] Reichmann T, Kißler M, Baumöl U et al. (2017) Controlling mit Kennzahlen: Die systemgestützte Controlling-Konzeption, 9. ed. Verlag Franz Vahlen, München

[10] Gladen W (2014) Performance Measurement: Controlling mit Kennzahlen, 6. ed. Springer Gabler, Wiesbaden

[11] Wiendahl H-P, Tönshoff K (1988) The Throughput Diagram — An Universal Model for the Illustration, Control and Supervision of Logistic Processes. CIRP Annals 37(1): 465–468

[12] VDI VDI 4400 Blatt 2 Logistikkennzahlen für die Produktion - 2000

[13] Kummer S, Grün O, Jammernegg W (eds) (2008) Grundzüge der Beschaffung, Produktion und Logistik. Pearson Studium, München

[14] Gorecki P, Pautsch P (2013) Praxisbuch Lean Management: Der Weg zur operativen Excellence. Hanser, München

[15] Nakajima S (1988) Introduction to TPM: Total productive maintenance. Productivity Press, Cambridge, Mass.

[16] Arnold D, Isermann H, Kuhn A et al. (eds) (2008) Handbuch Logistik, 3. ed. Springer, Berlin

[17] Gottmann J (2016) Produktionscontrolling: Wertströme und Kosten optimieren, 1. ed. Springer Gabler, Wiesbaden

[18] Kletti J, Schumacher J (2011) Die perfekte Produktion: Manufacturing Excellence durch Short Interval Technology (SIT). Springer-Verlag Berlin Heidelberg, Berlin, Heidelberg

[19] Wiendahl H-P (2010) Betriebsorganisation für Ingenieure, 7. ed. Hanser, München

[20] Eversheim W, Schuh G (eds) (1996) Hütte: Produktion und Management; Betriebshütte, 7. ed. Springer, Berlin

6 BIOGRAPHY

Prof. Dr.-Ing. habil. Peter Nyhuis, born 1957, studied mechanical engineering at Leibniz Universität Hannover and subsequently worked as a research assistant at the Institute for Production Systems and Logistics (IFA). After obtaining his Dr.-Ing. doctorate, he was habilitated before working as an executive in the field of supply chain management in the electronics and mechanical engineering industry. Since 2003 he was appointed Professor and has lead the Institute for Factory Systems and Logistics (IFA) at Leibniz Universität Hannover.

Marco Hübner, M.Sc., born in 1989, has been a research associate in the Production Management Division of the Institute for Production Systems and Logistics (IFA) at Leibniz Universität Hannover since 2015. He previously studied mechanical engineering at Leibniz Universität Hannover with exchanges at The University of British Columbia (UBC) and Keio Unitversity in Tokyo.

Statistical Certainties of Expert-Supported Quality Forecasts for Groups of Product Characteristics or Process Parameters

K. N. Kostyszyn[1], R. Schmitt[2]

[1]Fraunhofer Institute for Production Technology IPT, Department Production Quality, Aachen, Germany

[2]Laboratory for Machine Tools and Production Engineering (WZL), RWTH Aachen University, Chair for Quality Management and Metrology, Germany

Abstract

In small batch production, statistical methods for derivation of quality forecasts are rarely applied. The reason is that sample data available for statistical analysis is strongly limited. To increase available amounts, several scientific approaches recommend to create and to analyze groups of product characteristics or process parameters. However, those approaches do not quantify statistical certainties of resulting quality forecasts. Also, available expert knowledge and experience regarding process behavior is not applied although experts may regard specific quality forecasts as not realistic. In this case, those forecasts could be excluded from consideration. To overcome all mentioned deficiencies, this work introduces an approach to quantify statistical certainties of quality forecasts for groups of product characteristics or process parameters. Quality forecasts are expected to be based on typical patterns that can be observed in z-charts. Beside empirical distributions of available sample data, possible underlying infinite populations are taken into account. Experts can limit the scopes of considered infinite populations to those ones that are regarded as realistic. This enables a knowledge-based specification of quantified statistical certainties. All considered sample data is assumed to follow normal distributions. In this paper, sample values of an exemplary group are plotted in a chart where a trend pattern is detected. The statistical certainty for this trend will be calculated with the presented approach.

Keywords

Statistical Process Control, Statistical Certainties, Quality Control

1 INTRODUCTION

In various fields of industry, producing companies are facing an increase of individual customer needs. In combination with competitive pressure, the transformation from mass or large batch production to small batch production is gaining more importance [1]. Modern technologies and adaptive process chains are targeted to fulfill high efficiency and quality standards. For quality assurance, increasing small batch production leads to new challenges. In many process chains, quality control depends on sample measurements. Typical arguments against 100 % quality control are time-consuming quality control processes and limited capacities regarding personnel or technical equipment [2]. Thus, statistical analyses of measured samples are required. This can be done with Statistical Process Control (SPC) techniques that are widely used in modern quality management systems. However, sample measurements in small batches result in strongly limited amounts of samples that are available for analysis. In many cases, those amounts are even not high enough to provide reliable preliminary estimates regarding the output of an observed process. ISO 7870-2:2013 recommends to collect 25 samples of size 4 or 5 during a continuous process run in classical SPC application [3]. In contrast, small batches often consist of even less than 25 items.

In scientific literature, different approaches have emerged to transform classical SPC to application in small batch production. However, statistical certainties of quality forecasts based on pattern detections are not quantified although they could be considered as useful to support experts with their decision-makings. Moreover, available expert knowledge and experience regarding process behaviors is not considered.

In the following, a new approach is introduced for quantification of statistical certainties of quality forecasts for groups of product characteristics or product parameters. Those forecasts can be derived from patterns found in a z-chart which is a specific type of control chart [4]. To increase the amount of samples, a group of product characteristics or process parameters, in the following called *features*, can be observed in one chart. The group is not limited to one feature or to one nominal value but it is expected that all measured features are having systematic process influences in common. It is further required that measured samples of each feature are following a normal distribution.

The quantification of statistical certainties is presented with the aid of an example. Artificially generated normally distributed sample values are plotted in a z-chart. In this chart, a continuous process shift, also called *trend*, is detected. In a further step, expert knowledge is applied to limit the scope of considered infinite populations for quantification of statistical certainties. In this paper, all values have no units.

2 SHORT-RUN SPC

2.1 Approximation of infinite populations with empirical distributions

In classical SPC applications, the parameters describing the distribution of a feature are preliminary determined. For this, measurements of around 25 samples, also called *subgroups*, of sample size 4 or 5 during a continuous process run are required. Assuming that the process is stable, distribution parameters can initially be derived considering an amount of k samples:

$$X_k = \{x_1, x_2, \dots, x_k\}. \quad (1)$$

All samples considered in this paper are required to follow normal distributions. To describe the density function of a normal distribution, two parameters are required: Its mean μ and its standard deviation σ. In real production, those parameters cannot precisely be determined since this would require an infinite amount of available samples that have to be measured while constant conditions regarding influencing factors and factor levels are ensured. Due to this fact, the population that is characterized by those two unknown parameters can be denoted as *infinite population*. [5]

To approximate the distribution of an infinite population, empirical parameters can be calculated with the available amount of samples. The empirical mean $\bar{x}$ and the empirical standard deviation s are determined as follows:

$$\bar{x} = \frac{1}{k}\sum_{i=1}^{k} x_i\ , \quad (2)$$

$$s = \sqrt{\frac{1}{k-1}\sum_{i=1}^{k}(x_i - \bar{x})^2}\ . \quad (3)$$

The values of the unknown parameters can now be estimated through calculation of interval limits. With the empirical parameters, the available number of samples k, and a pre-defined statistical certainty of $p = 1 - \alpha$ where α is the so-called level of significance [2], it is:

$$\bar{x} - \frac{t_{k-1,b}s}{\sqrt{k}} \leq \mu \leq \bar{x} + \frac{t_{k-1,b}s}{\sqrt{k}}\ , \quad (4)$$

$$\sqrt{\frac{k-1}{\chi^2_{k-1,b}}}s \leq \sigma \leq \sqrt{\frac{k-1}{\chi^2_{k-1,a}}}s. \quad (5)$$

The resulting intervals are called *confidence intervals* [2]. t_{k-1} describes a Student t-distributed variable with $k-1$ degrees of freedom. With a statistical certainty of $p = 95\,\%$ $(\alpha = 5\,\%)$, its specific value $t_{k-1,b}$ is characterized by the cumulated density function of the Student's t-distribution with $F(t_{k-1,b}) = 1 - \frac{\alpha}{2} = 97.5\,\%$. Analogically, the specific values of the χ^2-distributed variable χ^2_{k-1} are characterized by the cumulated density function with $F(\chi^2_{k-1,b}) = 1 - \frac{\alpha}{2} = 97.5\,\%$ and $F(\chi^2_{k-1,a}) = \frac{\alpha}{2} = 2.5\%$.

Assuming empirical parameters $\bar{x} = 0$ and $s = 1$, the confidence interval limits can be calculated. In the following, the calculation was performed for two different sample amounts $k = 25$ and $k = 3$:

$$-0.41 \leq \mu \leq 0.41\ (k = 25), \quad (6)$$

$$-2.48 \leq \mu \leq 2.48\ (k = 3), \quad (7)$$

$$0.78 \leq \sigma \leq 1.39\ (k = 25), \quad (8)$$

$$0.52 \leq \sigma \leq 6.28\ (k = 3). \quad (9)$$

It is obvious that little amounts of samples only enable considerably inaccurate estimations of the distribution parameters. As a consequence, classical SPC cannot be applied in small batch production. Method adaptions are required.

2.2 Short-Run SPC and Grouping

The scientific literature and existing norms describe different approaches for quality control in small batch production. In most cases, the term *Short-Run SPC* is named to imply the application field. Short runs refer to production scenarios where only little amounts of product items are produced before a different item is then produced [6]. In contrast to classical SPC, the focus lies on the process and not on the product. That is because same or similar processes are commonly used for different products. One example is a milling process that is applied to different products with different shapes and materials with one machine. Short-Run SPC suggests to analyze aggregated samples from different features to describe the process output. Units, nominal values and spreads do not need to be the same.

ISO 7870-8:2017 recommends various control charts depending on specific sample data. It can be chosen with the help of a proposed decision tree. Most control charts for Short-Run SPC are restricted to a sample size of one [6]. Some control charts are specialized in samples of same features with nearly constant spreads but different nominal values. In those charts, only the distances of individual values to their nominal values are plotted. Other control charts are specialized in samples that describe different features. In those charts, different nominal values and different spreads are expected. Distances between individual and nominal values are standardized through division by expected ranges. In this way, samples with unequal spreads become comparable. A similar approach is presented with z-charts that even accept different sample sizes. The quantification approach presented in this paper assumes the application of z-charts. The applicability with other types of control charts is currently not validated.

For the grouping of features, different methods can be applied. ISO 7870-8:2017 recommends to make the grouping rely on common systematic process influences. Other publications complement this approach with statistical hypothesis testing to ensure equal means and standard deviations. Moreover, definitions of classification trees or coding systems are proposed [7-10]. Beside expert-driven groupings, also clustering algorithms can be applied [11]. For this paper, an exemplary group was created in advance. The quantification of statistical certainties of quality forecasts is independent from the previous grouping approach.

3 CERTAINTIY OF QUALITY FORECASTS

3.1 Statistical certainties of trend forecasts

In the following, the example data that will be analyzed to calculate the statistical certainty of a detected trend is introduced. Chapter 3.1.2 describes the theoretical approach for calculation and applies it to the available sample data. With chapter 3.1.3, an optional consideration of expert knowledge and experience is introduced.

3.1.1 Example data and z-chart

For the prepared example, available sample data can be formulated as

$$X_{i,k_i} = \{x_{i,1}, x_{i,2}, \dots, x_{i,k_i}, x_{i,k_i+1}, x_{i,k_i+2}, x_{i,k_i+3}\}, \quad (10)$$

where $i = 1, 2$ differs between two features mixed with other features in one group. The first k_i samples of each feature i are assumed to represent a stable process and to follow a normal distribution. In the described example, it is $k_1 = k_2 = 3$. The last three samples of each feature are a part of the detected trend.

As shown in (2) and (3), the empirical parameters of the normal distributions can be calculated with

$$\bar{x}_i = \frac{1}{k_i}\sum_{j=1}^{k_i} x_{i,j}\,, \quad (11)$$

$$s_i = \sqrt{\frac{1}{k_i-1}\sum_{j=1}^{k_i}\left(x_{i,j} - \bar{x}_i\right)^2}. \quad (12)$$

It is expected that the standard deviations of both features are not equal. To make process developments for both characteristics comparable in one chart, a z-transformation of the sample data is prepared. It is

$$Z_{i,k_i} = \{z_{i,1}, z_{i,2}, \dots, z_{i,k_i}, z_{i,k_i+1}, z_{i,k_i+2}, z_{i,k_i+3}\}, \quad (13)$$

with

$$z_{i,j} = \frac{x_{i,j} - \mu_i}{\sigma_i} \approx \frac{x_{i,j} - \bar{x}_i}{s_i}. \quad (14)$$

All sample data, calculated empirical parameters and z-values are listed in Table 1. Figure 1 illustrates the resulting z-chart respecting a time-based sequence of z-value measurements marked by an increasing ID. The values between ID 6 and 101 are from other features that are also part of the monitored group. In this example, the calculation of control limits is neglected.

First $k_1 = k_2 = 3$ values			Trend members			Empirical parameters	
Sample values of feature 1							
$x_{1,1}$	$x_{1,2}$	$x_{1,3}$	$x_{1,4}$	$x_{1,5}$	$x_{1,6}$	$\bar{x}_1$	s_1
-2.18	-1.16	-3.11	-3.37	-3.12	-1.50	-2.15	0.98
Sample values of feature 2							
$x_{2,1}$	$x_{2,2}$	$x_{2,3}$	$x_{2,4}$	$x_{2,5}$	$x_{2,6}$	$\bar{x}_2$	s_2
2.40	1.01	1.85	2.56	3.12	3.56	1.75	0.70
z-values of feature 1							
$z_{1,1}$	$z_{1,2}$	$z_{1,3}$	$z_{1,4}$	$z_{1,5}$	$z_{1,6}$		
-0.02	1.01	-0.98	-1.25	-0.99	0.66		
z-values of feature 2							
$z_{2,1}$	$z_{2,2}$	$z_{2,3}$	$z_{2,4}$	$z_{2,5}$	$z_{2,6}$		
0.92	-1.06	0.14	1.15	1.95	2.58		

Table 1 - Sample values, derived empirical distribution parameters and z-values

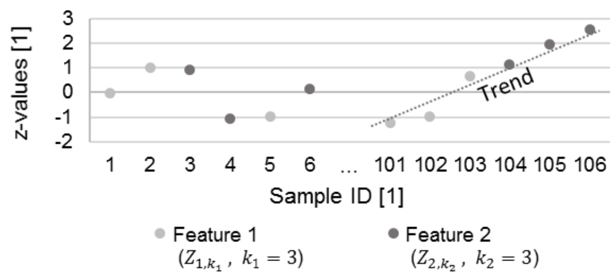


Figure 1 - z-chart with example data

The trend detection relies on a Nelson rule, that describes a trend as 6 increasing or decreasing values [12]. The implied quality forecast refers to an unstable process with a continuous shift of process means. Thus, the trend-signalizing samples are not considered for calculation of empirical means and standard deviations.

In chapter 2.1, exemplary calculations showed that empirical means and standard deviations are no precise approximations when being calculated based on little amounts of samples as typical for small batches. Since the z-values are approximated with such low-precision parameters, see (14), it cannot be ensured that the detected trend in this example is a real one. It is obvious that the first and last three z-values within this detected trend are continuously increasing since they each describe a single feature that is expected to follow a normal distribution. However, an increase from the third to the fourth z-value cannot be guaranteed. In most cases, the certainty is supposed to be considerably lower than 100 %. For the shown trend, this certainty will be calculated in the following.

3.1.2 Calculation of statistical certainty

The calculation approach considers the fact that different infinite populations could have led to the possible values of the available samples. In fact, these infinite populations are underlying a specific distribution since each of their possible parameters μ_i and σ_i are following specific distributions. All combinations of possible means μ_i and standard deviations σ_i will be considered for calculation of the statistical certainty. The integration of the densities of all combinations will finally lead to the required statistical certainty value.

As already visible in (4) and (5), the distributions of possible parameters μ_i and σ_i can be described with

$$\mu_i \sim \bar{x}_i + t_{k_i-1} \cdot s_i/\sqrt{k_i}, \quad (15)$$

$$\sigma_i \sim s_i \cdot \sqrt{k_i - 1}/\sqrt{\chi^2_{k_i-1}}. \quad (16)$$

μ_i follows a Student's t-distribution and σ_i follows a χ^2-distribution. Referring to (14), it can now be derived that a z-value follows a combination of both:

$$z_{i,j} = \frac{x_{i,j} - \mu_i}{\sigma_i} \sim \frac{x_{i,j} - \left(\bar{x}_i + t_{k_i-1}\frac{s_i}{\sqrt{k_i}}\right)}{s_i\frac{\sqrt{k_i-1}}{\sqrt{\chi^2_{k_i-1}}}} = c_{i,j}. \quad (17)$$

For a chosen value $c_{i,j}$, there is an infinite amount of possible combinations of t_{k_i-1} and $\chi^2_{k_i-1}$ values while each specific value is connected to a density value defined by its distribution. To calculate the statistical certainty of a real trend as detected in the z-chart, the densities of all combinations have to be integrated where it is

$$c_{2,1} \geq c_{1,3}. \quad (18)$$

Through this requirement, only those infinite populations are considered that ensure

$$z_{2,1} \geq z_{1,3}. \quad (19)$$

Before introducing the final integral, a transformation of (17) is suggested. According to literature [13-14], it can be stated that

$$t_{k_i-1}\sqrt{\chi^2_{k_i-1}} = \nu_i\sqrt{k_i - 1}, \quad (20)$$

where ν_i is following a standard normal distribution which is a normal distribution with a mean equal to zero and a standard deviation equal to one. Integrating (20) in (17) results into

$$\frac{x_{i,j} - \bar{x}_i}{\sqrt{k_i - 1}} \cdot \frac{\sqrt{\chi^2_{k_i-1}}}{s_i} - \frac{\nu_i}{\sqrt{k_i}} = c_{i,j}. \quad (21)$$

This can be transformed to

$$\nu_i = \frac{x_{i,j} - \bar{x}_i}{s_i}\sqrt{\frac{k_i}{k_i - 1}}\sqrt{\chi^2_{k_i-1}} - \sqrt{k_i}c_{i,j}. \quad (22)$$

The certainty of a real trend p_{trend} can now be formulated as an integral of densities for all combinations of ν_i and $\chi^2_{k_i-1}$ that fulfill this equation. Since (22) applies to both features that are normally distributed, a term with four integrals has to be calculated:

$$p_{trend} = \int_0^\infty \int_0^\infty \int_{I_{\nu_1,a}}^{I_{\nu_1,b}} \int_{I_{\nu_2,a}}^{I_{\nu_2,b}} f_{\chi^2_{k_1-1}}(x_1) f_{\chi^2_{k_2-1}}(x_2)$$
$$\cdot f_\nu(\nu_1) f_\nu(\nu_2) d\nu_2 d\nu_1 dx_2 dx_1. \quad (23)$$

$f_{\chi^2_{k_i-1}}$ and f_ν are the density functions of the χ^2-distribution and standard normal distribution:

$$f_{\chi^2_{k_i-1}}(x_i) = \begin{cases} \frac{x_i^{\frac{k_i-1}{2}-1} e^{-\frac{x_i}{2}}}{2^{\frac{k_i-1}{2}}\Gamma\left(\frac{k_i-1}{2}\right)}, x_i > 0 \\ 0, x_i \leq 0 \end{cases}, \quad (24)$$

$$f_\nu(\nu_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\nu_i^2}, -\infty < \nu_i < \infty. \quad (25)$$

Γ describes the so-called *gamma function* [15]. The outer two integrals of (23) are limited to the interval $[0, \infty)$ since the density functions of χ^2-distributions are zero for negative values. To define the limits of the inner integrals, $d\nu_2$ and $d\nu_1$ will be expressed through $dc_{2,1}$ and $dc_{1,3}$. With (22), it is

$$d\nu_2 = -\sqrt{k_2}\, dc_{2,1} \quad (26)$$

$$d\nu_1 = -\sqrt{k_1}\, dc_{1,3} \quad (27)$$

Claiming (18) to ensure a real trend, it is:

$$p_{trend} = \sqrt{k_1 k_2} \int_0^\infty \int_0^\infty \int_{-\infty}^\infty \int_{c_{1,3}}^\infty f_{\chi^2_{k_1-1}}(x_1) f_{\chi^2_{k_2-1}}(x_2)$$
$$\cdot f_\nu\left(\nu_1(c_{1,3})\right) f_\nu\left(\nu_2(c_{2,1})\right) dc_{2,1} dc_{1,3} dx_2 dx_1. \quad (28)$$

Here, $\nu_1(c_{1,3})$ and $\nu_2(c_{2,1})$ signalize the dependency of ν_1 and ν_2 from $c_{1,3}$ and $c_{1,3}$ as shown in (22). The integral can now numerically be solved. This can be done with Python 3.6.6 and SciPy 1.1.0 which offers the function *nquad* for multiple integration. Default settings were applied which include an estimated absolute error of $1.49 \cdot 10^{-8}$. The calculation took around 120 seconds.

The statistical certainty for a real trend situation as suggested by the z-chart in Figure 1 is $p_{trend} = 65.77\,\%$. The example describes a simple situation for a trend. A more complex situation would appear when 6 samples of 6 different distributions create a trend. In this case, the calculation would include 12 integrals with adapted boundaries. As a restriction, every member of a trend is required to offer at least two samples in advance that can be assumed to represent a stable process and therefore be used to calculate empirical parameters of their normal distribution.

3.1.3 Consideration of expert knowledge

In the previous calculation of the statistical certainty, all sets of possible means μ_i and standard deviations σ_i that could have led to the values of the measured samples are considered. However, in some real use cases of SPC, experts can estimate a range of realistic means and standard deviations with their knowledge and experience:

$$\mu_{i,min} \leq \mu_i \leq \mu_{i,max}, \quad (29)$$

$$\sigma_{i,min} \leq \sigma_i \leq \sigma_{i,max}. \quad (30)$$

In this way, the consideration of infinite populations can be limited. In the given example, those limitations can be considered by adapting the integral limits of the outer two integrals. If only the range for the standard deviation is limited, (16) has to be considered. If only the mean range is limited, (17) is applied. In contrast to the standard deviation, the limitation of the mean ranges leads to interval limits that depend on $c_{2,1}$ and $c_{1,3}$. If both parameters are limited, only the intersections of the separately defined intervals have to be considered.

The parameter limitations can also be interpreted as a truncation of their distributions. Typically, this requires a standardization of the calculated certainty value. Let the intervals of the unchanged inner two integrals of the given example be named $I_{c_{2,1}} = [I_{c_{2,1},a}, I_{c_{2,1},b}]$ and $I_{c_{1,3}} = [I_{c_{1,3},a}, I_{c_{1,3},b}]$ and the new limited intervals of the outer two integrals be named $I_{x_2} = [I_{x_2,a}, I_{x_2,b}]$ and $I_{x_1} = [I_{x_1,a}, I_{x_1,b}]$. Then, (28) can be rewritten as

$$p_{trend} = \sqrt{k_1 k_2} \cdot \int_{I_{x_1,a}}^{I_{x_1,b}} \int_{I_{x_2,a}}^{I_{x_2,b}} \int_{I_{c_{1,3},a}}^{I_{c_{1,3},b}} \int_{I_{c_{2,1},a}}^{I_{c_{2,1},b}} f_{\chi^2_{k_1-1}}(x_1) f_{\chi^2_{k_2-1}}(x_2) \cdot f_v\left(v_1(c_{1,3})\right) f_v\left(v_2(c_{2,1})\right) dc_{2,1} dc_{1,3} dx_2 dx_1. \quad (31)$$

The statistical certainty p^*_{trend} considering limited parameter ranges can now be calculated as

$$p^*_{trend} = \frac{p_{trend}\left(I_{c_{2,1}}=[c_{1,3},\infty), I_{c_{1,3}}=(-\infty,\infty)\right)}{p_{trend}\left(I_{c_{2,1}}=(-\infty,\infty), I_{c_{1,3}}=(-\infty,\infty)\right)}. \quad (32)$$

For both, nominator and denominator, the integration limits I_{x_2} and I_{x_1} stay constant. With limited ranges of $2.5 \leq \mu_i \leq 2.5$ and $0.8 \leq \sigma_i \leq 1.2$, the statistical certainty for a real trend can be determined as $p^*_{trend} = 77.91$ %. Compared to the integral with unlimited parameters, the certainty could be increased by 12.14 %. However, it has to be considered that a limitation of considered means and standard deviations does not automatically result into higher statistical certainties. The reason is that with limiting the parameters, the share of possible infinite populations that would fulfill (18) can be increased or decreased. For example, if the range for the standard deviations is further limited to $0.8 \leq \sigma_i \leq 0.9$ and the limitation of the mean is not changed, $2.5 \leq \mu_i \leq 2.5$, the statistical certainty decreases to $p^*_{trend} = 75.78$ %.

3.2 Statistical certainties of other forecasts

The quantification approach is also applicable to other forecasts, e.g. based on run and out-of-control patterns. Let 6 values that are all either above or below the mean shown in the control chart, describe a run pattern (9 values according to [11]). As z-values are expected to follow a standard normal distribution, the mean in a z-chart is zero. If the last 6 values in Figure 1 showed a run where all values lied above the mean, the requirement given by (18) would change to

$$c_{1,4} > 0 \wedge c_{2,4} > 0. \quad (33)$$

In this way, it can be ensured that the smallest values of each feature taking part in the run are above the mean. Considering (17), it can be seen that the integration over density functions of Student t-distributions is required. The denominator of the z-value with the χ^2-distributed variable can be neglected as it is always bigger than zero.

An out-of-control pattern is detected if one value can be found that either lies above the upper control limit or below the lower control limit of the chart. The lower and upper control limits are defined as $L_{CL} = -3$ and $U_{LC} = 3$ [14]. If the last value show in Figure 1 lied above U_{LC}, the requirement given by (18) would change to

$$c_{2,6} > 0. \quad (34)$$

To apply the quantification approach to other patterns, similar procedures are required.

4 CONCLUSIONS

This paper proposed a first approach to quantify statistical certainties of quality forecasts for grouped product characteristics and process parameters. Quality forecasts are derived from widely used patterns that can be found in z-charts. In other words: Applying experts receive an impression about the credibility gap of quality forecasts suggested by quality charts in Short-Run SPC compared to classical SPC.

By the optional use of expert knowledge and experience, limited ranges of means and standard deviations for possible infinite populations can be defined. In this way, only those possible populations that are seen as realistic are considered in the calculation.

As resulting terms can be complex, software support for calculation is required. The quantification of statistical certainties claims that all product characteristics or process parameters being part of a detected pattern provide at least two samples before the pattern has appeared. This is necessary to derive the empirical means and standard deviations of their normal distributions.

In future works, it will be investigated if the presented approach can also be adapted to sample values that do not follow normal distributions. Also, the applicability in other types of control charts will be examined. It can be estimated that statistical certainties of quality forecasts can be used to evaluate and optimize the group formation. This will be a part of further studies.

5 REFERENCES

[1] Reinhart, G. (2017). *Handbuch Industrie 4.0*. München: Carl Hanser Verlag, p.557.

[2] Pyzdek, T. and Keller, P. (2018). *The Six Sigma Handbook*. 5th ed. New York: McGraw-Hill Education, p.368, 444, 446.

[3] International standard ISO 7870-2 (2013). *Control charts - Part 2: Shewhart control charts*. Switzerland: International Organization for Standardization, p.15.

[4] International standard ISO 7870-5 (2014). *Control charts - Part 5: Specialized control charts*. Switzerland: International Organization for Standardization, p.9-10.

[5] MacNeil, D. (2013). *Fundamentals of Modern Mathematics*. Mineoly, New York: Dover Publications, p.115.

[6] International standard ISO 7870-8 (2017). *Control charts - Part 8: Charting techniques for short runs and small mixed batches*. Switzerland: International Organization for Standardization, p.1, 5.

[7] Al-Salti, M. and Statham, A. (1994). The Application of Group Technology Concept for Implementing SPC in Small Batch Manufacture. *International Journal of Quality & Reliability Management*, 11(4), pp.64-76.

[8] Zhu, Y., Wong, Y. and Lee, K. (2006). Framework of a computer-aided short-run SPC planning system. *The International Journal of Advanced Manufacturing Technology*, 34(3-4), pp.362-377.

[9] Lin, S., Lai, Y. and Chang, S. (1997). SHORT-RUN STATISTICAL PROCESS CONTROL: MULTICRITERIA PART FAMILY FORMATION. *Quality and Reliability Engineering International*, 13(1), pp.9-24.

[10] Jiang, X., Zhao, K., Tang, L. and Wang, W. (2008). Research on multi-type & small batch oriented process quality control system under network environment. *2008 Asia Simulation Conference - 7th International Conference on System Simulation and Scientific Computing*.

[11] Wiederhold, M. (2017). *Clustering of Similar Features for the Application of Statistical Process Control in Small Batch and Job Production*. Aachen: Apprimus Verlag.

[12] Nelson, L. (1984). The Shewhart Control Chart—Tests for Special Causes. *Journal of Quality Technology*, 16(4), pp.237-239.

[13] Mandenius, C. and Titchener-Hooker, N. (2013). *Measurement, Monitoring, Modelling and Control of Bioprocesses*. Berlin, Heidelberg: Springer Berlin Heidelberg, p.232.

[14] Goos, P. and Meintrup, D. (2016). *Statistics with JMP*. Chichester: Wiley, p.232.

[15] McShane-Vaughn, M. (2016). *The probability handbook*. Milwaukee, Wisconsin: ASQ Quality Press, p.112.

6 BIOGRAPHY

Kevin Nikolai Kostyszyn holds two M.Sc. degrees in Automotive Engineering and Transport, and Mechanical Engineering from RWTH Aachen University and Tsinghua University, Beijing. Since 2015, he is a research assistant at Fraunhofer Institute for Production Technology IPT in Aachen.

Robert Schmitt obtained his doctorate in Metrology and Communications Engineering from RWTH Aachen University. In 2004, he was appointed Professor at RWTH Aachen University and became head of the Chair for Metrology and Quality Management at Laboratory for Machine Tools and Production Engineering (WZL).

Predicting Acceptance of Novel Technology from Social Network Data - An Agent-based Simulation-Approach

A. C. Valdez[1], M. Ziefle[1]
[1]RWTH Aachen University, Human-Computer Interaction Center
Aachen, Germany

Abstract
With digitization of production technology changes in technology infrastructure will become more frequent and more important for the competitiveness of organizations. Here, a crucial factor lies in the acceptance of such novel technology by users. Technology acceptance models aim to predict the adoption of new technology in an organization. They are however static in nature and fail to capture to dynamic process of adoption. To overcome this limitation, we utilize quantitative data from a small organization to understand both acceptance patterns and social structure of the organization. Both are used in an agent-based simulation to predict acceptance integrating social effects of diffusion over time. Our simulation achieves very similar results as the quantitative real-world data.

Keywords
Technology Acceptance; Social Network Analysis, Agent-Based Modelling, Adoption Prediction

1 INTRODUCTION

The integration of digitization in production technology has many different names: Industry 4.0, Internet of Things, Internet of Production. The common ground of these terms is the understanding that a key resource in production and manufacturing lies in the understanding of data that is generated in production processes. Utilization of such data in the form of data science, smart agents, decision support systems, smart logistics and AI-based automation promises the optimization of production processes by integrating various uncertainties. However, still many companies, especially SME are unprepared for total digitization. Both technological and social resources are missing to cope with the disruptive changes of the coming years. In today's rapidly changing technological environments, adapting to new technologies quickly is a crucial organizational skill both from a technological infrastructure perspective and from a human resources perspective. While the first is often addressed in research, the latter is far less understood.

Here, human resources required for total digitization in production refers to staff that has adequate attitudes, skills, etiquette, culture, and acceptance. Workers and engineers must approve of digitization strategies. They must have the technological capacities to implement or communicate necessary changes. Further, they must have a holistic understanding of the accompanying different work practices that are connected to digital processes (e.g., digital communication etiquette, sharing culture, privacy culture, etc.). Lastly, changes in working infrastructure and thus work processes must find acceptance in both workers and engineers. Without social acceptance of these novel technologies, introduction of said technologies will be silently sabotaged undermining objective performance measurement. The evaluation through KPIs ultimately yields only numbers masked by social practices and acceptance.

In this article we investigate social acceptance by trying to simulate the introduction process of novel software using an agent-based model. Based on one of the most popular and recent technology acceptance models the *Unified theory of acceptance of technology* (UTAUT) we measured psycho-social predictors of acceptance in a real working group. We then measured acceptance of a digital communication tool (i.e. Slack). Based on sociometric data of the working group (n = 20) we modelled social influence using a graph-based representation—the underlying social network. We then simulated the decision-making process in an agent-based model utilizing the previously established psycho-social predictors and underlying social-network structure. Lastly, we compared the individual simulated acceptance with the real-world measured acceptance and found very low deviance and similar acceptance patterns.

2 SIMULATING TECHNOLOGY ADOPTION

In order to understand the technology adoption process and establish a simulation model we first have to look at existing technology acceptance models. We then look identify various methods to establish the underlying social-network structure from different data source. Lastly, we need to establish a simulation approach that allows incorporating both inter-individual differences and social network information.

2.1 Technology acceptance models

Research on technology acceptance has had a long history in social psychology and the decision sciences. The earliest roots can be traced back to theories of understanding voter behavior in the 1960s in the theory of reasoned action [1]. The underlying assumption of this model is that both attitudes and norms play a role in accepting a social practice, a political candidate or a novel technology. Later models incorporate effects of behavioral control (e.g., theory of planned behavior [2]). More modern models specifically focus on the adoption of information systems and even formalize the diffusion of an innovation in society (theory of the diffusion of innovations [3]).

Nevertheless, these models formulate social processes by establishing user typologies and their characteristics. If the set of "early adopters" is convinced to use a product, the next phase in diffusion (majority adoption) is started and so on. This perfectly addresses large-scale social acceptance questions but misses to predict individual acceptance in individual organizations. Here the different revisions of the technology acceptance model (TAM [4]) triumph. By modelling individual psychosocial antecedents of social acceptance these models predict the behavioral intention of a user. They predict the answer to the question: Do I intend to use the software?

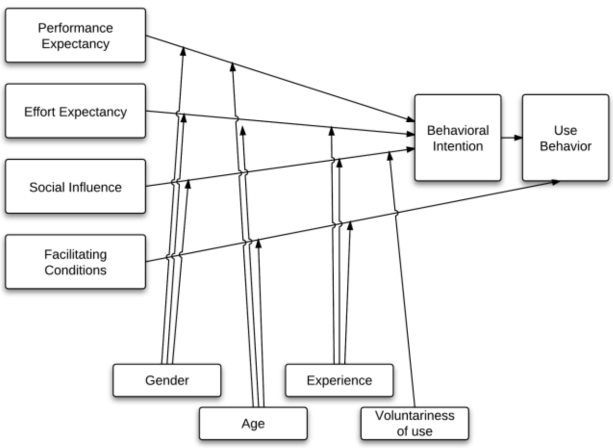


Figure 1 - UTAUT model visualization adapted from [5]

The most recent form of the model, the UTAUT [5] model (see **Error! Reference source not found.**), predicts the behavioral intention from four factors: performance expectancy, effort expectancy, social influence, and facilitating conditions. *Performance expectancy* refers to the users evaluation of whether they believe that the system will improve their work performance—will it improve my work? *Effort expectancy* refers to the users belief about the required effort to work with the system—is it easy to use? *Social influence* refers to the perceived acceptance of the system by other people that are important (e.g., bosses, colleagues). It addresses the question: Will I be the only one using it? *Facilitating conditions* refers to the infrastructural support of the system—will I get help when I run into trouble?

These four factors determine the behavioral intention. Some of them are moderated by additional inter-individual differences (age, gender, experience, voluntariness). The UTAUT model is applied by measuring the antecedents of behavioral intention and then estimating the individual acceptance. This is achieved by using a hierarchical model of linear regressions (structural equation model). The coefficients of which have been established from extensive research in technology acceptance.

The drawback of this model is that the social influence for each individual is measured *only once* during the assessment. However, adoption is a dynamic process characterized by diffusion. Thus, social influence changes over time, when important persons in an organized change their individual acceptance over time (e.g., after training courses). This dynamic is insufficiently modelled in static acceptance models.

2.2 Measuring social network structures

In order to improve on the static nature of acceptance models, one core source of dynamics, the social infrastructure, can be modelled, integrated, and used in simulation models. Such simulation approaches should be able to predict shifts and changes in social acceptance in the social structure of an organization. But, how do we measure and access such structures. The science of sociometry is the study of relationships in groups by mapping individual perspectives of relationships onto self- and other-centric maps of social structure. Initially designed to understand dynamics in school classes, the method can has been extended to various other group dynamic processes [6]. The downside of sociometry is the extensive research effort required to measure adequate social structures. Often social relationships are asymmetrical—I like you, but you don't like me. Other relationships are unknown to the public. This requires extensive anonymization procedures to ensure trust towards the research procedure and the experimenters.

A different approach is to utilize proxy relationships to determine an image of the real social structure. Such proxy relationships can be collaborations in work projects, office co-occupation, attendance in meetings, etc. The proxy that we use in our research is the co-authorship on publications available to the public [7]. The benefit lies in the public availability of the data. The data does not necessarily capture to sole affective structure of the underlying social network—who likes whom. What it does, is capture a conjoint measurement of social coherence, organizational role, and organizational authority. All of which are factors typically relevant for social influence in the UTAUT model. From such data we can derive the structure of the underlying social network and even quantify the intensity of relationships from repeated co-authorships. Whether such data has sufficient quality for simulation modeling, is part of the research question in this article.

2.3 Agent-Based Modelling

Assessing both antecedents of acceptance and the underlying social network gets us only so far. Individual decision-making on acceptance and on observing the decisions made by colleagues are hard to formalize in closed-form representations or formulas. Technically, each individual conducts a set of matrix multiplications to derive the individual behavioral intention. However, these calculations have to be conducted iteratively multiple times until the state stabilizes. As soon as each individual incorporates a mental model of the opinion of others or starts learning, closed-form representations break down [8].

For such cases, agent-based models have been proven both successful in replicating real-world data and successful in communicating results to laymen [9]. The central idea of agent-based modeling lies in programmatically modeling the individual as a template or *agent* and letting the independent agents make their own decisions based on their perception of the environment. In our case agents each have individual perceptions of the environment—in our case the other agents and the underlying social network. From these perceptions they derive their own behavioral intention, possibly influencing the neighbors in the social network in the next iteration of the run. By analyzing the outcome of several of such simulations, probability of organizational acceptance can be derived.

Agent-based models are often designed in specialized software toolkits (e.g., Netlogo [10]). These toolkits simplify formulation of agent behavior and include interfaces for visualizing simulation states, interacting with simulation parameters, and exporting simulation results. As additional tools, they provide the means to run simulations in batches and to search for optimal parameter configurations using different optimization strategies such as genetic algorithms.

3 METHOD

In order to study the effectiveness of agent-based models for acceptance research, we first conducted a survey containing all UTAUT measurement variables in a scientific working group (n = 20). These measurements were directed at the use of the software "slack"[1]. The items of the UTAUT scale can be found in the original work by Venkatesh et al. (cite). All measurements were taken on a six-point Likert scale. From each individual response we derived an agent for our agent-based model (see Figure 2). The agent-based model was generated in

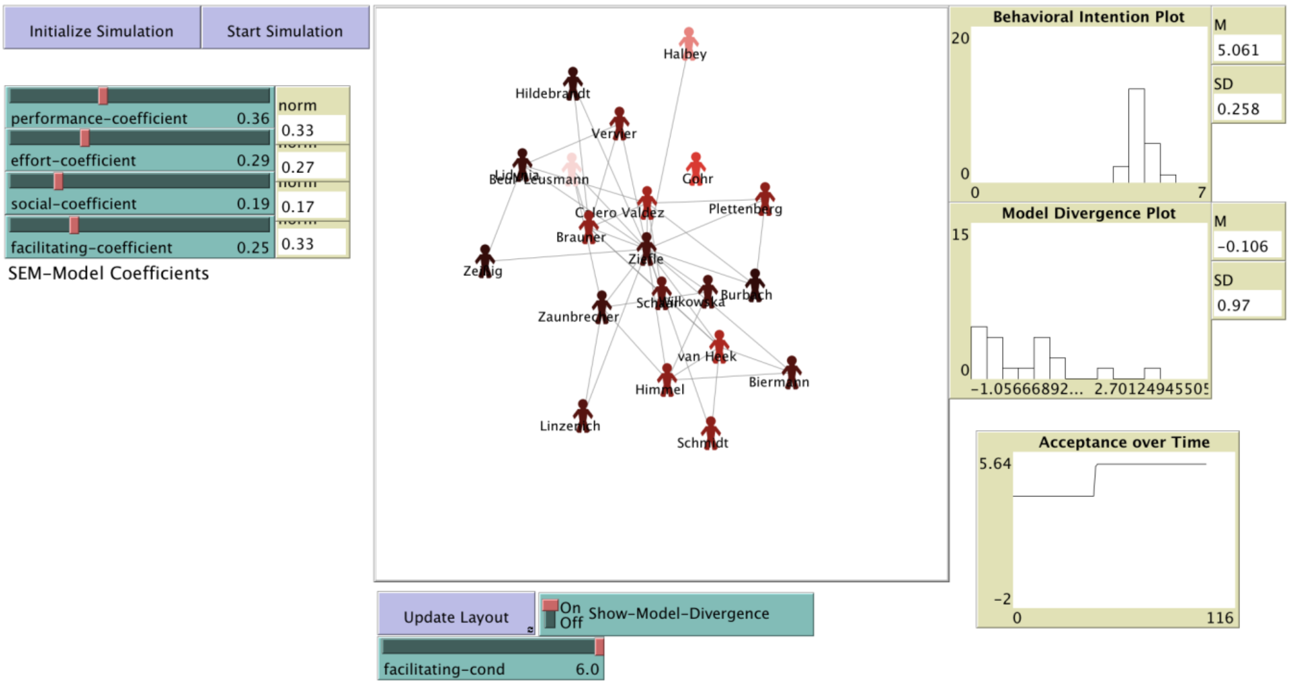


Figure 2 - Netlogo model used in this article. The left-hand side shows the coefficients used in the model, including a normalization. The center shows a graphical representation of the model. The right shows model evaluation parameters for an individual run.

performance, effort and the facilitating conditions. The perception of social influence is generated from

[1] https://slack.com/ - Slack is a social collaboration software suite.

Netlogo 6.0.1—an easy to use simulation software [10].

The survey was *not* anonymized to enable connecting the data with a social network structure. The social network structure was derived from the central publishing repository of the group, creating a connection between all agents, that were co-authors on any publication. Repeated co-authorship was not evaluated. All authors not present in the survey data were removed from the agent-based model. The target variable (i.e., behavioral intention) was modelled as the weighted sum of the variable's performance expectancy, effort expectancy, facilitating conditions, and social influence. While the first three were established from survey data, the social inclusion variable was determined as the averaged behavioral intention of all connected users. The behavioral intention of all agents was initialized as 3.5 which signifies a neutral stance.

As free parameters for the experiments the coefficients of the aforementioned weighted sum were selected. To ensure that the outcome variable stays within the original measurement range, all coefficients were normalized to add up to one. Coefficients can be chosen on a scale of 0 to 1 in 0.1 steps. As an additional free parameter, we let facilitating conditions be any integer on the scale of 1 to 6 to simulate different support conditions.

Using the *behavior space* simulation tool, we generated 161,051 parameter constellations with a ten-fold validation. This number is the number of unique *normalized* free form parameters from a total of 600,000 possible simulations. The random seed was uninitialized to allow for the influence of randomness in the model.

In each simulation run we measure the divergence of the simulated model from the real-world behavioral intention as measured in our survey. We do this for each individual agent and keep track of the mean model divergence and the standard deviation of the model divergence. For visual inspection, we implemented two visualizations, one to show the simulated behavioral intention, another to show the model divergence.

4 RESULTS

In order to see how the simulation results fare against the survey methodology we first look at the results of the survey and then compare the findings with the simulation results.

4.1 Survey Results

The survey of 20 participants, yielded 14 female and 6 male participants. The average age of participants was 32.7 years (*SD* = 7.6). The participants reported a high behavioral intention to use the software (*M* = 5.17, *SD* = 0.98). The software was seen as rather easy to use (*M* = 5.67, *SD* = 0.33) and was considered to be rather neutral with regard to increasing work performance (*M* = 3.91, *SD* = 0.67). The participants rated the social influence rather positively (*M* = 4.38, *SD* = 0.85) and found that the software was well supported (*M* = 5.35, *SD* = 0.52).

4.2 Simulation Results

Total simulation runtime was approximately 20 minutes. Each simulation reached the steady state in about 7 iterations. The output of the simulation extended to about 400Mb of comma separated values of free parameters and all tracked model parameters.

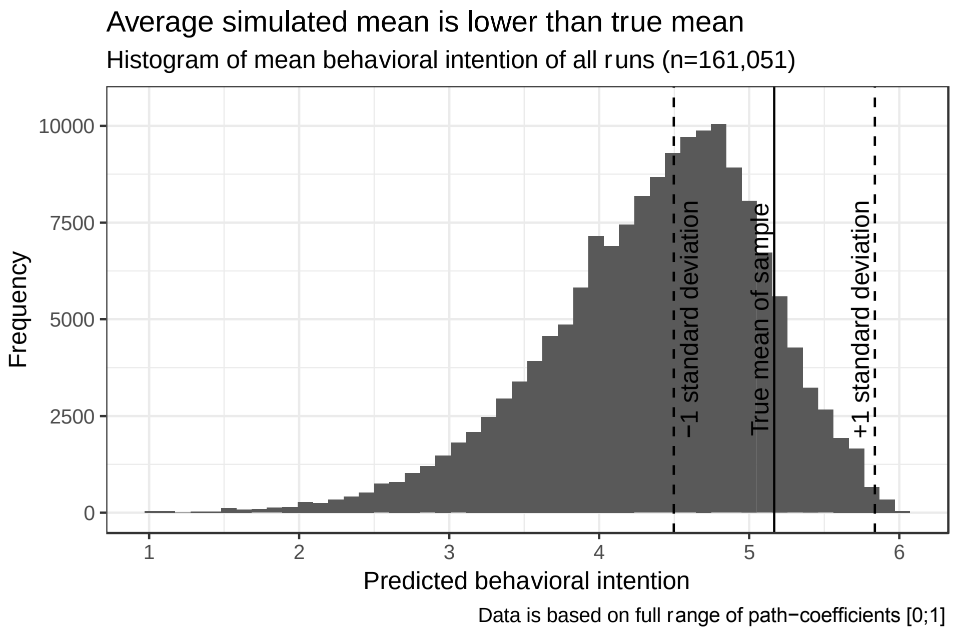


Figure 3 – Results distribution of all 161,051 simulations. The green area indicates the range of one standard deviation around the true mean of our sample.

We use the software R to analyze the resulting data. Using multiple linear regression analyses, we find that from our simulation, effort expectancy and social influence are the most determining factors for this particular working group (F(4, 161,046) = 1651, p < .001, adj. r^2 = .29, see Table 1).

Table 1 – Multiple linear regression result table from simulation results

Coefficient	Estimate	SE	*t*-value	*p*
(intercept)	3.47	0.19	18.63	<.001
PE	0.30	0.19	1.6	.11
FC	-0.25	0.19	-1.347	.18
EE	2.64	0.19	14.12	<.001
SI	0.91	0.19	4.86	<.001

PE=Performance Expectancy, FC=Facilitating Conditions, EE=Effort expectancy, SI=Social Influence

When we look at the attained behavioral intention from all simulation runs, we see that the achieved behavioral intention is on average slightly lower than the true sample mean, which was measured by the survey (see Figure 3). This means that independent of the coefficients of the model behavioral intention is strongly determined by the perceived low effort expectancy in this group. The mean model divergence was -0.79 (SD = 0.73).

When we limit the free parameters of the model to the established parameters from literature (±0.1), we get a better fit of results (see Figure 4). Mean model divergence from the data now only yields M = -0.37 (SD = 0.28). This indicates that the parameters established by Venkatesh et al. [5] in conjunction with our agent-based simulation help predict acceptance of novel software solutions.

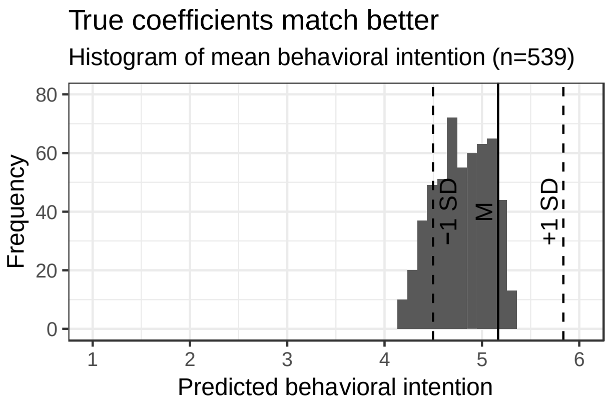


Figure 4 – Resulting simulations when the model coefficients are used from data

5 DISCUSSION AND LIMITATIONS

In this paper, we have simulated human decision making on the basis of the relatively simple UTAUT model. We surveyed twenty employees of a research group and modelled their virtual twins based on their survey results. We further used public co-authorship information to infer the underlying social network structure in the work group. The simulation of the acceptance of a collaboration software suite yielded similar results as the quantitative assessment of acceptance.

These results indicate that the agent-based model was able to *generate a similar prediction* as the structural equation-based approach. What was most interesting was that when looking at the individual network visualizations one could see that the users who reported the lowest behavioral intention still showed the lowest behavioral intention. The idiosyncrasies of the network were retained. Our approach was computationally not very challenging as the network was rather small in size.

The real-world data that we collected was a static snapshot of a long adoption process of two years. Since we did not collect this information regularly across time, we are unable to verify whether the adoption diffusion progressed similarly in the model as in the real world. Future research could try to *validate longitudinal* data—inherently available in the simulation—from simulation and the real-world.

The standard deviation of model divergence in some simulation runs was still rather large (between 0.9 and 1.1). It would be interesting to see, *who is responsible for these deviations*. Are individual users badly simulated or do some parameter constellations lead to overall bad results. Here, more detailed analyses would help in understanding where this simulation error can be attributed to.

So far, our model does not include a mental model for the individual agents. Thus, the *agents are unaware of trends*, or do not change their social network structure. Future research could investigate whether a model of trustworthy colleagues, whose social influence would then matter more, is able to predict adoption even better.

Our approach can, in theory, also be used to predict acceptance of novel technology in large organizations, given that a proxy for the social structure is available. Future research will have to *validate the simulation in larger organizational settings*.

6 CONCLUSION

Overall, we were able to simulate the diffusion process of acceptance in a small organization. The final outcome was relatively close to the observed real-world data. Our simulation was able to integrate dynamic behavior into a static model of technology acceptance. Applying such models can be useful in determining neuralgic users that are well connected and could drastically shift long-term adoption one way or the other. Understanding their motivation, will help in designing a support structure to persuade these key-users and gatekeepers, allowing organizations to adapt to the required changes of digitization more rapidly.

7 ACKNOWLEDGMENTS

This research was supported by the Digital Society research program funded by the Ministry of Culture and Science of the German State of North Rhine-Westphalia. Grant-Number: 005-1709-0006, project “Digitale Mündigkeit”, project-number 1706dgn017. The authors also thank the German Research Council DFG for the friendly support of the research in the excellence cluster “Internet of Production”.

8 REFERENCES

[1] Fishbein, M., & Ajzen, I. (2011). Predicting and changing behavior: The reasoned action approach. Psychology Press.

[2] Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes*, *50*(2), 179-211.

[3] Rogers, E. M. (2010). *Diffusion of innovations*. Simon and Schuster.

[4] Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: a comparison of two theoretical models. *Management science*, *35*(8), 982-1003.

[5] Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425-478.

[6] Moreno, J. L. (1951). Sociometry, experimental method and the science of society.

[7] Glänzel, W., & Schubert, A. (2004). Analysing scientific networks through co-authorship. In *Handbook of quantitative science and technology research* (pp. 257-276). Springer, Dordrecht.

[8] Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, *99*(suppl 3), 7280-7287.

[9] Kiesling, E., Günther, M., Stummer, C., & Wakolbinger, L. M. (2012). Agent-based simulation of innovation diffusion: a review. *Central European Journal of Operations Research*, *20*(2), 183-230.

[10] Tisue, S., & Wilensky, U. (2004, May). Netlogo: A simple environment for modeling complexity. In *International conference on complex systems* (Vol. 21, pp. 16-21).

9 BIOGRAPHY

André Calero Valdez is junior research group leader of the “Digitale Mündigkeit” (digital responsibility) research group he holds a PhD in Psychology from RWTH Aachen University and works as a senior researcher at the Human-Computer Interaction Center at RWTH Aachen University. His research deals with how users interact with technology.

Martina Ziefle is professor at the chair for Communication Science and founding member of the Human-Computer Interaction Center at RWTH Aachen University. Her research addresses human-computer interaction and technology acceptance in different technologies and using contexts, taking demands of user diversity into account.

Using a Discrete-event, Simulation Optimisation Optimiser to Solve a Stochastic Multi-objective NP-hard Problem

T. Bamporiki[1], J. Bekker[1], M. Yoon[1]
[1]Department of Industrial Engineering, University of Stellenbosch, South Africa

Abstract
This paper presents the use of the *cross-entropy method for multi-objective optimisation* (MOO CEM) metaheuristic and the *multi-objective moonyoung yoon* (MMY) procedure in a *decision-support system* (DSS) tool to solve multi-objective optimisation problems when simulating stochastic systems. The MOO CEM and MMY techniques are relatively new in the multi-objective optimisation field and have respectively shown quality results when used to solve bi-objective large-scale and small-scale stochastic optimisation problems. The DSS tool presented in this paper is a third-party optimiser, developed to solve discrete-event, large-scale stochastic problems using the MOO CEM, which is a *search algorithm*, and to guarantee the probability of correct selection of the results obtained by the MOO CEM using the MMY procedure, which is a *ranking and selection algorithm*. The tool is integrated with a commercial simulation software package and is used to solve the known NP-hard buffer allocation problem in a stochastic, multi-objective context.

Keywords
Optimiser, Multi-objective optimisation, Stochastic

1 INTRODUCTION

In this paper, two relatively new algorithms are presented and utilised as the backbones of a new *decision-support system* (DSS) tool developed to solve stochastic multi-objective optimisation problems.

Optimal decision making under uncertainty and in the presence of conflicting objectives (i.e. multi-objectives) is an important area of study in industrial engineering. Many problems in this area of study can be formulated under the following framework

$$\text{Minimise/Maximise } E[f_i(\boldsymbol{x},\xi)]$$
$$\text{Subject to}$$
$$\boldsymbol{x} \in S;$$

Where the expression $f_i(\boldsymbol{x},\xi), i = 1,\ldots,m$ represents different values that objective i can take on when solution $\boldsymbol{x}$ is selected in the presence of stochastic element ξ, which is responsible for the uncertainty or stochastic behaviour in the problem. Solution $\boldsymbol{x} \in S$ is a vector of decision variables values and S is the feasible solution space. When S is large such that all $\boldsymbol{x} \in S$ cannot be evaluated practically, the problem can be referred to as a large-scale problem, whereas when it is small such that all $\boldsymbol{x} \in S$ can be evaluated practically, the problem can be said to be a small-scale problem [1]. $E[f_i(\boldsymbol{x},\xi)]$ is the expected value of objective i. Because it is difficult to obtain the true value of $f_i(\boldsymbol{x},\xi)$ due to ξ, an estimate $E[f_i(\boldsymbol{x},\xi)]$ that can be obtained after a 'sufficient' number of N simulation replications (or observations) are made is used instead. Consider the notation $f_{ij}(\boldsymbol{x},\xi)$ where $j = 1,\ldots,N$ represents the jth observations made for objective i, then

$$E[f_i(\boldsymbol{x},\xi)] = \left(\frac{1}{N}\right)\sum_{j=1}^{N} f_{ij}(\boldsymbol{x},\xi) \qquad (1)$$

can be considered as an estimated value for objective i.

It is known that *multi-objective optimisation* (MOO) problems have more than one optimal solution. These are often referred to as *Pareto optimal solutions* (S_P) or non-dominated solutions [2] [3]. Goldberg [4] developed a Pareto ranking algorithm that finds S_P with respect to a user-defined threshold t_h, when given a set of M solutions and their respective objective values. t_h is an integer value that allows the algorithm to include in S_P, $\boldsymbol{x} \in S$ that is/are dominated by, at most, t_h solutions in S. The algorithm in [4] is used by both the algorithms presented in this paper.

To stay competitive under these conditions, decision makers often rely on state-of-the-art decision-support system tools. Among these tools are commercial *discrete-event simulation* (DES) products that have, as part of their software package, integrated built-in or third-party optimisers.

Optimisers (in this context) are tools that are generally used to assist DES products when solving large-scale problems. Third-party optimisers such as OptQuest [5] (a popular commercial product used in over 10 existing DES products [6]), are powerful tools in practice that utilise robust as well as smart solution algorithms [7] [8]. OptQuest, in particular, has been praised by many researchers for its quality results [6] [9]. However, these tools are also known to lack the statistical rigour that is needed, in principle, when dealing with problems under stochastic conditions [10] [11]. In effect, when using an optimiser, the 'sufficient' number of N simulation replications is often selected 'blindly' or in a manner that is too conservative.

In this paper, a new optimiser for stochastic multi-objective optimisation problems is presented. The product, namely the *multi-objective optimisation solver* (MOOSolver), was developed and validated in [12]. MOOSolver uses relatively new MOO solution algorithms [13] [1] in the fields of metaheuristic and *ranking and section* (RS) and makes an attempt to incorporate rigorous statistical analysis in its solution approach. The purpose of this paper is to provide a comprehensive description of the MOO optimiser and to show the relevance of its solution approach to large-scale *stochastic multi-objective optimisation problems* (here after referred to as *multi-objective simulation optimisation* (MOSO) problems), by solving the well-known, NP-hard, *buffer allocation problem* (BAP).

The paper is structured as follows: Section 2 details the solution approach utilised by the optimiser, and in Section 3, we use the optimiser to solve the BAP. In Section 4, we conclude the paper.

2 PROPOSED SOLUTION APPROACH

The solution approach proposed in this paper is a two-phase *simulation optimisation* (SO) process and is as follows: The metaheuristic, which is responsible for the search mechanism, and the RS procedure, which is responsible for the statistical analysis; Work independently in sequential order. In phase one, the metaheuristic is used to search for 'good' solutions with respect to an arbitrarily selected $N \geq 10$. A sample of the results obtained by the metaheuristic is then used as input candidate solutions for the RS procedure in the second phase. The process is thus an interactive one with the user involved in selecting the candidate solutions to be input into the RS procedure. Giving the user the opportunity to first see all the good solutions found by the metaheuristic before the RS procedure takes over, allows the user to select candidate solutions from the approximate Pareto set (which can sometimes be overwhelmingly large), which they may have biased interest in. Only these preferred solutions undergo further, more rigorous statistical analysis.

In summary, the two-phase interactive approach reduces the large-scale problem to a small-scale one where all $\boldsymbol{x} \in S$ are known to be good solutions, thanks to the metaheuristic. Because the metaheuristic cannot guarantee the statistical soundness of the approximated Pareto set, the RS procedure is used to clean the potential 'noise' in the candidate solutions selected by the user by providing better estimates and *guaranteeing the probability of correct selection* P(CS). Hence, in the second phase of the solution approach, the *actual* sufficient number of N simulation replications for all $\boldsymbol{x} \in S$ is used, thanks to the statistical rigour of the RS procedure.

The following sections detail the metaheuristic and the RS procedure used in the solution approach.

2.1 The MOO CEM metaheuristic

The *multi-objective optimisation method using the cross-entropy method* (MOO CEM) is a multi-objective adaptation of the cross-entropy method metaheuristic [9]. The algorithm was developed and validated in [13]. The MOO CEM is the principal algorithm of MOOSolver. The description of the metaheuristic presented here is based on the work in [9] and [13].

The MOO CEM works as follows: The algorithm is an iterative, population-based metaheuristic [7]. New populations are generated by creating a sequence of *probability density functions* (pdfs) for each *decision variable* (DV) in every iteration. To form a sample vector $\boldsymbol{X}_i$ from a pdf $\boldsymbol{H}_i(\cdot, \hat{\boldsymbol{v}}_{t-1})$ in iteration t, where $\hat{\boldsymbol{v}}_{t-1}$ is the pdf parameters in iteration $t-1$, a truncated normal distribution is used for each decision variable. For the n DVs defined over ranges $[l_i, u_i]$, l_i is the lower limit and u_i the upper limit of DV x_i, $1 \leq i \leq n$. The truncated normal distribution $\boldsymbol{H}_{\boldsymbol{T}i}$ defined in the range $[l_i, u_i]$ and having mean μ_i and variance σ_i^2, is given by

$$\begin{cases} 0, & x < l_i, \\ \dfrac{H_n(x)}{\int_{l_i}^{u_i} H_n(x)\mathrm{d}x}, & l_i \leq x \leq u_i, \\ 0, & x > u_i. \end{cases}$$

The function $H_n(x)$ is the normal pdf defined on $-\infty < x < \infty$.

Next, each of the objective functions is evaluated using the solutions found (i.e. the members of the population). The best combinations of objective functions are found by doing a Pareto ranking. The values of the decision variables in the non-dominated set provided by the Pareto ranking algorithm are used to construct a histogram for each decision variable. The histograms provide guiding information for the MOO CEM algorithm and are maintained while the algorithm is searching for non-dominated solutions.

The concept works as follows: For a decision variable that is defined in the range $[l_i, u_i]$, the lower boundary of the first class is set equal to l_i, and the upper boundary of the last class is set equal to u_i. Next, the upper boundary of the first class is set equal to the minimum value of the decision variable

x_i in the non-dominated set, i.e. $\min(S_P(\cdot,i))$. The lower boundary of the last class, on the other hand, is equal to the maximum value of the decision variable in S_P, namely, $\max(S_P(\cdot,i))$. A number of equal-sized classes are then formed between these two boundaries using $\left(\max(S_P(\cdot,i)) - \min(S_P(\cdot,i))\right)/r$ if r of these classes are formed, resulting in a total number of $r + 2$ classes (see Figure 1).

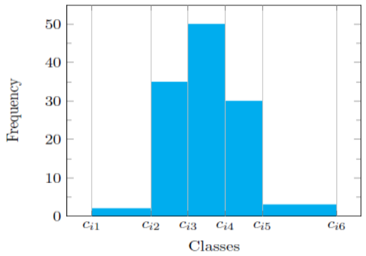


Figure 1 - Example of a histogram for decision variable x_i and $r = 3$

The class boundaries for the histogram of decision variable x_i are recorded in a vector $\boldsymbol{C}_i = \{c_{i1}, c_{i2}, \ldots, c_{i(r+2)}, c_{i((r+2)+1)}\}$, with $c_{i1} = l_i$ and $c_{i((r+2)+1)} = u_i$. Note that $\mathbf{C}_i$ contains $r + 3$ elements because the histogram has $r + 2$ classes, and that the class widths of the first class (i.e. $[c_{i1}, c_{i2}]$) and the last class (i.e. $[c_{i(r+2)}, c_{i((r+2)+1)}]$) can be different from each other and from the widths of the r classes.
The decision variable values are classified according to the following rule: x_{ij} belongs to the class $[c_{iw}, c_{i(w+1)})$ if $c_{iw} \le x_{ij} < c_{i(w+1)}$, $1 \le w \le r + 2$. The histogram frequency values are recorded in a vector $\boldsymbol{R}_i = \{\tau_{i1}, \tau_{i2}, \ldots, \tau_{i(r+1)}, \tau_{i(r+2)}\}$, where τ_{i1} is equal to the frequency count of decision variable x_i in the range $[c_{i1}, c_{i2})$, τ_{i2} represents the count in range $[c_{i2}, c_{i3})$, and so on.
In preparation for the next iteration of the algorithm, the new population of possible solutions is formed proportionally according to the class frequencies for each decision variable.

To ensure exploitation in the MOO context, the process described above is repeated o_l times as an outer loop of the algorithm. After each loop, the elite set S_P is ranked again and the number of classes of the histograms is incremented. Increasing the number of classes as the search progresses makes it possible to maintain good combinations of decision variable values as the resolution of the decision variable spaces becomes finer. The parameter vectors (μ_i, σ_i) are updated using

$$\hat{\mu}_{it} = a\tilde{\mu}_{it} + (1-a)\hat{\mu}_{i(t-1)} \quad (2)$$
$$\hat{\sigma}_{it} = a\tilde{\sigma}_{it} + (1-a)\hat{\sigma}_{i(t-1)} \quad (3)$$

where t is the iteration count index and $a = 0.7$ in all cases. This process is continued until the σ_i value of each decision variable has decreased below a common threshold *epsilon* (ϵ). On algorithm termination, S_P should contain a set of solutions that is a *good* approximation of the true Pareto set.
To support exploration and exploitation of the search, the initial ranking threshold t_h is relaxed and a value of two is selected. When a new loop starts and a new population is formed, the elite matrix is trimmed and the threshold is set to one. When the algorithm terminates, the existing elite matrix is refined a final time, and the threshold then used is zero, which means all solutions selected are non-dominated. The algorithm is presented in pseudo-code as Algorithm 1.

Algorithm 1 MOO CEM metaheuristic

1: Set $S_P = \emptyset$, $t = 1$, $o_l = 1$.
2: Initialise decision variable vectors $\boldsymbol{X}_i = \emptyset$, $1 \le i \le n$.
3: For each decision variable x_i, $1 \le i \le n$, initialise a histogram class vector $\boldsymbol{C}_i = \{c_{i1}, c_{i2}, \ldots, c_{i(r+2)}, c_{i((r+2)+1)}\}$ and histogram frequency vector $\boldsymbol{R}_i = \{\tau_{i1}, \tau_{i2}, \ldots, \tau_{i(r+1)}, \tau_{i(r+2)}\}$.
4: Set $i = 1$.
5: Set $w = 0$.
6: Increment w.
7: Do for frequency element τ_{iw} in $\boldsymbol{R}_i$.
8: Generate a class-based $\hat{\boldsymbol{v}}$ in the range $[c_{iw}, c_{i(w+1)})$, $1 \le w \le r + 2$.
9: Generate a subsample $\boldsymbol{Y}$ according to pdf $\boldsymbol{H}_{\boldsymbol{T}i}(\boldsymbol{x}_i, \hat{\boldsymbol{v}})$ with $\boldsymbol{X}_i \in [c_{iw}, c_{i(w+1)})$ and $|\boldsymbol{Y}| = \tau_{iw}$, $1 \le w \le r + 2$.
10: Append $\boldsymbol{Y}$ to $\boldsymbol{X}_i$.
11: If $w < r + 2$ return to Step 6.
12: Increment i.
13: If $i \le n$, return to Step 5.
14: Rank the performance values using the Pareto ranking algorithm with a relaxed $t_h = 2$ to obtain an updated S_P.
15: Form new histogram class vectors $\mathbf{C}_i$ and histogram frequency vectors $\boldsymbol{R}_i$ based on S_P, $1 \le i \le n$.
16: Use the values in S_P and compute $\hat{\boldsymbol{v}}_{it}$ for all i, $1 \le i \le n$.
17: Smooth the vectors $\hat{\boldsymbol{v}}_{it}$ for all i, $1 \le i \le n$, using (2) and (3).
18: If all $\sigma_{it} > \epsilon$ or less than the allowable number of evaluations has been done, increment t and reiterate from Step 4.
19: Rank S_P with $t_h = 1$.
20: Increment o_l.
21: If o_l is less than the allowable number of loops, return to Step 2.
22: Rank S_P with $t_h = 0$ to obtain the final non-dominated set S_P.

2.2 The MMY procedure

The *multi-objective moonyoung yoon* (MMY) procedure is the first MOO RS procedure with the indifference-zone approach in the literature that guarantees correct selection following the Bayesian probabilistic theory [1]. The algorithm was selected to equip MOOSolver with a modern, efficient multi-objective RS procedure. Procedure MMY was developed and validated in [1].

The following notation applies: S is the feasible solution set, i.e. $S = \{1, \ldots, M\}$, I is the set of solutions that are still in competition, K is the objective set, i.e. $K = \{1, \ldots, m\}$, $\bar{X}_{ik}(N_i)$ is the sample mean of solution i for objective k based on N_i observations, S_P^c is the observed non-Pareto set based on $\bar{X}_{ik}$ ($i \in S$ and $k \in K$), n_0 is the number of simulation replications at the first stage, δ_k^* is the indifference-zone value for objective k and P^* the minimum required value for P(CS).

Following are some definitions used in the procedure. Let

$$\delta_{ijk} = max\{\delta_k^*, \bar{X}_{jk}(N_j) - \bar{X}_{ik}(N_i)\} \quad (4)$$

and $\lceil x \rceil$ denotes the smallest integer greater than x. Consider a pair of solutions (i,j) where system i is observed as non-dominated and system j can be any other solution in S. This pair (i,j) $\left(i \in S_p \text{ and } j \in S, j \neq i\right)$ is relevant to Steps 4 and 5 in Algorithm 2.

For each pair (i,j) $\left(i \in S_p \text{ and } j \in S, j \neq i\right)$, let

$$K_1 = \{k | \left|\bar{X}_{jk} - \bar{X}_{ik}\right| \leq \delta_k^*, k \in K\} \quad (5)$$

and

$$k' = \arg max_{k \in K} \Phi\left(\frac{\bar{X}_{jk}(N_j) - \bar{X}_{ik}(N_i)}{\sqrt{\frac{S_{ik}^2(N_i)}{N_i} + \frac{S_{jk}^2(N_j)}{N_j}}}\right) \quad (6)$$

where Φ denotes the cumulative distribution function (cdf) of the standard normal distribution. Note that K_1 and k' should be defined for every pair of (i,j) $\left(i \in S_p \text{ and } j \in S, j \neq i\right)$. Step 4 in Algorithm 2 deals with (i,j) pairs when $K_1 = K$, that is, solution i and j are observed to be indifferent to each other, while Step 5 considers the case when $K_1 \neq K$.

The constants h_1 (in Step 4) and h_2 (in Step 5) are obtained from the following equations respectively:

$$\left[\int_0^\infty \left[\int_0^\infty \Phi\left(\frac{h_1}{\sqrt{(N_i - 1)\frac{1}{x} + (N_j - 1)\frac{1}{y}}}\right) f_1(x)\mathrm{d}x\right] f_2(y)\mathrm{d}y\right]^m = 1 - \gamma, \quad (7)$$

and

$$\int_0^\infty \left[\int_0^\infty \Phi\left(\frac{h_2}{\sqrt{(N_i - 1)\frac{1}{x} + (N_j - 1)\frac{1}{y}}}\right) f_1(x)\mathrm{d}x\right] f_2(y)\mathrm{d}y = 1 - \gamma, \quad (8)$$

where $\gamma = \frac{\beta}{M-1}$ and f_1 and f_2 denote the pdf of the χ^2 (i.e. Chi-square) distribution with $N_i - 1$ and $N_j - 1$ degrees of freedom, respectively. The solution that dominates solution j with the maximum probability can be found with

$$i = \arg max_{i' \in S_P} \prod_{k=1}^{m} \Phi\left(\frac{\bar{X}_{jk}(N_j) - \bar{X}_{i'k}(N_{i'})}{\sqrt{\frac{S_{i'k}^2(N_{i'})}{N_{i'}} + \frac{S_{jk}^2(N_j)}{N_j}}}\right). \quad (9)$$

Note that such i should be defined for every $j \in S_P^c$. This pair of solutions (i,j) $\left(i \in S_p \text{ and } j \in S_P^c\right)$ is considered in Step 7 in Algorithm 2. The constant h_3 in the same step of the algorithm is obtained from

$$\left[\int_0^\infty \left[\int_0^\infty \Phi\left(\frac{h_3}{\sqrt{(N_i - 1)\frac{1}{x} + (N_j - 1)\frac{1}{y}}}\right) f_1(x)\mathrm{d}x\right] f_2(y)\mathrm{d}y\right]^m = 1 - \beta, \quad (10)$$

where $\beta = \frac{\alpha}{M}$ and f_1 and f_2 denote the pdf of the χ^2 distribution with $N_i - 1$ and $N_j - 1$ degrees of freedom, respectively. Note that P^* is kept constant with a typical value of 90%. The RS procedure is presented as Algorithm 2.

Algorithm 2 Procedure MMY

1: Select the probability of correct selection requirement $P^* = 1 - \alpha$, δ_k^* for each objective $k \in K$, and $n_0 \geq 10$. Set $I = \{1, \ldots, M\}$; And $\beta = \alpha/M$.

2: Simulate n_0 replications for all M solutions, and calculate $\bar{X}_{ik}(n_0)$ and sample variances $S_{ik}^2(n_0)$ ($i \in S$ and $k \in K$). Let $N_i = n_0$.

3: Observe S_P and S_P^c based on $\bar{X}_{ik}(N_i)$ ($i \in S$ and $k \in K$) using the Pareto ranking algorithm.

4: For each solution $i \in S_P$ and $j \in S$ ($j \neq i$) with $K_1 = K$, check if the following two conditions are met:

$$N_i > \left\lceil max_k \left({}^{h_1 S_{ik}(N_i)}/_{\delta_{ijk}}\right)^2 \right\rceil \text{ and } N_j > \left\lceil max_k \left({}^{h_1 S_{jk}(N_j)}/_{\delta_{ijk}}\right)^2 \right\rceil \quad (11)$$

where h_1 is obtained from (7), and K_1 is defined in (5).

5: For each solution $i \in S_P$ and $j \in S$ ($j \neq i$) with $K_1 \neq K$, check if the following two conditions are met:

$$N_i > \left\lceil \left({}^{h_2 S_{ik'}(N_i)}/_{\delta_{ijk'}}\right)^2 \right\rceil \text{ and } N_j > \left\lceil \left({}^{h_2 S_{jk'}(N_j)}/_{\delta_{ijk'}}\right)^2 \right\rceil \quad (12)$$

where k' is the defined in (6), and h_2 is obtained from (8).

6: Delete solution i from I if conditions (11) or (12) are satisfied for all $j \in S$ ($j \neq i$).

7: For each solution $j \in S_P^c$, and solution $i \in S_P$ as defined in (9), check if the following two conditions are met:

$$N_i > \left\lceil max_k \left({}^{h_3 S_{ik}(N_i)}/_{\delta_{ijk}}\right)^2 \right\rceil \text{ and } N_j > \left\lceil max_k \left({}^{h_3 S_{jk}(N_j)}/_{\delta_{ijk}}\right)^2 \right\rceil \quad (13)$$

where h_3 is obtained from (10).

8: Delete solution j from I if conditions in (13) are satisfied.

9: If $|I| = 0$, then stop and present the current S_P as the final solution set. Otherwise, for each solution $i \in S_P \cap I$, that is, solutions in S_P that were not deleted from I in Step 6, add solution $j \in S$ $(j \neq i)$ to I if it does not satisfy conditions (11) or (12). Similarly, for each system $j \in S_P^c \cap I$, that is, solutions in S_P^c that were not deleted from I in Step 8, add the corresponding solution $i \in S_P$ to I if it does not satisfy (13). Go to Step 10.

10: Take one additional observation $X_{ik}(N_i + 1)$ from each solution $i \in I$, and set $N_i \leftarrow N_i + 1$ $\forall_i \in I$. Set $I = \{1, \ldots, M\}$ and update $\bar{X}_{ik}(N_i)$ and $S_{ik}^2(N_i)$ for all $i \in S$ and $k \in K$. Go to Step 3.

2.3 The MOOSolver optimiser

MOOSolver is a third-party optimiser that uses the MOO CEM and the MMY in its solution approach to user-defined MOSO problems of the form of the problem framework presented in Section 1. The optimiser is a callable module that can be integrated with discrete-event simulation software packages using the *inter-process communication* (IPC) technology in the C programming language.

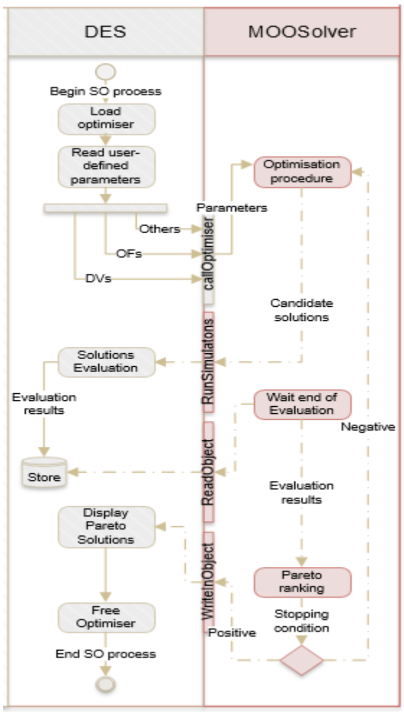


Figure 2 - IPC procedure between MOOSolver and the DES product

Currently, the optimiser has been successfully integrated with Siemens' Tecnomatix Plant Simulation [12].

Once the optimiser is successfully integrated and ready to be used, the user must provide the necessary parameters (see Section 3.1) of the simulation model to be solved. The flow diagram in Figure 2 illustrates the concept and shows the inter-process communication procedure between MOOSolver and the discrete-event simulation software package. When the SO process is complete, the optimiser outputs the results of each phase of the solution approach in table formats; These are illustrated as Tables 1 and 2 for phase one and phase two, respectively. Note that since the results from phase one can be very large, they can be better represented as a plot of the objective values for every solution in the approximate S_P, when $m = 2$. The plot is generally referred to as the Pareto front. Also note that T in phase one indicates the total number of solutions found by the metaheuristic i.e. the total number of elements in the Pareto front.

Solutions	Objectives		
$\boldsymbol{x}_1$	f_{11}	$\cdots$	f_{1m}
$\vdots$	$\vdots$		$\vdots$
$\boldsymbol{x}_T$	f_{T1}	$\cdots$	f_{Tm}

Table 1 - MOOSolver output table format for phase one

Solutions	Objectives			Variances			Rank	Runs	ID
$\boldsymbol{x}_1$	f_{11}	$\cdots$	f_{1m}	S_{11}^2	$\cdots$	S_{1m}^2	ρ_1	N_1	1
$\vdots$	$\vdots$		$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\boldsymbol{x}_M$	f_{M1}	$\cdots$	f_{Mm}	S_{M1}^2	$\cdots$	S_{Mm}^2	ρ_M	N_M	M

Table 2 - MOOSolver output table format for phase two

S_{ij}^2 is the variance information of objective function j of solution i, $j = 1, \ldots, m$. ρ_i is the Pareto rank of solution i indicating the total number of solutions in the set that dominate (i.e. are better than) solution i. N_i is the total number of simulation replications run for solution i.

In the following section, we use the optimiser to solve the buffer allocation problem.

3 THE BUFFER ALLOCATION PROBLEM

Production systems are often organised with machines connected in series and separated by buffers. This arrangement is often called a flow line or a production line [14]. The buffer allocation problem is a well-known problem in the design of production lines.

The basic setting of the BAP is the following [9] [15]. Consider a production line consisting of p machines in series, numbered $1, 2, \ldots, p$. Jobs are processed by all machines in sequential order. The processing times at all machines have, generally, a fixed exponential distribution with rate $\mu_i, i = 1, 2, \ldots, p$. The machines are assumed to be unreliable and the machine failures are *operation-dependent failures* (ODFs) with Poisson distributions having parameters λ_i. The repair times are exponential with parameters β_i. All repair, failure and processing times are assumed to be independent of each other.

The machines are separated by $p-1$ storage areas or niches, in which jobs i.e. *work-in-progress* (WIP) can be stored. The total number of storage spaces, or *buffer spaces*, is unknown and must be minimised. The required number of buffer spaces can be determined by estimating the average WIP rate.

The decision variables are the elements of the solution vector $\boldsymbol{x} = (x_1, \dots, x_{p-1})$, where x_i is the number of buffer spaces at niche i, $i = 1, \dots, p-1$. The objective functions in the stochastic, MOO context are throughput $T_R(\boldsymbol{x}, \xi)$ (to be maximised) and the WIP (to be minimised), denoted by $W_P(\boldsymbol{x}, \xi)$. ξ is the stochastic element in the problem caused by the machines' failures, as well as repair and processing times.

3.1 Specifics of the BAP solved

It was assumed for this variant of the BAP [1] that a total of q buffer spaces were available to be arranged in four niches (i.e. the number of machines in series in this case is $p = 5$. Every machine has a niche in front of it, except the first machine in the line). The problem therefore had a total number of $\binom{q+p-2}{p-2}$ candidate solutions. With q selected as 6 for the problem, the total number of potential solutions amounted to $\binom{9}{3} = 84$. (The BAP is NP-hard in the sense that the number of candidate solutions increases relatively fast for relatively small increases in q. For instance, $q = 10$ has a solution space that contains 286 elements, while for $q = 20$ the solution space contains 1771 elements.) From the 84 potential solutions, 10 were selected randomly to form the solution space S and analysed by the MMY procedure. Although the purpose for solving this BAP in [1] was simply to validate the results obtained by the MMY procedure over the small set of randomly selected solutions (and not to try to find the actual best solutions over the entire solution space), it is nonetheless assumed, for the purpose of this paper, that the approach used to solve the BAP in [1] is ineffective as the solution space is not explored in its entirety. The goal here is, therefore, to demonstrate the relevance of the solution approach proposed in Section 2 by solving the BAP as a large-scale MOSO problem.

Because the optimiser solves MOSO problems of the form described in Section 1, the BAP is now defined as

$$\begin{gathered}\text{Minimise } E[W_P(\boldsymbol{x}, \xi)], \\ \text{Maximise } E[T_R(\boldsymbol{x}, \xi)] \\ \text{Subject to} \\ 0 \le x_i \le 6\end{gathered}$$

and x_i are positive integers. In this formulation of the problem, each niche is allowed to have up to 6 buffers. Although doing so ensures that the 84 candidate solutions from the original formulation are included in S, this also leads to the creation of new candidate solutions whose q is greater than 6. S is hence larger in this case and contains a total of $7^4 = 2401$ elements. (It should also be obvious that for this formulation of the problem, S grows significantly faster than in the original formulation, hence the necessity of a search algorithm.) Note that the MOO CEM can handle, in a practical manner, solution spaces that are much bigger than those considered here [9] [12].

The parameters for the metaheuristic were set as follows: The maximum number of evaluations was made 2000, epsilon was made 0.1, the population size was made 100 and the maximum number of outer loops was made 100. For the RS procedure, $\delta^*_{W_P}$ was made 0.12 while $\delta^*_{T_R}$ was made 0.2. N for the first phase was selected as 15, and the system was observed over a simulation period of 100 days.

3.2 Results and discussion

Table 3 shows the set of candidate solutions that was selected in the original problem while Table 4 presents the results obtained after using the MMY procedure on the set.

Solution	1	2	3	4	5
(x_1, x_2, x_3, x_4)	(1,1,1,3)	(1,1,2,2)	(1,2,1,2)	(1,2,2,1)	(2,1,1,2)
Solution	6	7	8	9	10
(x_1, x_2, x_3, x_4)	(2,1,2,1)	(2,2,1,1)	(3,1,1,1)	(1,3,1,1)	(1,1,3,1)

Table 3 - Solutions selected for the BAP in [1]

Solution	T_R	W_P	$S^2_{T_R}$	$S^2_{W_P}$	Rank	Runs	ID
(1,1,1,3)	16.44	1.49	0.06	0.04	0	36	1
(1,1,2,2)	16.69	1.62	0.07	0.05	0	37	2
(1,1,3,1)	16.76	1.71	0.06	0.09	0	37	10
(1,2,1,2)	17.00	1.85	0.08	0.07	0	41	3
(2,1,1,2)	17.08	2.18	0.07	0.05	2	77	5
(1,2,2,1)	17.17	1.96	0.08	0.09	0	41	4
(3,1,1,1)	17.18	2.82	0.08	0.05	3	15	8
(1,3,1,1)	17.22	2.14	0.07	0.15	0	143	9
(2,1,2,1)	17.31	2.32	0.07	0.07	0	41	6
(2,2,1,1)	17.50	2.55	0.08	0.10	0	41	7

Table 4 - MMY output for solutions in Table 3

According to the results in Table 4, all solutions in the set except Solutions 5 and 8 can be considered as correct selections (with respect to the δ^*_k values selected by the user/decision maker).

We now consider the results of the two-phase solution approach using MOOSolver. Figure 3 illustrates the results from phase one of the proposed solution approach. A total of 127 solutions was found by the metaheuristic.

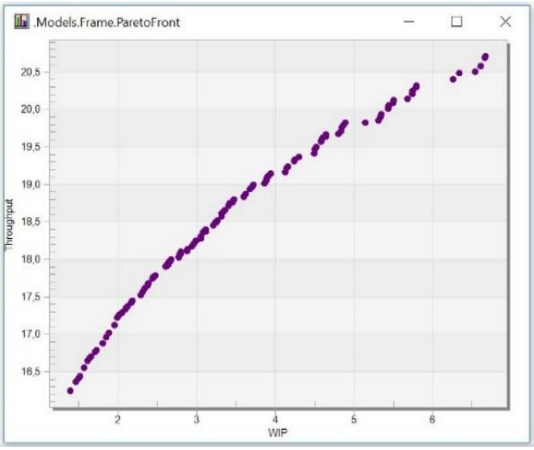


Figure 3 - Phase one results obtained by MOOSolver presented as a Pareto front in the DES product

In order to be able to compare the results obtained by MOOSolver with those in Table 4, it was assumed that the decision maker was only interested in solutions whose total buffer spaces q were equal or smaller than 6 for the second phase of the solution approach. These solutions are shown in Table 5 with their respective objective values from phase one. Observe that some of the solutions from Table 3 are members of the approximate Pareto set (see Table 5); For convenience purposes, therefore, we continue with the solution IDs numbering used in Table 3. Moreover, phase one results also inform us that Solutions 6, 7, 9 and 10 from Table 3 are not Pareto optimal as they do not appear in the approximate Pareto set. (This is because there now exist, in the approximate Pareto set, solutions that have better results than them.)

Solution	T_R	W_P	ID
(1,1,1,3)	16.42	1.50	1
(1,1,2,2)	16.64	1.61	2
(1,2,1,2)	16.96	1.85	3
(1,2,2,1)	17.12	1.95	4
(1,1,1,1)	16.25	1.39	11
(1,1,1,2)	16.37	1.46	12
(1,1,2,1)	16.55	1.57	13

Table 5 - Decision maker's preferred solutions for phase two

Given Table 5, M for phase two is thus equal to 7. Table 6 presents the results from phase two of the proposed solution approach.

Solution	T_R	W_P	$S^2_{T_R}$	$S^2_{W_P}$	Rank	Runs	ID
(1,1,1,1)	16.27	1.39	0.07	0.00	0	36	11
(1,1,1,2)	16.39	1.46	0.06	0.00	0	36	12
(1,1,1,3)	16.44	1.49	0.07	0.00	0	36	1
(1,1,2,1)	16.56	1.57	0.07	0.00	0	36	13
(1,1,2,2)	16.68	1.62	0.07	0.00	0	36	2
(1,2,1,2)	17.00	1.85	0.08	0.01	0	41	3
(1,2,2,1)	17.17	1.96	0.08	0.01	0	41	4

Table 6 - Phase two results obtained by MOOSolver for solutions in Table 5

These results give to the decision maker more accurate estimates for the selected set of solutions as well as a guarantee of at least 90% that the solutions whose Rank values are 0, constitute correct selection (with respect to δ_k^* values). In this case, all seven solutions can be selected.

4 CONCLUSION

This paper presented the use of relatively new solution algorithms (in the areas of metaheuristic and ranking and selection) in a novel optimiser developed to solve stochastic multi-objective optimisation or MOSO problems. The framework of the problems of interest was presented in the first section, followed by the proposed solution approach utilised by the optimiser to solve them. The optimiser was then used to solve the known buffer allocation problem, in the multi-objective context, to demonstrate the relevance of the solution approach proposed by the authors.

Large-scale MOSO problems can have an infinite number of potential solutions. Normally, an optimiser that can handle such problems would stop, in its solution approach, when a Pareto set has been approximated (by a search algorithm) using an arbitrarily, user-defined, number of observations during the SO process. In other words, despite the stochastic nature of the given problem, the optimiser would not include (in its solution approach) a rigorous statistical analysis. In effect, including rigorous statistical techniques such as RS algorithms in this case would make the SO process impractical given the fact that such algorithms can take long before statistical soundness can be guaranteed (see Section 3). The novelty of this paper, therefore, is in bringing RS in the large-scale MOSO context by, essentially, reducing the large-scale problem into a small-scale one first. And not merely arbitrarily; The solution approach gives, in effect, to the user the opportunity of selecting from the approximate Pareto set (which can be overwhelmingly large) only those solutions that they may have a biased interest towards (e.g. for business or practical reasons). As of now, the optimiser has been successfully implemented as an add-in to the commercial simulation software Tecnomatix Plant Simulation.

5 REFERENCES

[1] Yoon, M, 2018, New multi-objective ranking and selection procedures for discrete stochastic simulation problems, Ph.D. thesis, Stellenbosch University,.

[2] Coello Coello, C. A, 2009, Evolutionary multi-objective optimization: some current research trends and topics that remain to be explored, Frontiers of Computer Science in China, vol. 3, pp. 18-30.

[3] Deb, K, 2005, Multi-objective optimization using evolutionary algorithms, Wiley.

[4] Goldberg, D, 1989, Genetic Algorithms in search, optimization and machine learning, Addison-Wesley Publishing Company.

[5] Laguna, M and Marti, R, 2003, The optquest callable library. Optimization Software Class Libraries, pp. 193-218.

[6] Xu, J and Nelson, B. L and Hong, J, 2010, Industrial strength COMPASS: A comprehensive algorithm and software for optimization via simulation, ACM Transactions on Modeling and Computer Simulation (TOMACS), vol. 3, pp. 1-29.

[7] Amaran, S and Sahinidis, N. V and Sharda, B and Bury, S. J, 2014, Simulation optimization: a review of algorithms and applications, Annals of Operations Research, vol. 12, pp. 301-333.

[8] Laguna, M, 2011, Optquest: Optimization of complex systems, pp. 1-16.

[9] Bekker, J, 2012, Applying the cross-entropy method in multi-objective optimisation of dynamic stochastic systems, Ph.D. thesis, Stellenbosch University.

[10] Fu, M and Andradottir, S and Carson, J and Glover, F and Harrell, C and Ho, Y and Robinson, S, 2000, Integrating simulation and optimisation: research and practice, Proceedings of the 2000 Winter Simulation Conference, pp. 610-616.

[11] Fu, M, 2002, Optimization for simulation: Theory vs. practice, INFORMS Journal on Computing, vol. 14, pp. 192-215.

[12] Bamporiki, T and Bekker, J, 2018, Development of a discrete-event, stochastic multi-objective metaheuristic simulation optimisation suite for a commercial software package, South African Journal of Industrial Engineering, vol. 29, pp. 12-25.

[13] Bekker, J and Aldrich, C, 2011, “The cross-entropy method in multi-objective optimisation: An assessment, European Journal of Operational Research, vol. 211, pp. 112-121.

[14] Gershwin, S. B and Schor, J. E, 2000, Efficient algorithms for buffer space allocation, Annals of Operations Research, pp. 117-144.

[15] Alon, G and Kroese, D. P and Raviv, T and Rubinstein, R. Y, 2005, Application of the Cross-Entropy Method to the Buffer Allocation Problem in a Simulation-Based Environment, Annals of Operations Research, pp. 137–151.

6 BIOGRAPHY

Toussaint Bamporiki holds a B.Eng and an M.Eng degree in Industrial Engineering from the University of Stellenbosch.

James Bekker currently works at the Department of Industrial Engineering, Stellenbosch University. James does research in multi-objective optimisation of discrete-event simulation.

Moonyoung Yoon holds an M.Eng and a Ph.D. degree in Industrial Engineering from the University of Stellenbosch.

Extensible Callback Module Layering in Erlang

G.T. Hawkridge[1], K. Kruger[1], A.H. Basson[1]
[1]Mechatronic, Automation and Design Research Group, Department of Mechanical and Mechatronic Engineering, University of Stellenbosch, South Africa

Abstract
Intelligent manufacturing control is a key aspect of the Industry 4.0 paradigm. Erlang is a functional programming language designed for the development of soft real-time control systems. Erlang has many characteristics that are attractive when implementing manufacturing control. These characteristics include fault tolerance and ease of distribution. Erlang programs are built from multiple concurrent processes which interact through message passing. Code abstraction is provided by Erlang's behaviour mechanism, which separates code into a generic portion (the behaviour) and an application specific callback module. This paper presents an approach for the development of layered callback modules. Layered callback modules can be used to achieve process specialisation. To effectively layer callback modules, it is important that each layer be extensible. Extensible layers are callback modules that do not limit the functionality of subsequent layers. Some best practices relating to message handling and state access are presented. These best practices facilitate the development of callback modules that can be layered in a manner that assists modularity and code reuse. It is expected that this approach for process specialisation will be beneficial in Erlang based IIoT applications since many of the “things” in these systems will share common base functionality.

Keywords
Erlang, Code Abstraction, Software for Manufacturing

1 INTRODUCTION

Current progression, by both industry and academia, towards the realisation of Industry 4.0 has highlighted the need for intelligent manufacturing control [1] [2] [3] [4]. Key aspects of intelligent manufacturing control include decentralisation, cyber-physical systems (CPS) and the Industrial Internet of Things (IIoT) [3].

Erlang is a functional programming language designed for the development of fault-tolerant soft real-time control systems [5]. OTP (Open Telecom Platform) is a set of libraries included in Erlang that simplify the development of large complex systems [6, pg. 73]. Erlang has many attractive features, such as concurrency, scalability and fault tolerance, that are useful when implementing intelligent manufacturing control systems [7]. Erlang has been used to implement standby redundancy on embedded systems for a manufacturing station [8].

Code abstraction is a necessary feature when developing the control software for large systems. Erlang provides for code abstraction with the behaviour mechanism which separates code into a generic portion (the behaviour) and an application specific callback module.

This paper proposes an approach to achieve process specialisation in Erlang through callback module layering and recommends some best practices for improving the extensibility of callback layers. The paper begins with some background on Erlang followed by an overview of the behaviour mechanism. Callback layering is then described and some best practices for message handling and state access are discussed.

2 ERLANG BACKGROUND

The Erlang programming language is designed for applications that require concurrency, fault tolerance and distribution [9]. Erlang programs are built up as independent concurrent processes which cooperate with one another to achieve the systems goals. Erlang processes have the following properties: processes have sole access to their internal state, processes influence one another through asynchronous message passing and processes have the capability to spawn further processes [6, pg. 70].

This section introduces foundational concepts used in Erlang for the sake of readers not familiar with the language.

2.1 Syntax Overview

An overview of Erlang’s syntactic conventions is shown in **Table 1**.

2.2 Functional Purity

Erlang is a functional programming language; however, it is not necessarily pure. In a pure functional language, the output of a function is entirely dependent on the input arguments (i.e. it does not access global state). Furthermore, the effects of function execution are contained within the function. Therefore, pure functions do not permit side effects such as global state modification or I/O

Syntactic Element	Description	Syntactic Convention	Example
Variable	Construct for storing Erlang terms	Begins with an uppercase character	Variable, ThisVariable, That_Variable
Comment	Used to explain or annotate code	Preceded by a percentage symbol (%)	% Comment line here
Atom	A string literal constant	Begins with a lowercase character or enclosed with single quotes	atom, ok, 'An Atom'
Tuple	A fixed size set of terms	A set of terms enclosed with braces	{Term1, …, TermN}, {nested,{tuple}}
List	A dynamic length set of terms	A set of terms enclosed with square brackets	[Term1, …, TermN], [nested,[list]]
Function	A collection of related Erlang expressions	A function name (an atom) followed by arguments enclosed with round brackets	function(Arg1,…,ArgN) -> expression1, …, return_value.

Table 1 - Overview of Erlangs Syntactic Conventions

operations. An advantage of pure functions is that their code is easy to reason through and their correct functioning can be validated.

Since many practical applications require side effects, [10] recommends separating Erlang code into sections that are pure (calculations and data manipulation) and impure (message passing, file access, etc.). In this way the pure code can be validated, and the impure code can be thoroughly tested to ensure it functions as intended.

2.3 Dynamic Typing

Erlang is a dynamically typed language. Dynamic typing means that types are checked at run time instead of at compile time. Erlang typing is strict which means that type mismatches (i.e. a character used in a mathematical expression) will generate exceptions. **Figure 1** shows Erlang's dynamic typing used in a function that prints the value of the argument with which the function is called. This example shows that dynamic typing lets a single function definition serve multiple variable types.

Function Definition:

```
print_function(Var) ->
    io:format("The value is ~p",[Var]).
```

Example Usage:

```
> print_function(true).
The value is true

> print_function(3.14).
The value is 3.14
```

Figure 1 - Dynamic Typing Example

2.4 Pattern Matching

A core feature of Erlang is pattern matching. In pattern matching the input expression is compared to a pattern, and if the two are equivalent then the match succeeds. Pattern matching can be used to extract certain entries from a data structure as shown in **Figure 2**(a). In a pattern, underscores (`_`) and variable prefixed by underscores (`_Var`) are "match-any" wildcards that will match any entry. Pattern matching is frequently used in function heads as shown in the recursive factorial function in **Figure 2** (b). The input arguments are applied to the first function head that the pattern successfully matches. For the example, this means that while `N` does not match `0` the second function head matches and is executed. When `N` is `0`, the first function head matches and returns `1`.

a) Expression Pattern Matching:

```
{a,A,_,B}={a,b,4,1.2}.
% A=b, B=1.2
```

b) Function Pattern Matching:

```
factorial(0) ->
    1;
factorial(N) ->
    N*factorial(N-1).
```

Figure 2 - Pattern Matching Examples

2.5 Data Structures: Records and Maps

Erlang provides many different data structures; however, records and maps are usually the preferred options since their contents can be pattern matched in function heads. A record is a fixed length data structure which stores entries in named fields. A map is a dynamically sized key-value store. A significant difference between records and maps is that record field names are checked at compile time whereas incorrect map keys are only evident through incorrect behaviour or runtime errors.

3 BEHAVIOURS

3.1 Background

A behaviour is an abstraction that implements the generic portions of a common model or pattern.

OTP provides some standard behaviours including gen_server (that implements a client-server relationship), gen_event (that implements a publish-subscribe mechanism), gen_statem (that implements an event-driven state machine) and the supervisor model (that supervises child processes according to a selected restart policy). The application specific logic for these behaviours is implemented in the form of callback modules. An advantage of using these standard behaviours is that they have been thoroughly tested in many software products.

3.2 Behaviour and Callback Structure

The coupling between behaviour and callback modules is loose. Typically, the name of the callback module is passed to the start function of the behaviour module (which starts the process). The callback module name is stored in the process' state and is used to make calls to functions in that module using reflection.

This paper uses the agenda manager of a singulation unit as an example. The agenda manager is responsible for managing the singulation unit's bookings. Since booking management is a common element in manufacturing systems, that portion of the functionality can be implemented in a behaviour module while the singulation unit specific aspects can be placed in a callback module (i.e. how estimates are calculated when preparing proposals).

Figure 3 shows the code structure for a simple implementation of this example using the behaviour-callback pattern. The process is created when the start function in the callback module is called (Line 2). This function calls the start function of the behaviour module (Line 8) with the callback module's name (singulation_am) as an argument.

The behaviour module's start function spawns the new process (Line 9). The process starts execution with the setup function (Line 10) and then enters the loop function with the callback module and initial state as arguments (Line 12). The loop function (Line 13) enters a receive statement (Line 14) where it waits until a message is received. Received messages are matched against the specified patterns (Lines 15, 18 & 23). For certain messages, the behaviour module will handle the message internally. In the example, this happens when a booking request is received since it does not require application specific information (Line 16). If message handling requires application specific information, then this is obtained from one of the callback functions in the callback module. In the example this occurs when a proposal request is received (Line 19) and the behaviour calls the `get_estimate` function (Line 4) to obtain a performance estimate.

Once a message has been processed, the loop function recursively calls itself with the updated state in the arguments (Lines 17, 21 & 24) (Erlang features optimised tail recursion, so this does not lead to stack overflow). To ensure that the process' message queue does not get overloaded, the last entry in the receive statement is usually a "match-any" pattern (Line 23) so that messages that do not match any prior patterns are removed and ignored.

```
1)  -module(singulation_am). %Callback
2)  start(Args) ->
3)    generic_am:start(
        singulation_am,
        Args).
4)  get_estimate(RequestInfo) ->
5)    Estimate=%Calculate Estimate,
6)    Estimate.
```

```
7)  -module(generic_am). %Behaviour
8)  start(CBMod,Args) ->
9)    spawn( %Spawns a new process
        generic_am,
        setup,
        [{CBMod,Args}]).
10) setup({CBMod,Args}) ->
11)   InitialState=%Form Initial State,
12)   loop(CBMod,InitialState).
13) loop(CBMod, State) ->
14)   receive
15)     {request_proposal,Info} ->
16)       Estimate=
            CBMod:get_estimate(Info),
17)       %Create Proposal and Reply
18)       loop(CBMod,NewState);
19)     {request_booking,Details} ->
20)       NewState=%Add booking,
21)       loop(CBMod,NewState);
22)     _ ->
23)       loop(CBMod,State)
24)   end.
```

Figure 3 - Example Behaviour and Callback Modules for a Simple Agenda Manager

3.3 Behaviour and Callback Module Attributes

The behaviour module usually trusts that the callback module implements all the required callback functions: however, this can be checked using `erlang:function_exported` before making the callback function call. This capability can also be used to implement optional callback functions.

To help ensure that callback modules implement the required callback functions, Erlang provides the behaviour and callback module attributes. The behaviour module attribute is used to indicate the behaviour modules for which a callback module implements callback functions. The callback module attribute is used to specify which callback functions a behaviour module expects to be implemented. These attributes are used to generate compiler warnings for modules that do not implement the specified callback functions; however, these attributes are not required for a behaviour-callback pattern to be implemented.

3.4 Callback Layering

The behaviour-callback pattern is not limited to the case where there is a single behaviour module and a single callback module. A module can be implemented as both the callback module for a behaviour and act as the behaviour module for a different callback module. Several of such modules can then be layered to create a stack of layered callback modules with a single foundational behaviour at the bottom. Each layered callback module should be used to enhance the functionality provided by the layer below it. In this way, the layers form a stack of increasing specialisation.

Figure 4 illustrates how callback module layering can be applied to the agenda manager example. Since the agenda manager can be considered a server which provides bookings for clients, the `gen_server` behaviour is used as the foundation of the stack. It is recommended that the foundational behaviour in the stack be a standard OTP behaviour since these behaviours are compatible with OTP's built in debugging tools. The generic agenda manager functionality is built on top of the `gen_server` behaviour as a callback module. The singulation unit specific functionality is then added on top of the generic agenda manager layer.

To start a process which executes a layered callback stack, the start function of the top layer is called. Each callback module's start function calls the start function of the layer below it with the calling module's name as an argument. In this way each layer specifies the layer below it. Execution begins with the behaviour's initialisation function which calls the initialisation function of the layer above it. These calls are propagated up the stack with the return value of each layer being integrated into the return value of the layer below. Execution then enters the behaviour's primary loop. From the primary loop, callback functions are called in response to messages. These calls propagate up the stack until they reach the layer that provides the required level of specialisation. When the process is stopped, the terminate function is called in a manner similar to the initialisation function.

Callback layering contributes to code modularity since code for a specific functionality is contained within a single layer. Furthermore, a layer module can be replaced with a different module that exposes and requires the same interfaces without influencing the remainder of the stack. Callback layering also facilitates code reuse since processes with different roles can start with different top layer modules and converge to the same layers lower down in their behaviour stacks. (i.e. the agenda manager for a robotic arm could build on top of the generic agenda manager layer).

3.5 Behaviour and Callback Extensibility

Often when a behaviour is developed, the developer has a clear understanding of the functionality that will be implemented in the callback functions. It is therefore tempting, and often easier, to cater directly to that expected functionality. However, when that behaviour is later reused, the behaviour may need to be altered to facilitate additional functionality. This can be particularly problematic if the behaviour is not editable by the developer (i.e. part of a closed source 3rd party library).

Therefore, to maximise the modularity and code re-use benefits of using layered callback modules, each layer module should be extensible. The extensibility of a layer module here refers to the degree to which that module facilitates the implementation of additional functionality within the subsequent callback layers. In other words, a highly extensible callback layer does not make limiting assumptions about the implementation of the layer above it.

Two categories of assumptions that limit or prevent the implementations of layered callback modules have been identified. These two categories are: assumptions about the message handling requirements of subsequent layers and assumptions about the state access requirements of subsequent layers. The following sections will discuss these assumptions in greater detail.

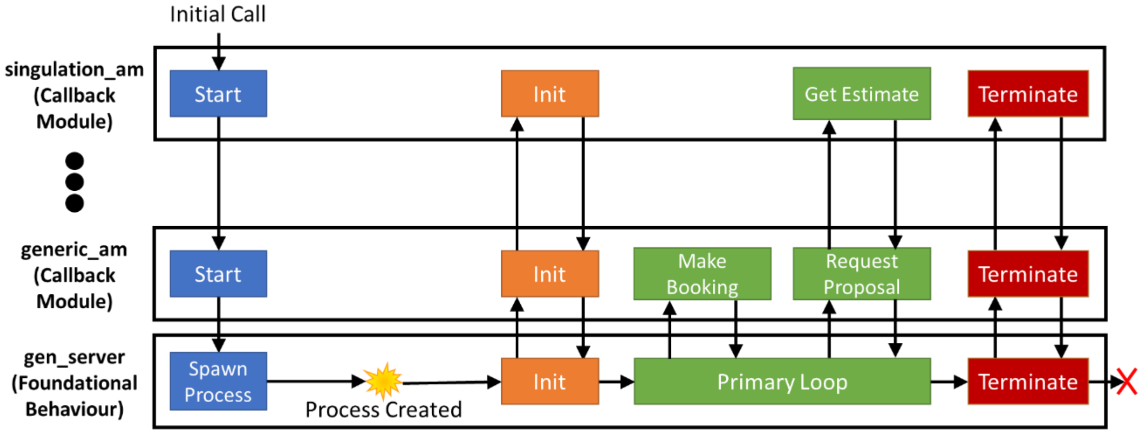


Figure 4 - Callback Module Layering

4 MESSAGE HANDLING

Interprocess communication is a foundational aspect of Erlang programs. It is therefore understandable that a behaviour that limits the message handling capabilities available to its callback module will limit the functionality that can be implemented in that callback module. For a layered callback module to be extensible, it must therefore not restrict the ability of the layer above it to send and receive messages. In the case of the agenda manager example, the `singulation_am` layer may need to request information from another process before it can generate a performance estimate.

Sending messages from different layers does not cause complications, as long as layers have access to all the data needed to formulate and send the required messages. However, complications arise if layers implement their own message receiving.

The presence of multiple receive statements in a layered callback stack can result in hanging or poor responsiveness (i.e. when blocking receive statements are used to achieve synchronous calls). Furthermore, there needs to be a "match-any" pattern in at least one of the receive statements so that messages that do not match any of the receive statements do not overload the message queue. However, this has the potential to erroneously remove messages that are intended for other layers. (i.e. a receive statement in `singulation_am` removing a booking request for `generic_am`).

Therefore, the recommended approach is to have a single receive statement in the foundational behaviour module. If a received message does not match any of the patterns for that module then the message should be passed on to the next layer in the stack using a predefined callback function (i.e. `handle_message`). The next layer then attempts to match it against its own patterns and passes it on upwards using a similar callback function if no match is found. If a message reaches the top of the stack without matching, it can then be safely ignored. This leads to a hierarchy since lower layers have message handling precedence over higher layers. Care should be taken to avoid (unintended) message pattern overlaps between layers.

If having additional receive statements within layers is unavoidable, then these receive statements should be non-blocking and only match messages intended for that layer.

A further advantage of having every layer in the stack use only the foundational behaviour's receive statement is that more of the callback functions in higher layer modules can be functionally pure and therefore easier to validate.

5 STATE ACCESS

State is a representation of a process' current condition. The ability to maintain state is a necessary requirement for any non-trivial process since it allows past events to influence the handling of future events. Erlang processes typically rely on local state. This local state is implemented as an argument to the foundational behaviour's primary loop. If a function requires access to this state, then it is passed as an argument to that function. Erlang provides facilities for a global process state (i.e. the process dictionary); however, the use of global state is generally discouraged as it leads to code that is functionally impure and therefore difficult to validate.

If a callback module does not have the facility to maintain its own local state, then it will either limit the functionality that can be implemented in that module or force the callback module to use global state. An extensible layer module should therefore provide local state storage facilities for the layer above it (regardless of whether that layer currently requires state storage).

In the case of the agenda manager example, the `gen_server` module provides the `generic_am` module with state storage facilities which are used to store the booking list, etc. However, if the `generic_am` module does not provide the `singulation_am` module with state storage facilities then this would limit that layer. For example, it would not be possible to implement functionality whereby the `singulation_am` receives status updates from the singulation unit's execution manager and stores the current status for consideration when generating performance estimates.

Erlang's dynamic typing is helpful in this regard. A layer can allocate a term in its own local state for the local state of the layer above without knowing the format or contents of that layer's state. This callback-state is then included as an argument whenever the layer calls a callback function in the layer above. Since function arguments are passed by value, callback functions should return an updated state, which the lower layer uses to replace the previous callback-state.

Another way that the achievable functionality of a layer can be limited is when the layer below it only includes certain state elements in callback function arguments. This may be because only a specific subset of the layer's state was required for the initial callback implementation.

In terms of extensibility, it may be desirable for a layer to have as much access to the state of the layers below it as possible. However, this will be detrimental to the modularity of the callback layer, since it could require knowledge of the state structure of lower layers. If a layer incorporates references to the layer below it, these interfaces would need to be validated if changes were made to that lower layer. Erlang maps are helpful in this regard since they do not require the higher layer to import a definition header file, but this advantage of maps comes with the risk of run-time errors, as pointed out Section 2.5.

Access to another layer's state here implies read access, since it is recommended that layers have sole write access to their own state. This prevents modifications in other layers from creating state inconsistencies. If modification of a state entry by higher layers is desired, then this can be implemented by allowing modify requests to be included in the return values of callback functions.

6 CONCLUSIONS

This paper proposes an approach for achieving process specialisation in Erlang through callback module layering. The use of extensible layers can improve the ease of development, modularity, code re-use and ease of modification of layered callback modules.

Extensible message handling is facilitated by limiting layers to using the foundational behaviour's receive statement. Extensible state access is facilitated by providing each layer with a mechanism for storing local state and by providing each layer with access to the relevant state of the layer below it. These best practices facilitate the development of extensible callback module layers.

Erlang has a strong heritage in telecom and internet applications. It is expected that this approach for process specialisation will be beneficial in Erlang based IIoT applications since many of the “things” in these systems will share common base functionality.

7 ACKNOWLEDGMENTS

The financial assistance of the National Research Foundation (NRF) towards this research (through funding a bursary, with reference number 103260, for the first author) is hereby acknowledged. Opinions expressed, and conclusions arrived at are those of the authors and are not necessarily to be attributed to the NRF.

8 REFERENCES

[1] Brettel, M., Friederichsen, N., Keller, M. and Rosenberg, M., 2014, How virtualization, decentralization and network building change the manufacturing landscape: An industry 4.0 perspective, International Journal of Mechanical, Industrial Science and Engineering, 8/1:37-44.

[2] Zhou, K., Liu, T. and Zhou, L., 2015, Industry 4.0: Towards future industrial opportunities and challenges, International Conference on Fuzzy Systems and Knowledge Discovery, IEEE, 12:2147-2152.

[3] Lasi, H., Fettke, P., Kemper, H.G., Feld, T. and Hoffmann, M., 2014, Industry 4.0, Business & Information Systems Engineering, 6/4:239-242.

[4] Zhong, R.Y., Xu, X., Klotz, E. and Newman, S.T., 2017, Intelligent manufacturing in the context of industry 4.0: a review, Engineering, 3/5:616-630.

[5] Armstrong, J., 1996, Erlang—a Survey of the Language and its Industrial Applications, In Proc. INAP 96.

[6] Armstrong, J., 2010, erlang, Communications of the ACM, 53/9:68-75.

[7] Kruger, K. and Basson, A., 2017, Erlang-based control implementation for a holonic manufacturing cell, International Journal of Computer Integrated Manufacturing, 30/6:641-652.

[8] Hawkridge, G., Basson, A.H. and Kruger, K., 2018, An Evaluation of Erlang for Implementing Standby Redundancy in a Manufacturing Station Controller, 8th Workshop on Service Orientation in Holonic and Multi-Agent Manufacturing.

[9] “What is Erlang” [Online] Available http://erlang.org/faq/introduction.html [Accessed August 29, 2017].

[10] Armstrong, J., 2003, Making reliable distributed systems in the presence of software errors (Doctoral dissertation, Mikroelektronik och informationsteknik).

9 BIOGRAPHY

Greg Hawkridge is a PhD candidate in Mechatronic Engineering and a member of the Mechatronics, Automation and Design Research Group at University of Stellenbosch, South Africa.

Karel Kruger obtained his PhD from Stellenbosch University, South Africa. He is a lecturer at the Department of Mechanical and Mechatronic Engineering and is part of the Mechatronics, Automation and Design research group.

Anton Basson obtained his PhD in Aerospace Engineering at Penn State University in 1992. He is a Professor in Mechanical Engineering and co-leader of the Mechatronics, Automation and Design Research Group at University of Stellenbosch, South Africa.

Method for the Identification of Potentials in Lean Production Systems as Basis for the Selection of Digitalization Technologies

G. Hoellthaler[1], S. Weissenborn[2], S. Braunreuther[1,3], G. Reinhart[1,4]

[1]Fraunhofer Research Institution for Casting, Composite and Processing Technology IGCV, Augsburg, Germany

[2]valantic Supply Chain Excellence AG, Munich, Germany

[3]Augsburg University of Applied Sciences, Augsburg, Germany

[4]Institute for Machine Tools and Industrial Management (*iwb*), Technical University of Munich, Munich, Germany

Abstract

Since the dissemination of the philosophy of lean production, there has been an unprecedented acceptance of this organizational set of rules in Western production companies. Ever since then, lean production has been used for systematization of production organization and for waste reduction in production processes. However, it has not evolved, even though markets are more demanding due to individualization and volatility. Digitalization provides tools to adapt lean production to aggravated market conditions by expanding its capabilities. Therefore, digitalization has the potential to enhance the realization of the philosophy of lean production.

In this context, the current status quo of company-specific lean production systems needs to be assessed and structural deficits have to be identified. Extensive knowledge about a company-specific lean production system and its deficits provides the basis for the selection of appropriate digitalization technologies. This paper introduces a method that allows the evaluation of lean production systems in the context of digitalization.

Keywords

Lean production system (LPS), maturity model, requirement capability comparison (RCC)

1 INTRODUCTION

Enterprises are constantly forced to improve their competitiveness due to the ongoing globalization and satisfaction of increasing customer demands. Lean production systems (LPSs) support to realize efficient production processes and thus help enterprises to be economically successful [1]. Since the organizational set of rules of lean production (LP) encourages mastering the complexity of individual production processes and organizing labor, it is widely used by production companies [2]. However, in recent years the change of production challenges and constraints has brought LPSs to the limits of their capabilities [3].

2 STATE OF THE ART

2.1 Digitalization: enabler for lean production

The elements of LP have their initial origin in the Japanese automotive industry [4]. With a publication at the beginning of the 1990s, Womack, Jones and Roos created the basis for an unprecedented proliferation of LP into Western production companies [2, 5]. Ever since then, this organizational set of rules consisting of principles, methods and tools has become more and more important to production companies [2, 6] due to its efficiency and value-added process design [7]. The elements of LP contribute to the avoidance of waste by reducing efforts regarding transport, inventory, motion, waiting, overprocessing, overproduction and defects [6]. While framework conditions and challenges for production processes have been intensifying, the implementation of the elements of LP has remained mostly the same and not evolved any further. In this context, digitalization technologies offer opportunities for other ways of implementing elements of LP. Here, information and communication technologies contribute to realize functionalities like data capturing, data transfer, data retention, data processing and data provisioning [8, 9]. Therefore, such technologies support individual production processes with necessary information and facilitate an intelligent interconnection of resources, employees and objects in a factory [10]. Various implementations of digitalization technologies like a digital Kanban system or a pick-by-light system represent examples of how to implement elements of LP in a digital way. These as well as other examples state that the digital implementation of elements of LP represents the subsequent and logical next step, since digitalization technologies contribute to a more efficient use of elements of LP [3, 11]. Therefore, a profound knowledge about the current state of an LPS is necessary in order to be able to select appropriate digitalization technologies. An assessment of a company-specific LPS needs to depict where action and improvement are required.

2.2 Assessment of lean production systems

There are different ways to assess LPSs. Karlsson and Åhlström (1996) focus on “assessing changes towards lean production”. Here, the focus is on describing probable effects on a production system by applying principles of LP [12]. Dombrowski et al. (2016) are more specific by developing an interdependency framework that depicts effects among LP methods using System Dynamics. Therefore, this framework depicts changes of lean methods within an LPS [13]. Moreover, “value stream mapping 4.0” is a holistic approach that assesses lean principles with respect to deriving opportunities for digitalization. This method identifies waste in information logistics in order to optimize information logistic processes [9]. Another approach is maturity-based and focusses on the assessment of established lean principles in a production system in the context of digitalization. Based on this, migration paths are derived for the identification of individual use cases for digitalizing production processes [10].

Altogether, these approaches depict generic as well as maturity-based assessment methods that are on different levels of granularity. It is appropriate to develop a maturity-based assessment method as an assessment of an LPS on a more detailed level and towards a desired state of an LPS. Therefore, a maturity based method contributes to a comparison of a desired state (requirement) and current state (capability) of an LPS. The disparity between requirement and capability of an LPS reveals the need for action and potential for improvement. Hence, the implementation of suitable digitalization technologies can contribute to the reduction of the identified disparity.

2.3 Maturity models

2.3.1 Definition

Maturity models are specific step models that assess an current development state of a focussed object by using defined maturity degrees [14, 15]. The maturity degrees result from the requirements of a focussed object that need to be fulfilled and that are consecutively constructive [16]. Therefore, these maturity degrees need to be defined in regard to prerequisites for the development stages of an object in order to describe necessary requirements for reaching each subsequent degree [14, 17].

2.3.2 Characteristics

Different maturity models are conceptualized in regard to each specific contemplation focus. Depending on their purpose, these models can assess whole companies, certain departments or specific processes such as production processes. Maturity models have in common that their maturity degrees are used for the assessment of current states regarding focussed objects. Therefore, a structured analysis reveals probable deficits regarding a desired condition. Requirements to be met are defined for each maturity degree and the improvement measures for the realization of a higher maturity degree of an object can be derived precisely. [14, 16, 18]

2.4 Concepts for the requirement capability comparison

As a base for the development of a requirement capability comparison (RCC) method in the context of LP, the following subchapters briefly describe existing concepts.

2.4.1 Quality Function Deployment: QFD

QFD is a structured procedure model used in the product development process. Its superior objective is to guarantee that customer requirements are converted into product features. Customer requirements and their realization need to be separated precisely in a matrix for their comparison and alignment, which is also known as House of Quality (HoQ). With a HoQ for every phase of the product development process, the realization of customer requirements can be monitored and deviations can be identified in order to prevent erroneous developments soon enough. [19]

2.4.2 Business process modelling: EDEN

EDEN describes a maturity model for a company-wide process management that assesses different departments on organizational as well as on processual levels. An identified status quo reveals deficits and provides the basis for deriving a desired condition and therefore for deriving a strategy for the realization of the transformation from the status quo to a desired condition. This model consists of 170 single criteria from the following nine dimensions goals, strategy, methods, organization, measurement, competences, communication, documentation and information technology that are assessed on a scale consisting of six maturity degrees. [20]

2.4.3 Lean Logistics Maturity Model: LLMM

LLMM focusses on the evaluation of production processes regarding intralogistics. In regard of organizational as well as subject-specific aspects, this maturity model assesses four hierarchically structured objectives: configuration fields, process areas, configuration criteria and peculiarity degrees. When the maturity degree of all of the subordinate objectives has increased, the superordinate objective rises to the next level of its specific maturity degree scale. [21]

2.4.4 Lean Capability Model: LCM

LCM supports the introduction and sustainable development of LP on organizational as well as on technological levels within production companies. By evaluating the current state of the implementation of LP, development aspects are depicted that have to be improved to be able to benefit from its advantages in the long term. This maturity model therefore assesses the design on lean structures for example on an organizational level and the performance of single LP methods on a five-step maturity degree scale. [22]

2.4.5 Lean Enterprise Transformation Maturity Model: LETMM

LETMM is a lean-specific maturity model, which focusses on the assessment of business processes. This model has differences in three sections: lean transformation, processes of the product life cycle and supportive infrastructure. Each of these sections consists of subordinate classes that need to be assessed according to five generic maturity degrees. Despite these maturity degrees, this model allows the depiction of the status quo in company-specific contexts, since responsible employees from different hierarchical levels of different departments give their professional opinion on their processes. Based on this fact, LETMM allows the identification of process-specific potentials for improvement. [23]

3 FUNDAMENTALS FOR THE DEVELOPMENT OF A REQUIREMENT CAPABILITY COMPARISON METHOD

3.1 Pivotal characteristics of the method

In order to provide a method, which is applicable on individual LPSs, the development focusses on different characteristics that the method needs to fulfill (figure 1). These characteristics should contribute to an adaptive method that can be conducted and evaluated by representatives from the production environment. Furthermore, applicability for different companies with individual peculiarities of their LPS should be facilitated as well.

3.2 Analysis of existing concepts for the requirement capability comparison

The existing concepts as described in chapter 2.4 are predominant in different application areas. Regarding the requirements for a RCC method in the context of LP, their applicability needs to be analyzed. All concepts except QFD are maturity-based. Moreover, LLMM, LCM and LETMM already have a lean connection whereas QFD and EDEN have not been used in the context of LP so far. This analysis regarding relevant characteristics, which is depicted in figure 1, differentiates between three degrees of fulfillment.

Characteristics \ Concepts	QFD	EDEN	LLMM	LCM	LETMM
Intuitive usability [21]	●	◑	○	◑	◑
Generality [24]	◑	●	●	●	◑
Scalability [25]	◑	●	◑	○	◑
Evaluability of different states [26]	○	●	●	○	●
Feasibility of detail levels [25]	◑	◑	●	◑	●
Graphical evaluation [26]	○	●	●	○	○
Focus on lean production	○	○	◑	◑	◑
Precise rating scale	◑	◑	●	●	●

○ Not fulfilled ◑ Partly fulfilled ● Entirely fulfilled

Figure 1 – Analysis of existing concepts

In summary, there is no method at all that fulfills the required characteristics to their full extent. Therefore, none of the listed methods are directly suitable for the RCC in the context of LP in its current peculiarity. This is an inducement for the development of a RCC method that focusses on LP. Elements of the individual approaches of the concepts described will therefore be used for the development of the method.

3.3 Requirements for production processes

Varying production challenges bring LPSs to the limits of their capabilities. Nevertheless, lean structures are a solid base for the implementation of digitalization technologies. Due to these aspects, a RCC for assessing the level of implementation of an LPS is needed. This alignment needs to rate the level of fulfillment of production requirements by the capability of the methods that are defined as part of LP.

Main goals for production processes focus on the minimization of cost and time and maximization of quality [6]. In order to provide a more specific clustering of requirements for production processes, these goals are augmented in regard to flexibility, logistics, environment and human needs. These goals are the superior clusters for further specification that was conducted in this context and resulted in 40 requirements that are not mentioned in this paper in detail. These goals can however be adapted according to the LPS to be assessed.

4 REQUIREMENT CAPABILITY COMPARISON METHOD

4.1 Derivation of maturity degrees

Based on the previously described concepts, a new RCC method is derived in the context of LP. This method requires a rating scale to evaluate the current as well as the target states regarding the fulfillment of requirements for production processes in regard to the capability of LP methods. Six maturity degrees (0 to 5) are therefore specified (figure 2). As mentioned in chapter 2.3.1, maturity degrees are consecutive and characterized by requirements regarding focussed objects that have to be fulfilled to reach the

	Definitions: The method…
0	…is not implemented or has no influence on the requirement.
1	…was implemented as part of a pilot project to support fulfilling a requirement.
2	…is implemented in several production areas to support fulfilling a requirement.
3	…is entirely implemented and contributes slightly to fulfill the requirement.
4	…is entirely implemented and contributes significantly to fulfill the requirement.
5	…is entirely implemented and contributes to fulfill the requirement completely.

fulfillment
implementation

Figure 2 – Maturity degrees

next degree. Nonetheless, maturity degrees can be skipped in case of big and rapid steps regarding the fulfillment of requirements for production processes.

By definition, all stages of the implementation of lean elements have to be completed. Due to this fact, all theoretically possible specifications regarding the fulfillment of requirements are considered in the definition of maturity degrees. The evaluated characteristics become more numerous as maturity degrees increase. In case of the RCC, a higher maturity degree represents a greater contribution to the fulfillment of requirements as well as the implementation of a larger scale of LP methods than lower degrees.

4.2 Elements and functionality of the requirement capability comparison method

The RCC method serves to assess the contribution of LP methods in regard to a systematic fulfillment of requirements for production processes. A matrix-based model is developed for the conduction of this method and is used for documentation and analysis of the RCC. The requirements for production processes are listed line-by-line on the left of the matrix and are grouped by their main categories. The LP methods are listed column-by-column at the head of the matrix and are grouped by their principles. In the matrix, the evaluation concerning the fulfillment of requirements is performed. Two key figures are defined for analyzing the result of the RCC. The degree of fulfillment of requirements is calculated by dividing the maturity degree of the current state by the target state for each requirement for production processes. The degree of establishment for each LP method is calculated similarly to the degree of the fulfillment of requirements. This key figure is calculated column-by-column and compares the current state with the target state of maturity degree in regard to the establishment of an LP method. Due to the categorization of requirements for production processes and LP methods, the analysis is possible on various levels of detail. Thus, the model supports the gathering of all relevant information for the RCC. Furthermore, it allows a mathematical analysis as well as a graphical analysis of the result by using a cobweb chart. An Excel-supported model was developed to perform the method of the RCC in practice.

The procedure starts with defining relevant requirements for production processes. Therefore, a selection of several requirements that are adaptable or expandable is provided. LP methods that have already been implemented in a company's production system have to be determined subsequently. As an option, a prioritization of requirements can be set for production processes. The assessment of the current degree and the target degree of the fulfillment of requirements is the main objective of the RCC method. The fulfillment of requirements by an LP method is assessed one-by-one according to each of the six maturity degrees,

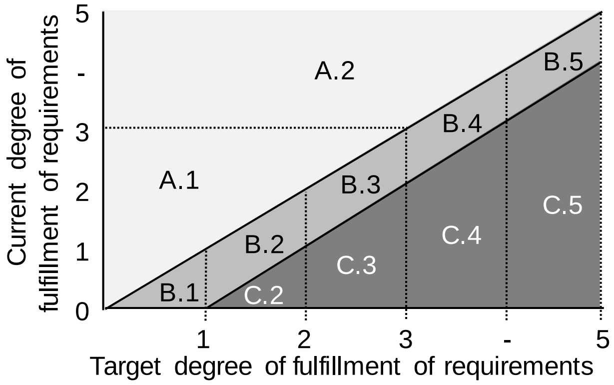


Figure 3 – Grid for clustering the results of the requirement capability comparison

whereas the current state is assessed for each requirement in comparision to the desired target state. By defining the target state, the future direction of production is determined with the lean philosophy in mind. Not until the evaluation of an LP method is completed, can the next one be considered. After completing the evaluation, the automatic calculation of the result by the model provides an immediate analysis. Hence, discrepancies between the current state and the target state can be identified. These discrepancies are potential for optimization, which can be realized by appropriate measures or by the implementation of digitalization technologies. To check the effectiveness of defined measures, the RCC method has to be conducted iteratively.

4.3 Generic guidance items

To realize the identified potentials for optimization and thus achieve the target state of fulfillment of requirements, the following guidance items have been defined:

- Implementation of further LP method(s)
- Expansion of the implementation of the LP methods
- Staying at current maturity degree
- Decreasing the current maturity degree to avoid an unnecessary use of resources
- Support of the LP method with a digitalization technology

As not all generic guidance items fit and promise optimization for all identified potentials, the RCC results are clustered for the assignment of appropriate guidance items (figure 3). The threshold value for clustering the results is defined by the discrepancy of the assessed value for the current state and the target state. The angle bisector therefore represents the optimal state, where the current state and target state are identical. For all of the results above the angle bisector, the current state is higher than the target state. For all of the results below the angle bisector, the current state is inferior than the target state. The clustering is divided into two steps. The rough clustering first determines whether there are measures absolutely necessary (C.x), not absolutely necessary (B.x) or depending on the future target state (A.x). On the basis of the detailed clustering (A.1, A.2, B.1, etc.), which depends on the assessed current state, appropriate guidance items are determined for all individual results.

5 USE CASE

5.1 Subsumption of the use case

For the validation of the RCC method, it was conducted in regard to the LPS at two production sites of a company from the german medium-sized businesses that develop, produce and distribute filter elements for filter presses. A company-specific LPS had been implemented several years ago as well since there have been other projects focussing on LP in the past.

5.2 Conduction of the requirement capability comparison method and analysis of the result

The RCC method was conducted in a three hour workshop with the lean experts for the german factories. After a short description of the method, its procedure and maturity degrees, the fulfillment of requirements was assessed. The automatic calculation of the result by the model supporting the method allowed a first analysis at the end of the workshop.

22 of the 35 defined LP methods according to [6] are implemented. The analysis in regard to the clustering of the results shows that measures for the optimization of the current state of three methods are absolutely necessary. The three methods revealing potentials for optimization are process standardization, quick changeover and value stream planning. Due to the clustering, the process standardization and the value stream planning can be optimized company-wide by extending the implementation of the methods. Supporting the LP method with a digitalization technology is a possible approach for improving the method quick changeover. By observing these guidance items the target state for the improvement of this LPS can be achieved.

5.3 Critical reflection of the method

The use case revealed insights concerning the RCC method. At the beginning, the result quality grows rapidly as the number of participants increases since more perspectives of experts are respected. Another insight for a small number of participants depicts that the development of the expense-benefit-ratio shows the same behavior as the result quality. But at a certain number of participants, the expense-benefit-ratio decreases rapidly. Hence the optimal expense-benefit-ratio can be achieved at the number from approximately 5 to 8 participants. This number depends on the company size, number of relevant requirements for production processes and number of implemented LP methods. Finally, the guidance items are not very detailed. The RCC method can therefore be re-developed to be more process-related. If the degree of granularity of the method is higher, more specific measures can be set for optimizing concrete processes.

6 CONCLUSION AND OUTLOOK

Digitalization technologies offer new opportunities for the productivity of LPSs, since elements of LP can be implemented differently. In order to identify the areas where an implementation of such technologies is most promising, LPSs need to be assessed first in regard to the identification of their deficits.

In order to derive an assessment method for LPSs, different elements of existing concepts used for the comparison of requirements and capabilities are combined and adjusted for this specific context. The RCC method determines the individual status quo of enterprises regarding the implementation of LP and reveals discrepancies between the current state and the target state of fulfillment of requirements with the capability of LP methods. Measures have been defined to realize the identified optimization potentials. In the end, a validation of the developed method demonstrates its practical applicability. In summary, a method for assessing the level of implementation regarding the LP methods has been developed. This method therefore reveals potential deficits in existing LPSs. These deficits can be indications for implementing digitalization technologies in order to support implemented LP methods.

The developed method assesses the implementation level of LP methods regarding their degree of fulfillment according to production requirements. In a subsequent step, this method could be developed further in regard to its level of granularity in order to identify potential deficits not only in the method layer but also in the process layer in LPSs.

7 REFERENCES

[1] Reinhart G., Zühlke D., 2017, Von CIM zu Industrie 4.0, In: G. Reinhart (Ed.): Handbuch Industrie 4.0. München, Hanser:XXXI-XL.

[2] Staufen AG, 2016, 25 Jahre Lean Management.

[3] Kolberg D., Zühlke D. 2015, Lean Automation enabled by Industry 4.0 Technologies. IFAC, 48/3:1870-1875.

[4] Ohno T., 1988, Toyota Production System. New York, Productivity Press.

[5] Womack J.P., Jones D.T., Roos D., 1990, The Machine that Changed the World. New York, Macmillan Publishing Company.

[6] VDI e.V. 2870-1, 2012, Beuth, Ganzheitliche Produktionssysteme. Grundlagen, Einführung und Bewertung. Berlin.

[7] Dombrowski U., Richter T. 2016, Ganzheitliche Produktionssysteme und Industrie 4.0. ZWF Zeitschrift für wirtschaftlichen Fabrikbetrieb, 111/12:771-774.

[8] Roy D.T., 2017, Industrie 4.0 - Gestaltung cyber-physischer Logistiksysteme zur Unterstützung

des Logistikmanagements in der Smart Factory. Diss. TU Berlin (2017). Berlin, Universitätsverlag TU Berlin.

[9] Meudt T., Metternich J., Abele E. 2017, Value stream mapping 4.0: Holistic examination of value stream and information logistics in production. CIRP Annals - Manufacturing Technology, 66:413-416.

[10] Pokorni B., Schlund S., Findeisen S., Tomm A., Euper D., Mehl D., Brehm N., Ahmad D., Ohlhausen P., Palm D. 2017, Produktionsassessment 4.0. ZWF Zeitschrift für wirtschaftlichen Fabrikbetrieb, 112/1-2:20-24.

[11] Wagner T., Herrmann C., Thiede S., 2017, Industry 4.0 impacts on lean production systems, In: Elsevier B.V. (Ed.): Procedia CIRP, 50th CIRP Conference on Manufacturing Systems:125-131.

[12] Karlsson C., Åhlström P. 1996, Assessing changes towards lean production. International Journal of Operations & Production Management, 16/2:24-41.

[13] Dombrowski U., Ebentreich D., Krenkel P., 2016, Impact analyses of lean production systems, In: Elsevier B.V. (Ed.): Procedia CIRP, 49th CIRP Conference on Manufacturing Systems:607-612.

[14] Kamprath N. 2011, Einsatz von Reifegradmodellen im Prozessmanagement. HMD Praxis der Wirtschaftsinformatik, 48/6:93-102.

[15] Akkasoglu G., 2013, Methodik zur Konzeption und Applikation anwendungsspezifischer Reifegradmodelle unter Berücksichtigung der Informationsunsicherheit. Diss. FAU Erlangen-Nürnberg (2013).

[16] Ahlemann F., Schroeder C., Teuteberg F., 2005, Kompetenz- und Reifegradmodelle für das Projektmanagement. Osnabrück.

[17] Fraser P., Moultrie J., Gregory M., 2002, The use of maturity models / grids as a tool in assessing product development capability: a review, In: IEEE (Ed.): Proceedings, International Engineering Management Conference:244-249.

[18] Bruin T. de, Freeze R., Kulkarni U., Rosemann M., 2005, Understanding the Main Phases of Developing a Maturity Assessment Model, In: ACIS (Ed.): Proceedings, 16th Australasian Conference on Information Systems.

[19] Knorr C., Friedrich A., 2016, QFD - Quality Function Deployment. München, Carl Hanser.

[20] Allweyer T., Knuppertz T., Schnägelberger S., 2009, EDEN - Reifegradmodell.

[21] Höltz N., 2012, Lean Logistics Maturity Model. Diss. BTU Cottbus (2012).

[22] Jørgensen F., Matthiesen R.V., Nielsen J., Johansen J., 2007, Lean Maturity, Lean Sustainability, In: J. Olhager et al. (Ed.): Advances in Production Management Systems, International Federation for Information Processing. Boston, Springer:371-378.

[23] Nightingale D.J., Mize J.H. 2002, Development of a Lean Enterprise Transformation Maturity Model. Information Knowledge Systems Management, 3/1:15.

[24] Meis J.-F., 2017, Produktionsseitiges Anforderungsmanagement. Diss. TU München (2017).

[25] Ouyeder O., Straube F., 2011, ModuLoSys - Modularisierung logistischer Systeme. Berlin, Universitätsverlag TU Berlin.

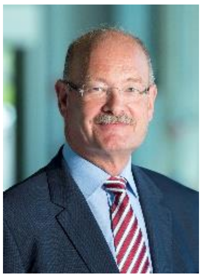
[26] Mayer A., 2007, Modularisierung der Logistik. Berlin, Universitätsverlag TU Berlin.

8 BIOGRAPHY

Georg Hoellthaler, M. Sc., graduated in Mechanical Engineering at the Technical University of Munich. Since 2016 he performs as a research associate at the Fraunhofer IGCV. In the department of production planning and control he focusses on the realization of digitalization in production companies.

Sebastian Weissenborn, M. Sc., graduated in Mechanical Engineering at the Technical University of Munich. Since 2018 he works as a consultant at the valantic SCE AG. In the department for logistics and supply chain planning, he optimizes logistical processes of enterprises.

Prof. Dr.-Ing. Stefan Braunreuther is head of the department for production planning and control at the Fraunhofer IGCV in Augsburg. Since 2017 he is Professor at the Augsburg University of Applied Sciences, where he lectures regarding topics like factory operations and production management.



Prof. Dr.-Ing. Gunther Reinhart holds the chair for Industrial Management and Assembly Technologies at the *iwb* of the Technical University of Munich. Furthermore, he is Executive Director of the Fraunhofer IGCV in Augsburg. The *iwb* and the Fraunhofer IGCV conduct research into assembly technology, automation, robotics as wells as production management and logistics.

Cloud-Based Automation Architecture for Competitive Manufacturing Systems

C. Faller[1], L. Friedrichsen[1], P. Stenkamp[1]
[1]Department for Automation Engineering,
Campus Velbert-Heiligenhaus, Hochschule Bochum, Germany

Abstract
The competitiveness of future manufacturing systems depends on their flexibility and availability. The development of the last years within the topic of Industry 4.0 have shown that the convergence of the information technology and operational technology (IT-OT-Convergence) is key to ensure this aim. Cyber-Physical-Systems (CPS) and Industrial Internet-of-Things (IIoT) are core technologies to realize the IT-OT-Convergence. This paper shows a future-oriented, decentralized architecture model for CPS based on the synergy between elements of classical automation technology and modern IT paradigms to leverage service-orientation down to the edge-level.

Keywords
Industrie 4.0, Cyber-Physical-Systems, Industrial Internet of Things, Cloud Services, System Theory

1 INTRODUCTION

Looking at the developments within automation technology over the last few years, it becomes obvious that the trend is increasingly moving towards individual production ('batch size one', cf. e.g. [1]).

Along with the currently ongoing shift from an industrial to post-industrial market in the consumer industry, manufacturing is, in direct consequence, undergoing drastic changes as well.

Whereas previously in a quasi-stable market situation with a more or less guaranteed acceptance of goods the focus in production was on maximizing output volumes while simultaneously minimizing costs, today flexibility and versatility are becoming increasingly important. Manufacturers need to be able to adapt to customer wishes as quickly as possible in order to remain competitive in an oversaturated environment [2].

This shift is, however, accompanied by the need to rethink existing automation technology paradigms such as the traditional automation pyramid in order to meet the requirements imposed by the recent market developments.

In this context, blueprints based on the further evolution of the existing approaches can be found, for example, in the work of the 'Industrie 4.0 Platform' in the form of the 'Reference Architecture Model for Industry 4.0 (RAMI4.0)' [3] or in the research agenda CPS of the VDI/VDE society of measurement and automation technology [4].

In this paper, however, a more far-reaching approach is presented, which, instead of merely focusing on the further refinement of existing models, synergistically integrates principles of modern software development.

2 CONVENTIONAL DESIGN OF AUTOMATION SYSTEMS

Historically, conventional production systems were initially designed with the goal to optimize repetitive, always identical processes. As aforementioned, since the acceptance of goods was guaranteed within the scope of the manufacturer's acquisition potential, priority was given to maximizing the output quantities while at the same time minimizing the cost. Due to the latter, all elements not essential for the production process were streamlined or eliminated.

This resulted in a lean, highly specialized architecture with real-time capabilities as the main focus, as the individual manufacturing steps were highly synchronized to maximize output. This involved the monolithic centralization of the process logic on the field layer to further reduce the reaction times as well as the system's complexity as the core principle.

This pattern is also reflected in the traditional automation pyramid. Even though today, layers 3 and above are gradually dissolving and becoming increasingly decentralized (as described in [4]), the control layer and below remain largely unchanged. This remains true as well for the improved model as defined in the RAMI4.0 guidelines [3], which adds additional life-cycle and infrastructure perspectives to the automation pyramid (making it a three-dimensional cube) and proposes changes concerning the hierarchical structure, e.g. logically merging the field and the control level, but does not break up the centralized process logic.

The problem with this is, that for a complex and highly dynamic market situation where flexibility and versatility are key to stay competitive [5], such a monolithic architecture is far too rigid. Today, the strong specialisation in the Taylorist industrial model is proving counterproductive in terms of adaptability

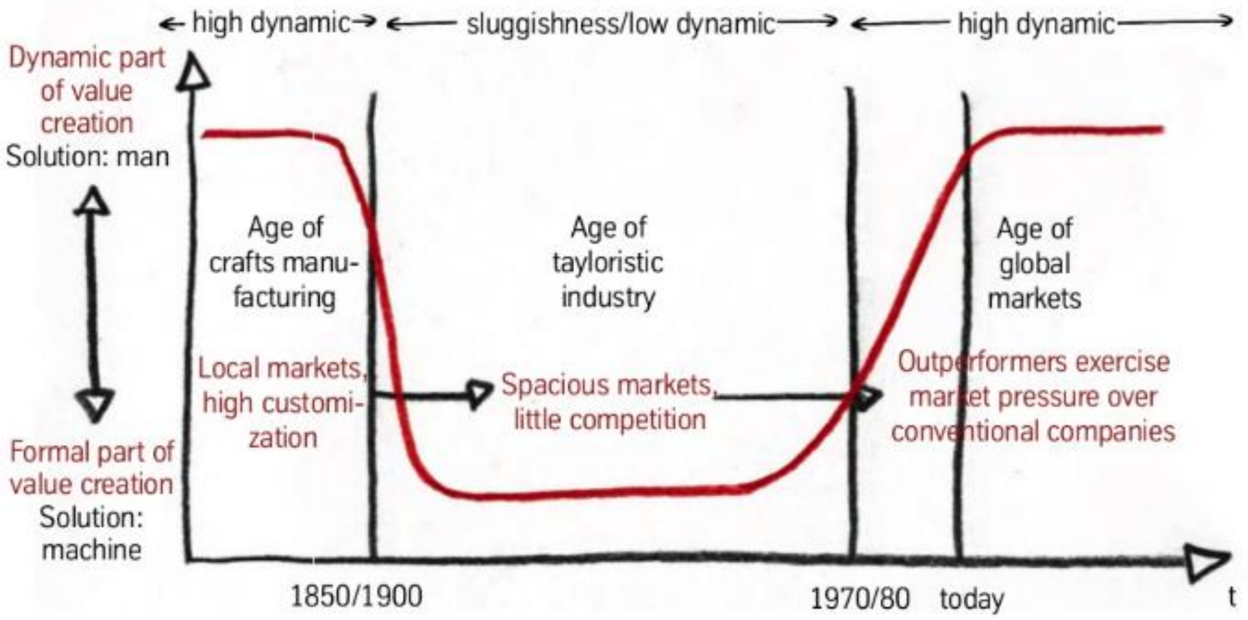


Figure 1 - The historical course of market dynamics and the recent rise of complex and dynamic markets [2]

since the original framework conditions no longer apply. New, alternative approaches are needed.

3 DECENTRALIZATION AND MODULARIZATION

One such alternative and sensible approach with regard to system flexibility requirements consists in the modularization of process logic.

This methodology is in line with the procedures established in the IT industry for the development of complex, highly dynamic software systems (cf. e.g. 'Microservices Architecture Pattern', [6]).

Contrary to the widespread assumption in automation technology, decentralization is not necessarily synonymous with a decline in system safety. Although the inherent complexity increases, this is only a consequence of the fact that a system cannot be depicted losslessly by a model of a lower complexity level [5]. This measure does thus not increase the general susceptibility to errors any more than the retention of the existing monolithic design with respect to the changing environment would do. Merely the type of errors changes; meta-level deficiencies are replaced by potential technical deficiencies.

While the former eventually proves to be fatal from a business perspective, the latter is preventable. As modularization is a common procedure in software development, there are already many guidelines that describe how to create fault tolerant distributed systems (cf. e.g. [7]).

Consequently, it comes down to the decision whether to stick to the centralized approach in favor of a lower system complexity and thus to hazard shortcomings on the strategic side, or whether to adapt to the shifting market situation and accept the inherent, yet avoidable technical pitfalls associated with decentralization of process logic.

If one decides in favor of modularization, it is furthermore intuitive to detach oneself from the coventional models coupling hierarchical layers to physical equivalents as they reach their limits with regard to service-oriented architectures (as illustrated in [4]). Consequently, this means to move away from the traditional automation pyramid to a perspective that focusses mainly on organizational dependencies.

In the following, one approach to achieve decentralization of process logic is shown using the development of a testbed for innovative architectures in automation technology as an example.

4 CASE STUDY: CONCEPT OF A CLOUD-BASED ARCHITECTURE FOR MANUFACTURING SYSTEMS WITH CASCADED PROCESS LOGIC

4.1 Background

The aforementioned testbed was developed as part of the research into future-oriented architectures for competitive manufacturing in the department for automation engineering at the Bochum University of Applied Sciences, Campus Velbert-Heiligenhaus. It is intended to serve as a basis for investigating the effects of service-oriented architecture in automation technology under realistic conditions.

4.2 Architectural design

Following the previous argumentation, a modular, decentralized and logically cascaded approach was chosen as the basis for the implementation.

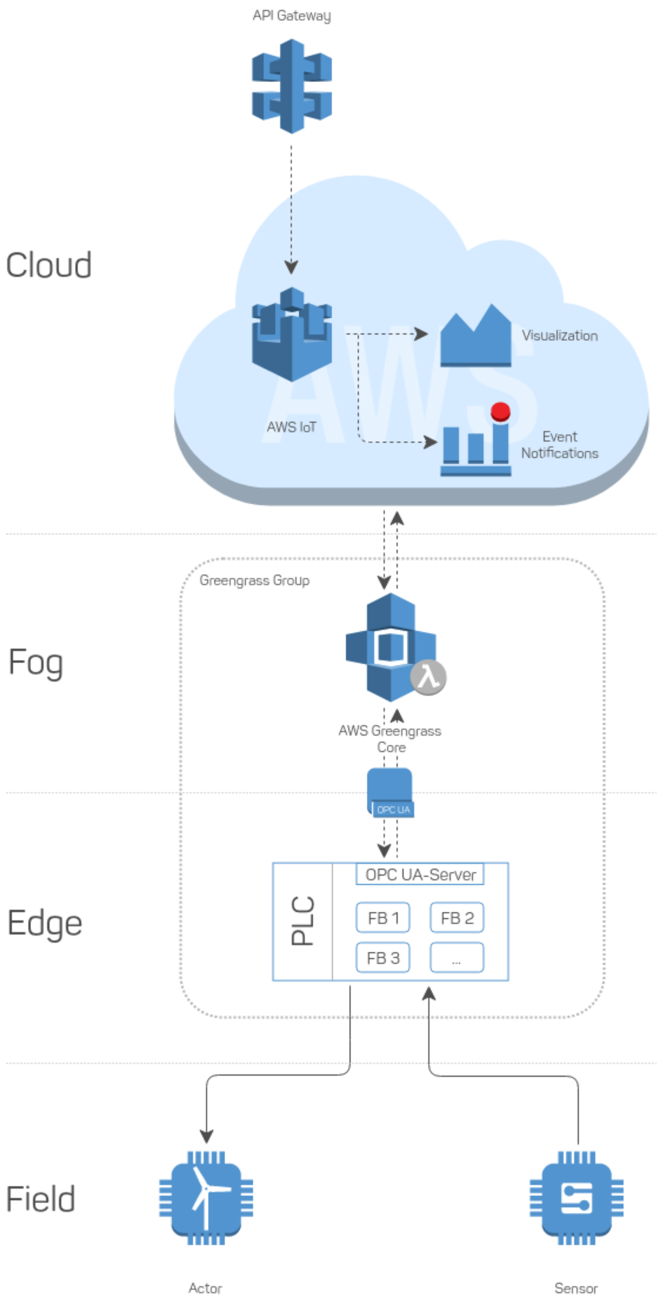


Figure 2 - Conceptual architecture with layers

As illustrated in figure 4-1, the organizational layout is based on the infrastructure layers known from software development at its core. Contrary to the traditional automation pyramid, this model does not imply immanent hierarchical links between the individual levels.

In order to achieve the greatest possible flexibility with regard to the extensibility of the testbed and to facilitate the subsequent integration of external services, a cloud solution was chosen for the back end. In this this case, the selection process favored **Amazon Web Service (AWS)**, since its IoT-stack supports the chosen decentralized approach with **AWS IoT Core** and **AWS Greengrass** well across all levels of infrastructure.

The latter serves to leverage part of the service landscape from the cloud to the fog layer, thus not only providing offline capabilities and drastically reducing latencies – which is crucial in terms of systems safety – but also enabling a service-oriented programming paradigm in a local environment. Furthermore, it offers an integrated deployment mechanism for software updates. This is an important feature as it allows for the application of subsequent security patches as well as for fast validation of effects from measures.

On the field layer, it was decided to use the **Modular Production System** by Festo Didactic as it is well suited for a testbed application due to its modularity and extensibility.

To ensure maximum flexibility with regard to the later adaptability and modifiability of the application, no conventional PLC running proprietary software was used. Instead, the setup is controlled by a Linux based programmable logic controller (PLC), specifically a **RevPi Core 3** by KUNBUS, providing full, unhindered access to all underlying processes.

As shown in figure 4-1, an OPC UA server based on the open source project **open62541** was set up on the PLC as an interface between the edge and the fog layer. OPC UA is well suited for this task since it already supports asynchronous and event-based communication by default, which is a core requirement for a service-oriented architecture.

Special to this implementation is, that it was carried out right from the start according to the 'API first' principle by making use of the 'UA Methods' defined in the OPC UA standard. The result is a so-called 'API as a product' [8], an abstraction layer against which can be developed from the outside and which, in contrast to a pure mirroring of the PLC's input and output states on the server, allows for elements to be facaded towards the outside. Thus, for example, all real-time critical program sections can run locally encapsulated with guaranteed safe execution, while the actual process flow is defined via the API at the fog or cloud layer. This enables effective cascading of the logic across the various infrastructure levels. From the perspective of a manufacturing system, such a design is highly advantageous, as it significantly increases the immanent flexibility and safety in a service-oriented architecture.

In the following, the OPC UA information model as well as the cloud-based services and the communication interface between the fog and the edge layer shall be discussed in more detail.

4.3 Implementation

4.3.1 OPC UA information model

Building on the aforementioned ‘API first’ principle, the key requirements for the creation of the OPC UA information model were as follows: First, all IOs of the PLC shall be accessible through open62541 while also being easily configurable without changing the

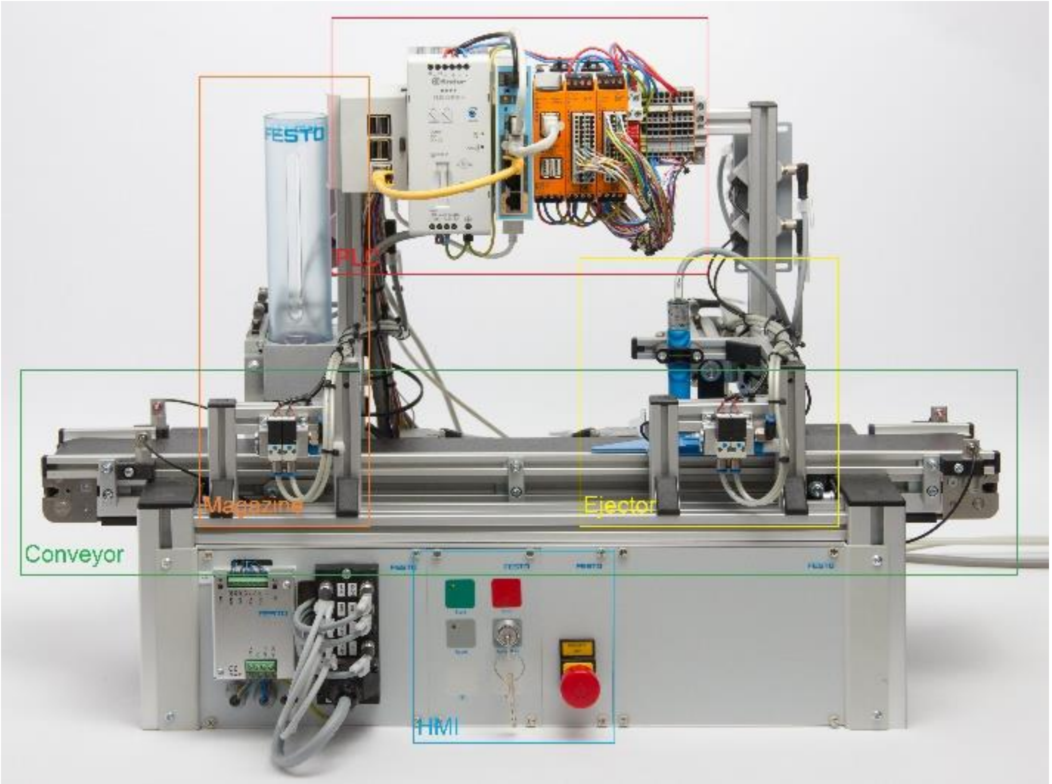


Figure 3 - The testbed, build on the Festo Modular Production System and a RevolutionPi 3 with additional IO

code of the server or editing configuration files on the PLC. Secondly, all IOs as well as the states of core modules like indicators, sensors and actuators have to be provisioned with getter- and setter-methods to provide a basic level of API.

Individual IOs shall not be writable directly by a client, but rather be linked with the core modules they belong to. These core modules can perform basic functions which they provide in form of methods to clients via OPC UA. Picture a pneumatic cylinder which is controlled by two valves and which position can be monitored via two sensors. Not only does the information model have to represent the position of the cylinder; it also has to provide methods to command a change of position. Therefore, the PLC needs to know what the current position is, which inputs the sensors are connected to and which outputs must be set to a given state to move the cylinder to the target position.

This information shall be provided by the model itself to allow for easy adaptation and modification of the system by application engineers while sparing them the time and effort changing and testing the program itself would require. This is achieved by utilization of OPC properties containing all the necessary information. These properties are evaluated at runtime by the server. If necessary and permitted, these can be edited at runtime via the same tools used to monitor the system (e.g. a regular OPC UA client), allowing for a reconfiguration of individual modules without shutting down the overlying system.

At the current state of implementation, the information model only provides methods for the IO and core modules as well as for safety critical elements. Higher level functions, for example the process logic required to transform a pneumatic cylinder and a number of sensors into a system module like a part magazine, are implemented at the fog and cloud layer.

4.3.2 Cloud-based services and communication architecture

As mentioned before, the core cloud services used in the testbed application are AWS IoT Core and AWS Greengrass.

The former provides the classic features of an IoT platform like secure communication, data processing and routing, and device management. In the implementation process, it was used for the creation of a so called ‘Digital Twin’ (cf. e.g. [3]) of the testbed, meaning a virtual representation of the physical object, as well as an interface to other cloud services.

For this purpose, the submodules of the MPS were recreated in the device management section as virtual devices. Through the constant synchronization with Greengrass and under usage of the AWS Shadow functionality, they were then configured to be updated automatically when a state change on the physical testbed is registered via the OPC UA server. As a result, the state of the digital twin is always consistent with the ‘real’ MPS. Other cloud services can access the virtual devices and thus obtain the live data from the manufacturing process.

The other way around, it is also possible to define a so called ‘desired’ state of the digital twin through the Shadow protocol. This will cause a gateway application – an AWS Lambda function hosted locally on the Greengrass device (cf. figure 4-1) – to be notified which then again resolves the specified condition into a function call and invokes the respective method on the OPC UA server. Thus a fully bidirectional synchronization between the virtual representation and the physical testbed is achieved. Since the whole communication process is event based, this harmonizes well with a service-oriented architecture.

For demonstration purposes, the application was additionally extended by a service-based automatic mode for the MPS as well as by a voice control option based on Amazon Alexa. The former is hosted locally on the Greengrass device alongside the OPC UA facing gateway in order to ensure fail-safe operation even in the event of a network failure while the latter is located in the cloud and serves as an example for the integration of further services into the described architectural design.

5 CONCLUSION

Already during the implementation process, several positive aspects of the chosen architectural layout became apparent: The high degree of modularization allowed for a parallel and agile development of core software components.

Due to the 'API first' design approach and the resulting cascading of functional entities like the core modules of the MPS, changes in the process logic of the PLC could be made without requiring an adaptation in the information model or the gateway functions and vice versa. One disadvantage of this approach however is the redundant curation of the information model on the edge and the fog layer. These are currently modelled in different formats, although this could possibly be resolved by the use of an additional abstraction layer or common format in the likes of AutomationML (cf. [9]) in the future.

Another big advantage of this highly modular design is the minimal effort required to integrate additional services on all layers from edge to cloud into the application. The deployment of these services on the AWS infrastructure as well as the local Greengrass instance has proven itself as very versatile and intuitive.

Overall, the immanent synergies between automation technology and IT became apparent and proved to be very advantageous for the development process. For example, the utilities provided by the latter greatly facilitated the state synchronization between the 'real' testbed and the digital twin on the technical level.

Concluding, the chosen approach has proven very promising with respect to the initially formulated requirements. The decentralisation of process logic significantly enhanced the flexibility of the system. In general, the adaptability increased drastically due to the service-oriented architecture. Choosing a cloud-stack as the backend for the application augmented this aspect even further since the extensive service landscape available makes it possible to deploy and evaluate changes in a fast and simple way.

Even though, while this work provides first findings regarding the feasibility of such a design approach in the domain of industrial automation, it does this only in the scale of a small and independent system. While this is sufficient in the context of an initial case study, a quantitative analysis of the scalability as well as potential risks of this approach in a production environment require further research.

Nevertheless, one thing that this case study shows for sure is that in the future the job profile of the automation technician will change drastically. With eventually increasing system complexity, the number of elements from conventional software development in this area will rise as well. In the long run, at least a partly merging of both domains, also referred to as IT-OT-Convergence, is inevitable. On the other hand, some traditional components of automation technology like the automation pyramid will most likely be abandoned or evolved as their initial framework conditions no longer apply.

From a corporate perspective, this means that employees have to be actively educated and trained to enable them adapt to the shifting circumstances. However, as assimilating such disruptive changes is a very time-consuming process, it is advisable to start as soon as possible.

6 REFERENCES

[1] NN: Implementation Strategy Industrie 4.0 – Report on the results of the Industrie 4.0 platform, Pages 26-28, Bitkom, VDMA, ZVEI, 2015

[2] Pfläging, Nils, Herrmann, Silke, Vollmer, Lars, Carvalho, Valérya: Organize for Complexity, BetaCodex Network White Paper No. 12, BetaCodex Network, 2012

[3] NN: Implementation Strategy Industrie 4.0 – Report on the results of the Industrie 4.0 platform, Pages 40-70, Bitkom, VDMA, ZVEI, 2015

[4] NN: Cyber-Physical Systems: Chancen und Nutzen aus Sicht der Automation, VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik, 2013

[5] Snowden, Davd J., Boone, Mary E.: A Leader's Framework for Decision Making, Harvard Business Review, 2007

[6] Wolff, Eberhard: Microservices – Grundlagen flexibler Softwarearchitektur, dpunkt.verlag, Heidelberg, Baden-Wuerttemberg, 2016

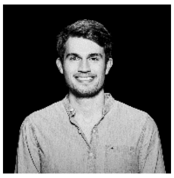
[7] Hanmer, Robert S.: Patterns for fault tolerant software, John Wiley & Sons Ltd, England, 2007

[8] NN: Technology Radar Nov ‘16, Page 5, ThoughtWorks Inc., 2016

[9] Faller, Clemens., Höftmann, Max.: Service-oriented communication model for Cyber-Physical-Production-Systems, 11th CIRP Conference on Intelligent Computation in Manufacturing Engineering, 2017

7 BIOGRAPHY

Prof. Dr.-Ing. Clemens Faller studied Mechanical Engineering at the Ruhr Universität Bochum and researched for his doctorate in the field of industrial communication standards. In 2007 Mr. Faller changed to the French company Schneider Electric. Since April 2013 he is professor for automation technology and researches in the field of digitization of production, industrial IT and industrial communication technologies.



Lukas Friedrichsen obtained his bachelor’s degree in Engineering (Mechatronics and Information Technology) from the Bochum University of Applied Sciences as part of a cooperative dual-study programme with integrated apprenticeship. He is currently working as a research assistant in the Department of Automation Technology while simultaneously completing his master's degree.

Philipp Stenkamp holds a bachelor’s degree in Engineering (Mechatronics and Information Technology) from the Bochum University of Applied Sciences. He is currently working as a research assistant in the Department of Automation Technology while pursuing his master's degree.

Classification and Use of Product Related Services in the Automotive Industry

D. Palm[1], K. Fedorovskaia[2], L. Narinsky[3], E. Brinster[1]
[1]ESB Business School, Reutlingen University, Germany
[2]Accenture GmbH, Austria
[3]T-Systems Austria GesmbH, Austria

Abstract
Because of saturated markets and of the low profit margins in the sales of cars, car manufacturers focus more and more on profitable product related services. This paper deals with the question how to classify product related services in the automotive industry and which characteristic product related services are offered to the end-users (consumers) in a standardized format. Two research studies on the provided product related services in 2010 und 2017 by 15 car manufacturers and 20 exemplary automotive brands in Germany revealed that the application degree by the OEM (original equipment manufacturers) in these years increased considerably. While in 2010, the average range of services only amounted to 33%, the value in the automotive industry increased until 2017 to 57%.

Keywords
Product related services, automotive, classification

1 INTRODUCTION

According to Bullinger, product related services (PRS) are those services, which are provided by producing enterprises to respond to the need or to solve a problem of the customer in connection with the product [1]. This service is provided in addition to the primary product pre-sales, sales, or after sales and either refers to the primary product lifecycle [2] or is independent on it [3]. The automotive industry mainly focuses on the primary product "car" even if some car manufacturer adapt the primary product to "mobility" in response to the changed customer behaviour in recent years [4,5]. In this respect, terms like "Product Service System (PSS)" or "Customer Solutions" were defined in literature. It designates the marketable bundle of product and services which is characterized by the fact that the benefit for the customer is higher than the value of each component in common [6]. In delimitation thereof, PRS in the automotive industry are defined as services that can be marketed as a service product in connection with the car in all phases of the product lifecycle or independent on it, which in this way generates an added-value to the primary product for the user to satisfy a need of the customer or to solve a problem of the customer. Hence, the sales and distribution of new or used cars in this sense is a primary performance and no product related service.

Due to saturated markets and low profit margins in the sales of cars, the car manufacturers see their profitability focus more and more in PRS [1]. The objective of this document is to analyse the offer of product related standardized services by the car manufacturers to private end-users and to identify the changes in the last years. To answer this question the public offer (websites of the manufacturers) of different OEM was analysed. A systematic analysis requires to first describing the scope of PRS in the automotive industry. For this purpose, first a scheme of features was generated. This feature scheme forms the basis for the validation of the PRS of the automotive industry as well as for the statistic execution of a typology.

2 TYPOLOGY OF SERVICES

To be able to have a closer look at the diversity of services it is necessary to classify them in different groups. The segmentation of services knows different classifications or typologies. A classification is the result of a division of objects in groups or classes. The objects are grouped by the characteristics of selected features. The characteristics form the classes. [7]

In a typology, features and characteristics of features can have relations and unclear boundaries to each other. In contrast to a classification, a typology has the additional property that it does not need to be complete. Some objects can belong to several types and others in turn do not belong to any type at all. In spite of the disadvantage of incompleteness, a topology suits to structure complex and vast objects by far better than a classification. [8]

To generate a model for the product related services of the automotive industry, the typology method is used for classification, for the following reasons:

- PRS show a large number of features.
- Between the features, strong relationships may occur. This property is identified by the typology types, which are described below.
- With services in general and with PRS in particular, multiple characteristics of features can be fulfilled in one feature.

Typologies and services in the literature are divided into two- or multidimensional models. Schmenner [9] divides the services by the features “working intensity” and “interaction intensity” and describes four types of services with two characteristics, each (high, low). The service typology of Barth, Hertweck and Meiren [10] divides the services by the features “contact intensity” and “variant diversity” and refer to the same four types. The feature scheme of Jaschinski [11] (see Figure 1) was derived from the effect carriers of the service products which are used by the company, the external factors which are mostly coming from the customers in the production process, the interaction of factors with each other as well as the result and profit of the service production.

Feature Scheme for the classification of services			
Product type	Individual product	Module product	Standard product
Main application factors	Human performance	Technic, devices	Information-/ Communication systems
Main object of service	Customer	Material objects	Immaterial objects
Product scope	Individual performance	Performance package	
Product type	End-user-/ consumer related	Enterprice related	
Planning of ustomer order	Short (< 1 day)	Middle (< 1 month)	Long (> 1 month)
Provision period	Short (< 1 day)	Middle (< 1 month)	Long (> 1 month)
Interaction site	Supply oriented	Demand oriented	Separate location
Process stability	Low	Middle	High
Role of customer	Actor	Audience	Without direct participation

Figure 1 - Typology of services [11].

3 DEVELOPMENT OF A FEATURE SCHEME FOR PRODUCT RELATED SERVICES OF THE AUTOMOTIVE INDUSTRY

The cluster analysis is a procedure, which summarizes the investigated objects in groups [12]. This procedure recognizes patterns, structures and relationships in a comprehensive and possibly vast area, from which further information can be gained and handling fields be recognized. The objective of cluster analyses is to create groups as homogenous as possible in itself and as heterogeneous as possible to other groups. Consequently, the objects in a group should be as similar as possible with reference to the used feature characteristics. Objects, however, which belong to different groups, should be as different as possible in their characteristics. [12]

The range of the PRS of the automotive industry forms the solution space for the generation of the feature scheme. A solution space contains all sources of revenue of the relevant area. Consequently, the solution space includes all PRS, which are provided by the automotive industry. The following three areas are used to determine the features in this space:

- Services,
- Product related services and the
- Automotive industry.

The PRS frame is a subset of the service space. The PRS of the automotive industry which, according to the set theory, form the intersecting set between the automotive industry and the PRS space, are a subset of the PRS range.

$PRS \subseteq SR$

$(AI \cap PRS) \subseteq PRS$

PRS … Product related service range

SR … Service range

AI … Automotive industry

Analog to the theory of inheritance of properties, a subset is described by subset features as well as by other specific features of the subset. It describes the PRS range in the automotive industry by service-, product related service range and by domain-specific features. The form of this inheritance is shown in Figure 2.

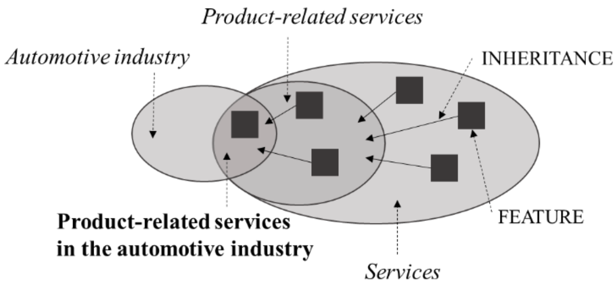


Figure 2 - Definition of product related service range in the automotive industry [13]

The investigation by features is done separately in the three observation spaces. Then, the typology of Jaschinski was transferred to the domain of the automotive industry where the non-relevant features were eliminated and the following features were derived:

- Internal factors (from main application factors)
- External factors (from main object)
- Product scope
- Provision period
- Interaction location with the occurrences: service provider, customer, separated location and footless services (customer and service provider go to the location where the service is executed)
- Role of customer

In addition to the features from the typology acc. to Jaschinski, other features were found in the areas of service production and product related services. The feature “sales form” describes in how far a connection exists between two product related services in sales. If a sales related connection to another PRS is missing then it is designates as “independent”, else as additional service (add-on). An example would be a replacement car in connection with the repair of the own car.

“Sales cycle” describes the time point of service provision in the sales cycle of the main product and/or the car. The sales cycle is divided into the phases Pre

Sales, Sales und After Sales [14]; PRS, however, can also be offered independently on the sales cycle.

The need of a customer often goes beyond the purchase of a tangible good. To provide the customer with an added value and to meet the requirement of the customer in its totality, PRS are offered in addition to the object. The offer of PRS orients to the customer‘s benefit and can be subdivided into the following five customer benefits:

- Product (additional products to main product, e.g. merchandising articles)
- Function /Product oriented (functional capability of the product, e.g. oil service or cosmetic repairs)
- Benefit oriented (e.g. insurances)
- Result oriented (e.g. Roadside-Assistance)
- Service (pure service, e.g. pick-up and delivery service)

In addition to the already described features, further features were determined from the literature referring to the domain of the automotive industry. Services, which offer mobility benefits, provide the users with human and technical performances to satisfy cognitive and emotion dominant mobility needs. These needs can be divided into three areas [15]:

- Mobility creating services (e.g. leasing)
- Mobility securing services (e.g. full comprehensive insurance)
- Mobility extending services (e.g. telemetry services)
- Mobility independent services (e.g. plant or factory tour)

The feature “car type” describes in how far the product related services are linked with the primary product (car) of the automotive industry.

From the customer’s point of view, a new car is understood to be a purchased car, independent on new or used. These features show those product related services, which can only be used in connection with the purchase of a car. To them belong, among others, car credits, credit protection insurances, and guarantee contracts.

To the area “rented cars” belong all product related services where rented cars are delivered as internal production factors. Examples are travel events where the customer gets a car. The feature “own car” includes those product related services which are linked with the ownership of a car and which are primarily offered in the After Sales phase. To this area belong, among others, repair, maintenance and sales of spare parts and accessories.

Figure 3 gives a summary of the feature scheme for the classification of product related services in the automotive industry.

Feature Scheme for the typology of product-related services in the automotive industry					
Service-specific features					
Internal factors	Human performance	Technik, devices	Information-/ Communication systems		
External factors	Customer	Material objects	Immaterial objects		
Product scope	Individual servce	Service package			
Provision period	Short (< 1 Tag)	Middle (< 1 month)	Long (> 1 month)		
Interaction site	At service provider	At customer	Separate location	Footless services	
Role of customer	Actor/active	Audiance/passive	Without direct participation		
Planning of customer order	Short (< 1 day)	Middle (< 1 month)	Long (> 1 month)		
Product-related specific features					
Sales form	Additional service (Add-on)	Independent			
Sales cycle	Pre Sales	Sales	After Sales	Independent	
Customer benefit	Product	Function oriented	Application oriented	Result orientged	Service
Domain-specific features					
Mobility	Mobility creating	Mobility securing	Mobility extending	Mobility independing	
Car type	New car	Own car	Rented car	None	

Figure 3 - Feature scheme for classification of PRS in the automotive industry [13]

The first survey was executed from October to December 2010 with 15 car manufacturers and 20 brands. The companies were selected by their sales quantity in 2008 and included American, Asian, and European manufacturers to also include cultural differences. When selecting the brands, attention was given to the representation of all market segments. In total, the first survey identified 658 product related services, which were offered on the German market. A second survey followed from April to June 2017 with the same selection of automotive companies and brands on the German market.

The data was checked for statistical independence. The calculation of the cluster based on the distance scale of Lance and Williams and on the Ward Clustering procedure, and first formed nine clusters (2010) which was expanded because of the extended service offer to ten (2017) for PRS in the automotive industry.

4 ANALYSIS OF PRODUCT RELATED SERVICES IN THE AUTOMOTIVE INDUSTRY

With reference of the unused potential, different analyses can now be executed. First, a brand-specific offer of PRS in relation to the total offer of all brands can be presented. Furthermore, a delta results between the total potential used in the automotive industry, and the actually possible theoretic potential, as well as a potential of the future need. The three potentials are shown in Figure 4.

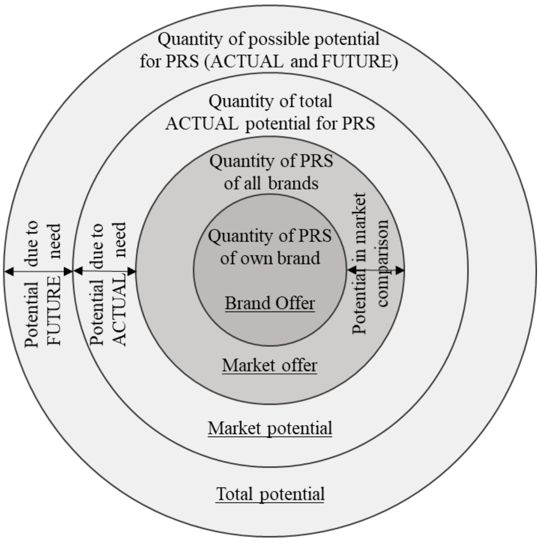


Figure 4 - Potential product related services (PRS) in the automotive industry [13]

In response to the brand-specific use compared with the potential total offer of the automotive industry, a definition of the characteristic PRS is necessary.

After elimination of context-related similar services, 60 characteristic PRS of the automotive industry were defined (see Figure 5). They represent the current total offer of the automotive industry (in Germany).

Characteristic product related services (PRS) in the automotive industry

Type 1 - Non-automotive related PRS Number: 12
- change requests for existing contracts
- car sharing
- account and securities transactions
- credits
- charging station
- merchandising
- online games
- parking aid
- publications
- phone contact
- ticket sale
- insurance

Typ 2 - Repair and maintenance Number: 11
- installation and modification
- storage
- express service
- collection and delivery service
- inspection
- roadside assistance
- car care
- repair
- restoration
- accident handling
- maintenance

Typ 3 - Adventure Number: 7
- adventure park
- vehicle pick-up
- guided tours
- rental spaces
- museum
- restaurant
- events

Typ 4 - Travel and training Number: 3
- automobile travel
- driving trainings
- course

Typ 5 - Rental Number: 2
- replacement car
- rental car

Typ 6 - Insurance Number: 7
- gap insurance
- warranty extension
- car and fully comprehensive insurance
- instalment protection insurance
- legal expenses insurance
- racing insurance
- maintenance contrac

Typ 7 - Distribution of vehicle parts and accessories Number: 9
- driving aids
- navigation updates
- original parts
- care products
- wheels and tyres
- technical literature
- tuning parts
- certificates
- accessories

Typ 8 - Financing Number: 2
- loans
- leasing

Typ 9 - Telematics and information technology services Number: 4
- internet
- communication service / apps
- navigation service
- radio

Typ 10 - Conversion and processing of information Number: 3
- data collection
- anti-theft protection
- fault detection

Total Number: 60

Figure 5 - Potential product related services (PRS) in the automotive industry [13]

With reference to the analysis in 2010, this space increased from 54 to 60 used services, i.e. by 11%. This growth took place in Type 1 (non-automotive product related services) and in Type 10 (conversion and processing of information). The latter type with three characteristic services was not offered by the automotive industry in 2010 at all. The degree of use of the service frame in the automotive industry increased on the average for the observed 20 brands from 33% in 2010 to 57% in 2017. The best brand value in 2010 amounted to 67% and in 2010 to 90%.

The highest increase of a single brand between two observation dates was 40%. The offer of the investigated single brands is shown in Figure 6.

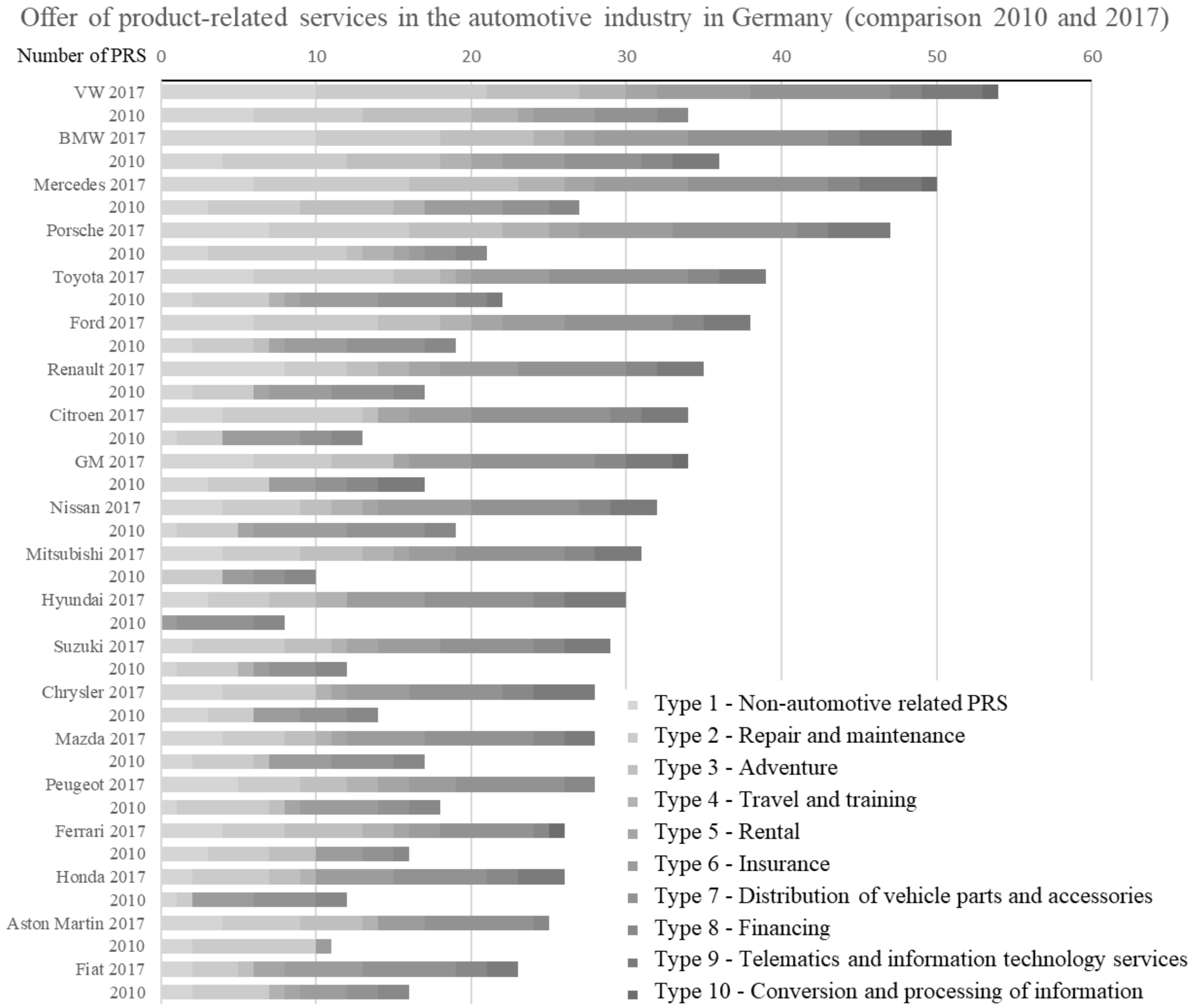


Figure 6 - Offered product related services in the automotive industry (number of PRS in Germany, comparison 2010 and 2017)

5 SUMMARY

The objective of this document is the introduction of a feature scheme to classify product related services (PRS) in the automotive industry and to define characteristic PRS in the automotive industry. This enabled to make statements about the degree of use in the automotive industry. The investigation was executed in 2010 and 2017. It could be stated that the degree of use by the car manufacturers in these 6.5 years increased significantly. While in 2010 only 33% on the average of the offered range was used, the value increased to 57% until 2017. In particular, the German car manufacturers increased their offer on the domestic market by 35% and achieved the highest value of about 90%.

PRS to use information and knowhow could be identified in the meantime in 2017 in comparison with 2010, although free-of-charge offers still prevail. While in 2010, the product range offered by the car manufacturers seemed partially to be rather wilfully, a structured proceeding seems to prevail in 2017. In total, it can be derived from the considerable efforts of all car manufacturers and brand to complete the offer, that the product related services will develop from a brand-differentiating feature to an important own competitive factor.

6 REFERENCES

[1] Bullinger, H.-J., van Husen, C.. Aktuelle Chancen und Trends im Servicegeschäft. In K. Barkawi, A. Baader, & S. Montanus, Erfolgreich mit After Sales Services - Geschäftsstrategien für Servicemanagement und Ersatzteillogistik. Berlin Heidelberg: Springer Verlag, 2006.

[2] Rösner, J. Service, ein strategischer Erfolgsfaktor von Industrieunternehmen? Hamburg: Steuer- und Wirtschaftsverlag, 1998.

[3] Kapmeier, J. Benchmarking in der Automobilindustrie: Der Service als Positionierungsinstrument eines Nutzfahrzeuges. Hamburg: Diplomatice Verlag, 2009.

[4] Silberer et.al. Mobile commerce – Grundlagen, Geschäftsmodelle, Erfolgsfaktoren. Gabler, Wiesbaden, 2002.

[5] Heinrichs et.al. Mobiles Internet: Auswirkung auf Geschäftsmodelle und Wertkette der Automobilindustrie, am Beispiel MINI Connected. In: Proff H., Schönharting J., Schramm D., Ziegler J. (eds) Zukünftige Entwicklungen in der Mobilität. Gabler Verlag, 2012.

[6] Böhmann, Tilo; Krcmar, Helmut: Hybride Produkte: Merkmale und Herausforderungen. In: Bruhn, Manfred; Stauss, Bernd (eds): Wertschöpfungsprozesse bei Dienstleistungen. Betriebswirtschaftlicher Verlag Dr. Th.Gabler, Wiesbaden, 2007, S.240–255.

[7] Dörner, D. Thought and Design - Research Strategies, Single-case Approach and Methods of Validation. In E. Frankenberger, P. Badke-Schaub, & H. Birkhofer, Designers. The Key to Successful Product Development. Berlin, New York: Springer Verlag, 1998, 3-11.

[8] Konrad, G. Theorie, Anwendbarkeit und strategische Potenziale des Supply Chain Management. Wiesbaden: Deutscher Universitäts-Verlag, 2005.

[9] Schmenner, R. W. How can service business survive and prosper? In C.Lovelock, Managing Services. Marketing, Operations, and Human Resources, Englewood Cliffs, Vol. 27, 1992, S. 31-42.

[10] Barth, T., Hertweck, A., & Meiren, T. Typologisierung von Dienstleistungen: Basis für wettbewerbsorientierte Strategien im Rahmen eines erfolgreichen Service Engineering. In H. Barske, A. Gerybadze, L. Hünninghausen, & T. Sommerlatte, Das innovative Unternehmen: Produkte, Prozesse, Dienstleistungen. Wiesbaden: Gabler-Verlag, 2000.

[11] Jaschinski, C. Qualitätsorientiertes Redesign von Dienstleistungen. Shaker Verlag, 1998.

[12] Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. Multivariate Analysemethoden - Eine anwendungsorientierte Einführung. Berlin Heidelberg: Springer-Verlag, 2008.

[13] Narinsky, L., Fedorovskaia, K.: Potentialanalyse für produktbegleitende Dienstleistungen der Automobilindustrie. TU Wien, 2011.

[14] Baethge, S. Ziele und Instrumente des After Sales Service. Norderstedt: Grin Verlag, 2007.

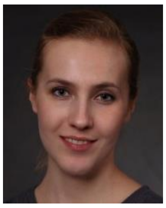
[15] Rennert, C. Dienstleistungen als Elemente innovativer Betreibungskonzepte im Automobilhandel. Ottobrunn: Autohaus Verlag, 1993.

7 BIOGRAPHY

Daniel Palm, Prof. Dr. techn., Dipl.-Ing., is professor at the ESB Business School, Reutlingen University, Germany for logistics management. He is a leader of the research centre for logistics and value creation at Reutlingen University.

Dipl.-Ing. Karina Fedorovskaia is IT Strategy Manager at Accenture GmbH, Austria. She graduated Industrial Engineering Computer Science at Vienna Technical University in 2011.

Dipl.-Ing. Leonid Narinsky graduated at Vienna Technical University in Industrial Engineering Computer Science. He is Service Delivery Manager at T-Systems Austria.



Elena Brinster, BSc graduated ESB Business School of Reutlingen University in 2017. She studied International Logistics Management and is currently at EDAG Engineering GmbH

Operational Site Planning and Development in Metropolitan Regions

M. Krause[1], D. Kik[2], P. Nyhuis[1], T. Spengler[2]

[1]Institute of Production Systems and Logistics, Leibniz University Hannover, Germany

[2]Institute of Automotive Management and Industrial Production, Technical University Braunschweig, Germany

Abstract

With increasing globalization, urbanization and rural depopulation, the development of metropolitan regions can be observed. They are considered to be of high economic and social relevance. Metropolitan regions are characterized by different spatial structures, whereby the location qualities of the different types of spatial structures diverge significantly. Companies are therefore faced with the challenge of identifying the potential within a region and further developing their locations in line with their requirements. In addition to aspects such as budget targets and opportunities of site development by municipalities, the interdependencies that occur between different site factors must also be taken into account. This article provides insights into a problem-adequate planning model to be developed for site planning and the improvement of existing sites. In addition, first results of the empirical survey of companies and municipalities conducted for this purpose will be presented.

Keywords

Spatial structures, MetroPlant, Integrated planning model

1 INTRODUCTION

Continuing megatrends such as globalization and urbanization are leading to changes in spatial structures. In this process the establishment and improvement of so-called metropolitan regions plays a decisive role [1]. Metropolitan regions are spatial and functional areas, which are considered to have an important function in economic and social development [2, 3]. High-performance infrastructures and dense production networks are typical characteristics of metropolitan regions [3]. With the associated high job and employment potential, the development of metropolitan regions is regarded as an important parameter for strengthening performance in international competition [1, 2]. However, with the different spatial structure types of urban, suburban and rural areas, there are explicit regional quality differences with regard to economically relevant location factors within a metropolitan region [4, 5].

Up to now, the task of operational location planning has been to select a location that offers optimal conditions for sustainable corporate success [6, 7]. Since a location is described by a multitude of location factors, which need simultaneous consideration, the location decision is usually multicriterial and thus a weighing decision between location factors in conflict with each other [8, 9]. The availability of space for new settlements in metropolitan regions is often limited, so that the development of existing locations is constantly gaining in importance. However, in order to identify the potential of the spatial structures in the metropolitan regions and to further develop locations according to one's own needs, it is also important to include company-related development measures. In addition, municipal-driven options for location development as well as interdependencies between regional location factors can be observed, which can change the location quality and must therefore be taken into account.

According to RICHTER these aspects are marginal or only insufficiently considered in existing site planning models [5]. As a result, there is a clear need for research in the area of operational site planning in metropolitan regions [10]. To address this need for research, the German research foundation (DFG) is supporting the project "MetroPlant" from the Technical University Braunschweig and the Leibniz University Hannover.

2 RESEARCH OBJECTIVES

The aim of the research project "MetroPlant" is to create an integrated planning model for operational site planning and development in metropolitan regions. This should support the site decision process based on company requirements for regional location factors. In doing so, both the municipal and the company's possible measures for site development are to be included. The planning procedure therefore includes a multi-criteria evaluation and decision model (Figure 1).

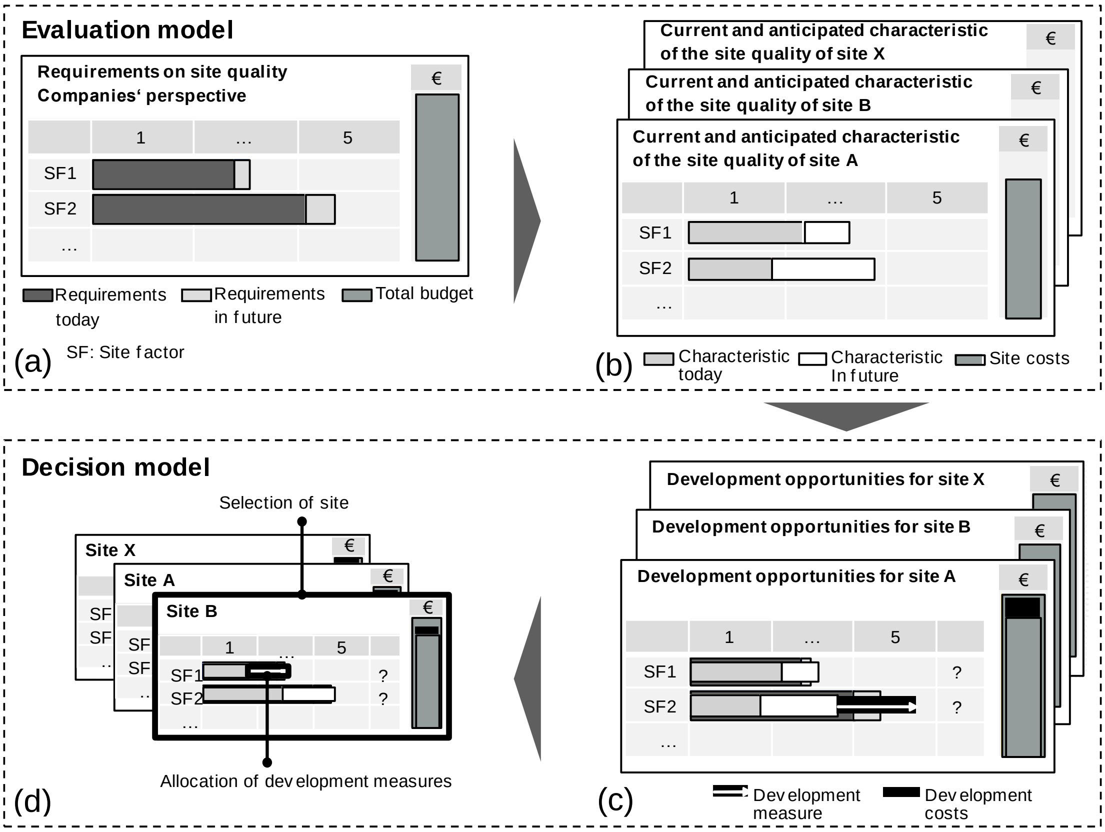


Figure 1 - Integrated planning model based on [11]

The evaluation model enables a target-is comparison of alternative sites on the basis of economically relevant regional location factors. In the first step, the current and future requirements for a company's location as well as the available total budget are recorded (a). In addition, the current as well as the forecasted site quality of each location is analyzed (b). A decision model is to be developed on the evaluation model. In the first step, the decision model is used to develop the future location characteristics of the individual locations. This is based on the allocation of corporate and municipal measures to influence the quality of the site. It should be noted that all company requirements are met as far as possible. The allocation of measures is based on a multi-criteria evaluation of the site factors. The site qualities and requirements as well as budget compliance are taken into account (c). In a second step, the best location for the company is selected from the original site candidates. On the other hand, the development measures to be planned at the site are also selected (d) [11].

A unique selling point of the planning process is the integrated allocation of development measures at the regional level of the decision. The specified budget, which is available for site-specific and measure-dependent development payments, is also considered. The approach also includes dependencies in the development of location factors. The associated planning model is very complex, which results from the possible combinations of locations and development measures [11].

Figure 2 shows the research questions to be answered in the research project in order to successfully develop the planning model.

Research question 1	Which dependencies exist between quantitative and qualitative regional location factors?
Research question 2	Which development measures exist for companies and municipalities to develop regional site factors and how do they affect the regional site factors?
Research question 3	How can alternative sites be evaluated on the basis of quantitative and qualitative regional site factors?
Research question 4	How can an approach for the selection of a site be designed taking into account uncertain municipal location developments as well as company-related development opportunities with integrated allocation of measures?
Research question 5	What potential results from the newly developed planning approach for site decisions?

Figure 2 - Research questions

The relevant regional site factors were defined in a preliminary study. First, the relevant dependencies between these site factors must be identified. These must be taken into account when assessing sites and site development measures (research question 1).

The dynamic planning model should include the development opportunities of municipalities and companies. This requires knowledge of the development measures of companies and municipalities and their effects on regional site factors (research question 2). The evaluation of a site by comparing the requirements of the company with the characteristics of regional location factors is a core element of the planning approach. In addition to the quantitative regional site factors, the qualitative factors must also be assessable (research question 3). The decision model selects an alternative site, based on the evaluation model. It also includes existing company requirements and restrictions. In doing so, both the uncertain community-driven developments and the development possibilities of company-related measures must be observed (research question 4).

The planning approach serves to evaluate and select company sites, with regard to municipal site developments and opportunities for influence on the company side. This shows additional degrees of freedom compared to previous methods, which leads to research question 5.

3 DETERMINATION OF THE EMPRICIAL BASIS

To answer the research questions, an empirical study conducted identifies possible development measures for improving regional site factors. In order to assess further effects initializing a development measure on other site factors, it is also necessary to include all interdependencies between the regional site factors. The aim of the empirical study therefore is on the one hand to identify the influence of regional site factors and on the other hand to determine suitable measures for the development of site qualities. In addition, synergetic and conflictual interdependencies between regional site factors are to be identified. The found development measures, including all identified interactions, have to be transferred to a generic catalogue of measures.

The empirical study is based on results from an interdisciplinary preliminary study on factory, location and network planning in metropolitan regions [12]. Economically relevant qualitative and quantitative regional site factors were identified in advance and aggregated into suitable clusters. The existing literature as well as a survey of companies and municipalities in a German metropolitan region were used. The clusters with the associated site factors can be seen in Figure 4 later in this paper.

The empirical study is based on the identified site factors and follows a two-stage survey concept. The survey is targeted at companies and municipalities in two German metropolitan regions. The first stage includes a quantitative online survey with a questionnaire. Here, the respondents assess their influence on the development of regional location factors and the associated interactions. The second stage serves to qualitatively supplement the

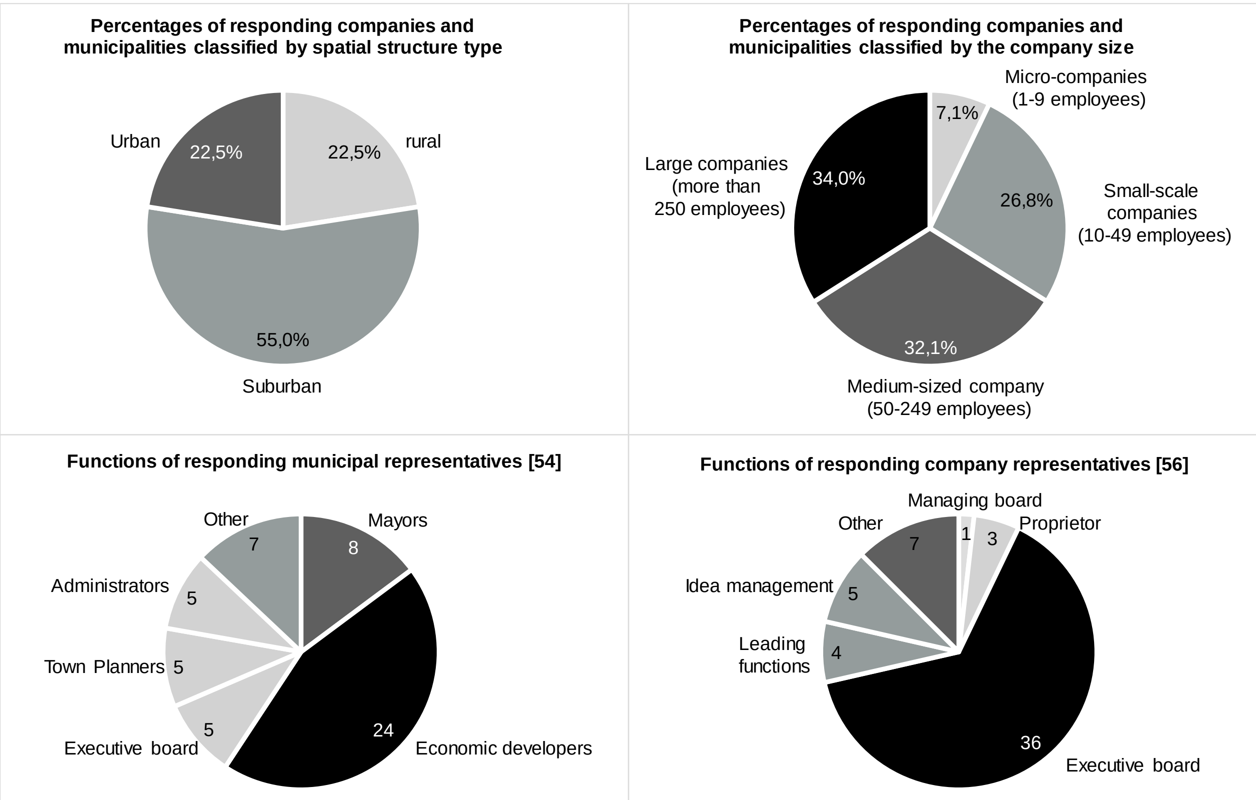


Figure 3 - Statistics of the respondents based on [11]

anonymous results through interviews with selected experts (Figure 3).

In the first stage, a representative sample of 110 completed questionnaires was obtained. This corresponds to a response rate of 21.4%. With a share of 50.9% representing companies and 49.1% consisting of municipalities, the sample of respondents is heterogeneous. Figure 3 illustrates the percentages of the company and municipality side according to spatial structure type, function and company size for the completed questionnaires. [11]

While 22.5% of companies and municipalities are located in urban and rural areas, the majority (55.0%) are located in suburban areas (a). Concerning the size of companies (b), large companies (34.0%) responded most frequently, followed by medium-sized (32.1%), small-scale (26.8%) and micro- companies (7.1%). Most of the municipalities and company representatives work in positions directly related to location planning. They also have experience and the necessary authority to make decisions in the context of site planning. Management boards and various management positions on the company side as well as economic developers and mayors on the municipal side, which underlines the relevance of the regional site problems, answered the questionnaires in particular (c and d). In addition to the questionnaires, seven expert interviews with selected company and municipal representatives were conducted in the second stage in order to supplement the results with qualitative statements. Special emphasis was placed on the selection of interview partners in heterogeneous types of spatial structures (rural, suburban, urban) and with heterogeneous company and municipalities sizes. The analysis of the expert interviews will be part of further publications and will not be discussed in this paper [11].

4 FIRST RESULTS OF THE EMPIRICAL STUDY

During the online survey, municipal and company representatives were able to assess their influence on regional location factors. Additionally, the regional location factors were compared so that existing interdependencies between the location factors could be evaluated. Initial results from both parts of the online questionnaire are presented in this section. The estimates regarding the influence of regional site factors by companies and municipalities were aggregated and compared using the six clusters in Figure 4.

With reference to KIK, in general, companies on average rate their influence on regional site factors lower than municipalities. While companies see their main influence in the cluster of employees, municipalities assess the site factors of municipal politics as highly influenceable. The evaluation of site factors in the cluster traffic infrastructure shows that municipalities and companies consider the connection to road and public local transport to be significantly easier to influence than the connection

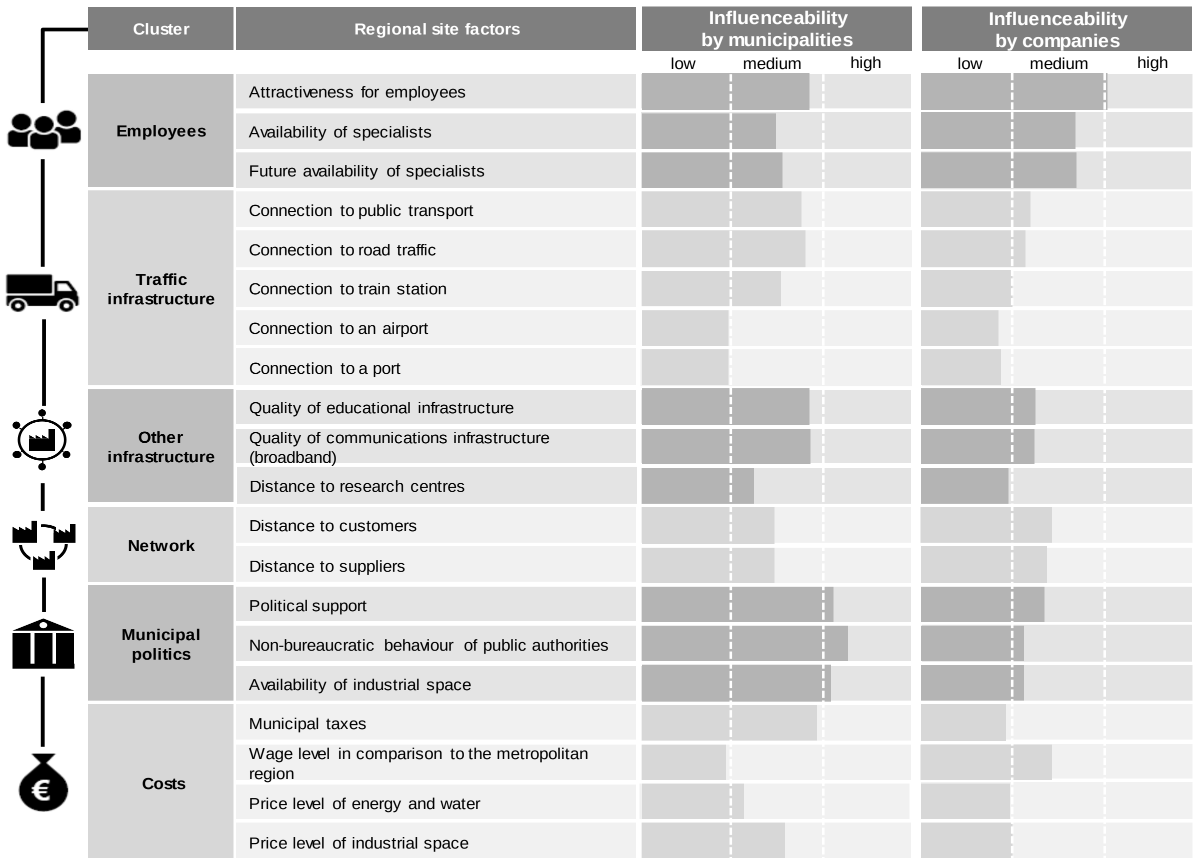


Figure 4 - Influenceability of site factors based on [5] and [11]

to the airport or harbor. In this context, municipal influence on the connection to road and public local transport as well as the railway station is greater than that of companies. The other infrastructure cluster shows that the quality of the educational and communications infrastructure can be influenced more strongly by companies and municipalities than the distance to research centers. The ability of companies and municipalities to influence the site factors in the cluster network is evaluated similarly in absolute terms. However, this influenceability belongs to the upper third of the ranking for companies and to the lower third for municipalities in comparison to other site factors. In the costs cluster, there are significant differences in the influence of wage levels in comparison to the metropolitan region and local taxes out of the perspective of companies and municipalities. Companies rate their impact on wage levels higher than municipalities, while municipal taxes are more strongly influenced by municipalities. Municipalities also point out a higher influence on the price level of trade and industrial sites as well as of energy and water. [11]

In the second part of the online questionnaire, company and municipal representatives had the opportunity to evaluate dependencies between regional site factors. The question was whether the improvement of one site factor had an impact on another site factor. Figure 5 shows the identified dependencies between the location factors at the cluster level.

The first evaluation of the interdependencies between the clusters of regional location factors shows that the employee is at the center of possible changes. Both the improvement of the traffic infrastructure and the other infrastructure have a decisive influence on the attractiveness for employees and the prospective and current availability of qualified employees. The respondent companies and municipalities also see strong dependencies within the employee cluster. Increasing the attractiveness for employees, for example, plays an important role in the availability of qualified employees. Inbetween the other clusters, only minor dependencies are discernible, which are also shown in Figure 5.

5 CONCLUSIONS

Continuing megatrends such as globalization and urbanization are leading to changes in spatial structures. In this context, the emergence of metropolitan regions can be observed, which have a high economic and social relevance and are characterized by different spatial structures. In order for companies to use the potential of existing spatial structures and develop locations according to their specific requirements, these aspects must be taken into account in the context of operational location planning. In addition, the availability of free space for new settlements is limited, so that the improvement of existing company locations is becoming increasingly important. In order to exploit these identified potentials, an integrated planning model for operational location planning and development in metropolitan regions is required.

This paper addresses the need for research for the first time by developing an integrated planning procedure for operational site planning and development in metropolitan regions. For this purpose, an empirical survey of companies and municipalities in two selected German metropolitan regions is presented. This allows the identification of the influence of regional site factors on the companies and municipalities, corresponding development measures and expected interdependencies. The results of the empirical study will be incorporated into the planning model to be developed in the future.

Future research activities of the "MetroPlant" project will focus on three areas. Initially the empirical results are to be reflected on the experiences of the discussion partners and evaluated with regard to their plausibility in interdisciplinary expert rounds with representatives from companies and

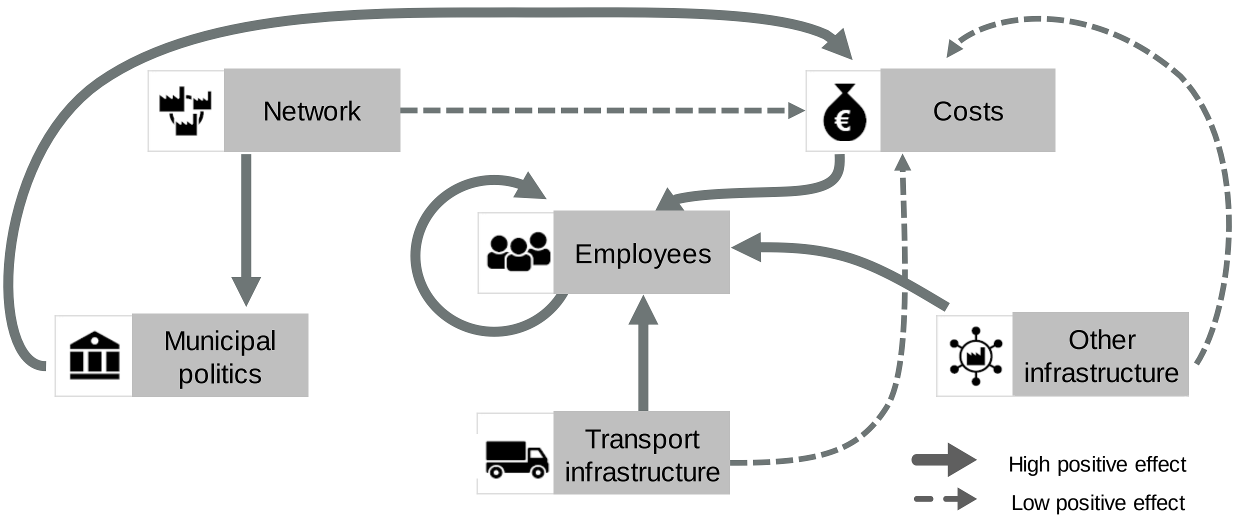


Figure 5 - Dependencies between regional site factors at cluster level

municipalities. In addition, a problem-adequate model of operational site planning and development needs to be developed. Finally, case studies and workshops with companies should ensure the transfer of the developed planning model into practice and serve to evaluate the application potentials in comparison to existing approaches. The overall aim of the planning process is to strengthen the decision-making process in municipalities and companies and to improve the future viability of and in metropolitan regions.

6 FUNDING

This work was supported by the German Research Foundation [grant number NY 4/60-1].

7 REFERENCES

[1] Blotevogel, H. H., 2002, Deutsche Metropolregionen in der Vernetzung, Bundesinstitut für Bau-, Stadt- und Raumforschung (BBSR), Bundesamt für Bauwesen und Raumordnung (BBR), 6/7.

[2] BMB Bau (Bundesministerium für Raumordnung, Bauwesen und Städtebau) 1995, Raumordnungspolitischer Handlungsrahmen – Beschluss der Ministerkonferenz für Raumordnung in Düsseldorf am 8. März 1995“.

[3] Adam, B., Göddecke- Stellmann, J., 2002, Metropolregionen – Konzepte, Definitionen und Herausforderungen, Bundesinstitut für Bau-, Stadt- und Raumforschung (BBSR) im Bundesamt für Bauwesen und Raumordnung (BBR) - Informationen zur Raumentwicklung, 9.

[4] Lutter, H., 2005, Raumordnungsbericht, Bundesamt für Bauwesen und Raumordnung (BBR) – Berichte, 21.

[5] Richter, L., 2017, Betriebliche Standortplanung auf regionaler Entscheidungsebene, Berichte aus dem IFA, 2, ISBN 978-3-95900-177-9.

[6] Domschke, W., Drexl, A., 1996 Logistik: Rundreisen und Touren, 3, ISBN: 9783486590937.

[7] Rüschenpöhler, H., 1958, Der Standort industrieller Unternehmungen als betriebswirtschaftliches Problem - Versuch einer betriebswirtschaftlichen Standortlehre, 6, ISBN 978-3-428-01266-4.

[8] Farahani, R. Z., Steadieseifi, M., Asgari, N., 2010, Multiple criteria facility location problems - A survey, Applied Mathematical Modelling, 34/7:1689 – 1709.

[9] Current, J., Min, H., Schilling, D., 1990, Multiobjective analysis of facility location decisions, European Journal of Operational Research, 49/3:295 – 307.

[10] Richter, L., Lübkemann, J., Schmidt, M., Nyhuis, P., 2014, Standort- und Netzwerkplanung in Metropolregionen – Aktueller Forschungsbedarf bei Entscheidungsmodellen, werkstatttechnik online 104/4:202 – 205.

[11] Kik, D., Krause, M., Wichmann, M., Spengler, T., Nyhuis, P., 2018, Betriebliche Standortplanung und -entwicklung in Metropolregionen, Zeitschrift für wirtschaftlichen Fabrikbetrieb, 7-8/113:503 – 507.

[12] Richter, L., Johannes, C., Schmidt, M., Nyhuis, P., Spengler, T., 2017, Standortplanung in Metropolregionen - Ergebnisse einer explorativen Unternehmens- und Kommunenbefragung in der Metropolregion Hannover-Braunschweig-Göttingen-Wolfsburg.

8 BIOGRAPHY

Marilena Krause M. Sc. studied mechanical engineering at Technical University Braunschweig with focus on production technology and systems engineering. Since 2018 she has been working as a research associate at the Institute of Production Systems and Logistics at Leibniz University Hannover within the research group Factory Planning.

David Kik M. Sc. studied Technology-oriented Management at Technical University Braunschweig with a focus on production and logistics management. Since 2018 he has been working as a research associate at the Institute of Automotive Management and Industrial Production at Technical University Braunschweig.

Prof. Dr.-Ing. habil. Peter Nyhuis studied mechanical engineering at the Leibniz University Hannover and afterwards worked as research associate at the Institute of Production Systems and Logistics. After receiving his doctorate in engineering, he also received his postdoctoral degree followed by his employment as manager in the field of supply chain management in the electronic and mechanical engineering sector. Since 2003, he is head of the Institute of Production Systems and Logistics (IFA) at Leibniz University of Hannover.

Prof. Dr. Thomas S. Spengler holds a diploma in Industrial Engineering and a PhD in Business Administration from the University of Karlsruhe (TH). He is Professor of Production and Logistics and head of the Institute of Automotive Management and Industrial Production at Technical University Braunschweig.

Holistic Product Optimization based on Design Space Exploration.

A. Zech[1], R. Stetter[1], M. Till[1]
[1]Hochschule Ravensburg-Weingarten, Weingarten, Germany.

Abstract
This paper is concerned with a systematic optimization framework for preparing and supporting design space exploration for products and manufacturing systems. In contrast to other publications which concern design space exploration this research literally deals with the design space, i.e. a certain volume within which a certain product, product component or manufacturing system can be realized. Such design spaces are very common in modular design, as remaining spaces my already be taken be predetermined modules of a modular technical system. Very often engineers need to develop product components with certain required characteristics and functions within a given volume. A specific support for this kind of product development task is not addressed in current design science. This paper seeks to explain a possible framework for a systematic analysis and representation of important aspects for this specific task in product development. A product example is used for explanation: an automotive bonnet (front hood) together with its manufacturing processes. The exemplary framework was realized using graph-based languages which employ the unified modelling language (UML) for a holistic digital representation of products as well as manufacturing and assembly processes and systems. Such languages were developed in recent years into a powerful tool for product development engineers and production planners; they offer several synergies and advantages for design space exploration. The digital generation of the bonnet starts with requirements (such as global stiffness or head injury criterion (HIC)) and functions (such as open/close). The available design space is defined by interface lines and areas. A reinforcement structure of the bonnet (consisting e.g. of struts and junctions) is automatically generated within this given design space using techniques based on coulomb repulsive forces and voronoi diagrams. The result is realized digitally using an interface to a commercial cad modelling system. The automated process allows holistic optimization processes.

Keywords
Integrated development, Optimization, Simulation

1 STATE OF THE ART

Kang et al. [1] define design space exploration (DSE) as follows: „design space exploration refers to the activity of exploring design alternatives prior to the implementation“. The basic idea of DSE is to replace a variant based procedure scheme by a design variable based procedure scheme [2], [3]. Practical experience in an industrial surrounding [4] as well as an investigation led by Tomiyama et al. [5] show, that expert teams can develop important system interrelationships in a variant based procedure scheme, but cannot arrive at with a complete system understanding. This fact underlines the necessity of DSE in product development processes.

When applying DSE, the existence of a quantitative system model is a prerequisite for an unequivocal evaluation of solution variants. Each changeable parameter of the system or a sub-component is referred to as design variable (DV) [4]. Several researcher report applications of DSE in the field of computer hardware [6], [7], [8], [9].

A DSE allows far reaching optimization techniques such as topology optimization. The procedures applied for identifying optimal product topologies range from heuristics with full-factorial experiment designs such as Pesce [10] over scaling procedures (e. g. Eghtessad [11], Felden et al. [12] to multi-criterion optimization with different algorithms. Moses [13] applies genetic algorithms for the optimization of electrical vehicles which are modelled using neuronal nets. Betram and Herzog [14] use a particle swarm approach. Goncalves et al. [15] describe the design of an unmanned aerial vehicle (UAV) and point out that design-space exploration techniques should be adopted for the component design steps. However, detail information about the design space exploration in the product concept phase is not given.

The design space in the publications mentioned above is spanned by two or more design variables and can be multidimensional. Each possible product topology is a point in this space. This understanding of DSE could be referred to as general design space exploration and is shown in Figure 1 on top for 3 dimensions (design variables). In contrast to this understanding, the research in this paper literally deals with the design space, i.e. a certain volume within which a certain product or product component can be realized. This understanding of DSE could be referred to as spatial design space exploration in mechanical engineering and is shown in Figure 1 - lower part.

The importance of this kind of DSE is underlined by the fact that engineers frequently need to develop product components with certain required characteristics and function within a given volume. Such design spaces are very common in modular design, as remaining spaces may already be taken by predetermined modules of a modular product.

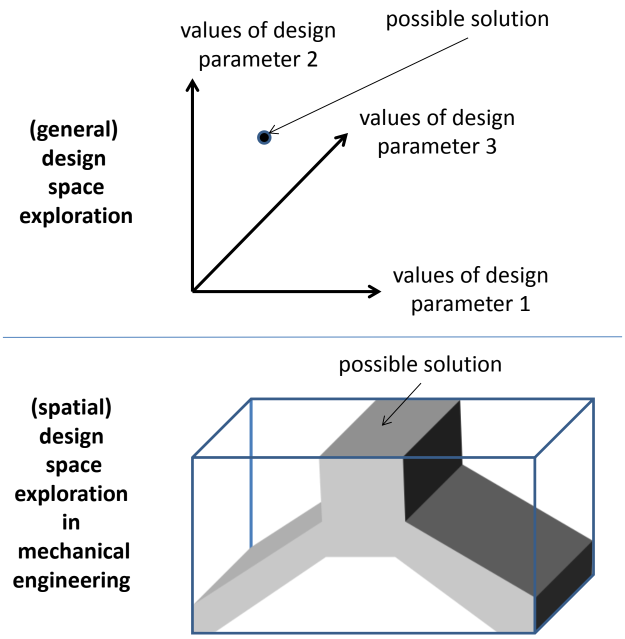


Figure 1 - General understanding of design space exploration

2 FRAMEWORK FOR THE DESCRIPTION OF DESIGN SPACES

In theory, design space exploration is also possible for complete products. However, in this case the number of degrees of freedom and the overall complexity is very high. Therefore, it is usually sensible to apply it to modules of a modular product. It is usually sensible, as well, to determine the general functionality, the general spatial arrangement and predetermined modules of the product up front using conventional product development methods. Following this stage a design space exploration stage using graph-based design languages may open the path towards optimized products. Once a general modular structure is determined, certain entities can be defined which have to fulfil a certain task or mission and for which a certain design space can be allocated.

2.1 Entity

In this context an entity is understand as a structure which is located within an allocated design space and has to fulfil a certain mission. Examples for entities can be brackets holding a certain load or an electrical drive system. In this paper the entities used for explaining the presented approach are an automotive bonnet. The task which an entity has to fulfil can be described in a mission statement.

2.2 Mission statement

In the modular structure of a product, each module or entity has to fulfil certain functionalities and has to respect certain requirements. A conscious clarification of this mission allows designers an abstract view of their module and decreases the risk to forget important aspects of the module. Additionally, this **mission** contains the criteria for later **optimization** cycles. The mission statement of an entity should therefore contain a description of the required functionalities together with certain requirements which may concern certain required characteristics of the module such as “with corrosion protection”. Figure 2 explains the notions “design space”, “entity” and “mission statement”.

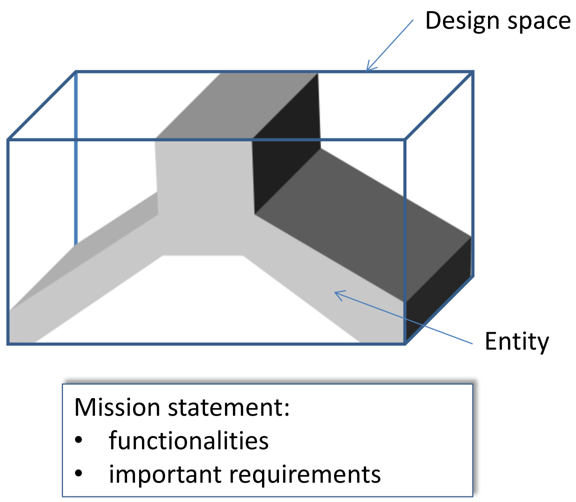


Figure 2 - Important notions in design space exploration

2.3 Design Space

While determining the general spatial arrangement of a product, certain spaces can be allocated for certain modules. This space can be considered to be the possible design space for a certain entity. Rather often relatively simple spaces can be assumed, e. g. boxes or cylinders. For complex products more complicated geometrical forms can be allocated as design space, for instance when the space requirements of adjoining modules are already predetermined at the stage of design exploration for a certain entity. Usually it is sensible to describe these more complicated spaces with spatial CAD data such as step files.

2.4 Interfaces

Generally everything getting in or coming out of such spaces can on an abstract level be described with matter, energy or signal (using the entities described in flow oriented function structures compare Pahl et al. [16], Ehrlenspiel and Meerkamm [17], Eisenbart et al. [18], Ramsaier et al. [19]. The geometrical position of such interfaces can be already determined by some structures outside the design space – in this case the notion “interface point” is sensible. Sometimes it might be possible to change the location of the interface in a certain region of the surface of the design space – this can be indicated using the notion “interface area”.

In extreme cases, when no surrounding structures were defined earlier the interface area for certain flows may even be the entire surface of the design space. In this case the optimization algorithms can chose any position for the interface – it only needs to be present. Very often the general dimension of the interface will be known – e. g. for certain electrical connectors or can easily be determined by calculating the amount of flow needed and the possible and sensible velocity of the flow. Figure 3 shows different interfaces.

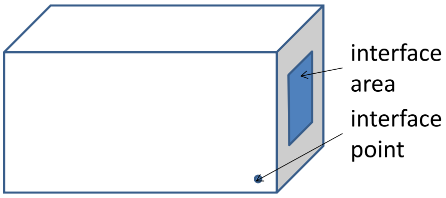


Figure 3 - Sample illustration of design space interfaces

2.5 Loads

For design space exploration it was however found to be sensible to add another category which might already be represented in the energy flow, but is so prominent for the optimization that it may be addressed separately – the load. A load can be coming from the outside, e. g. some weight resting on the future entity in the design space or may go to the outside, e. g. the weight of the future entity itself which needs to rest on something else outside the design space.

Similar to the general interfaces, the position of the load transmission can be already determined by some structures outside the design space – in this case the notion “load point” is sensible. In certain situation it might be possible to move the load transmission on a straight or curved line – in this case the notion “load line” can be used. Furthermore, it might be possible to move the load transmission in a certain region of the surface of the design space – this can be indicated using the notion “load area”. Figure 4 explains the different kinds of loads.

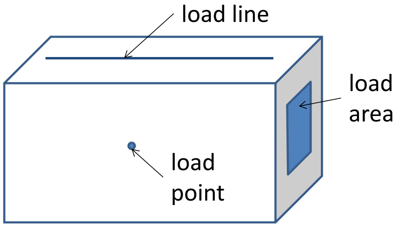


Figure 4 - Sample illustration of design space loads

This kind of loads can be described using stress tensors which contain normal and shear stresses. Sometimes it is also sensible to describe the loads using force and moment vectors in order to allow an optimization of the size of the load transmission area. It is usually sensible to distinguish static and dynamic loads and to describe how frequent the load will be acting on the interface point, line or area.

2.6 Required Elements

Very often it is clear upfront that certain elements are necessary for the mission of the future entity in the design space. It may be necessary that the entity includes a battery of certain dimensions. For required elements the following cases can be distinguished:

- Predetermined elements: it is exactly know upfront which element has to be used.
- Selection from an element family: elements can be chosen from a selection of possible elements, e. g. a modular system of such sub-elements.
- Parametrical elements: the dimensions of necessary elements can be changed in order to optimize the usage of the design space, but usually certain restrictions have to be obeyed (e. g. battery capacity often corresponds to space requirements).

3 PRODUCT EXAMPLE: AUTOMOTIVE BONNET

The presented research results are parts of the research project DiP (Digital Product Life Cycle) which attempts to represent the entire product life cycle of industrial products in a fully digital manner in order to make product development processes more effective and efficient. As one of the sample products serves an engine hood (bonnet) for automobiles (compare [20]). Bonnets are mounted in automotive car bodies and are used to cover engine components. In the project graph-based design languages are used.

3.1 Graph-Based Design Languages

An innovative approach for product development is to use a common language to describe the product lifecycle by means of the Unified Modeling Language (UML). Through vocabularies (UML classes) and rules (UML activities) a design language can be generated which leads to an inherent consistency. Rudolph [21] firstly proposed to use graph-based design languages for engineering design.

One important advantage in the underlying project is the availability of the “DesignCockpit43©”. This compiler is developed by the IILS mbH (http://www.iils.de) in cooperation with the University of Stuttgart, which is one of the cooperation partners in this project. DesignCockpit43 (DC43) is an eclipse-based software tool for the compilation of design languages formulated in UML.

3.2 Geometry Synthesis

Initially a variety of different topological structures for automotive bonnets could be identified. The different designs are mainly driven by strength and stiffness requirements. In addition to the consideration of real bonnet structures, different surface and solid modeling techniques were examined. A procedure for the creation of robust surfaces was created. This procedure can start with rather simple features which may considered as design parameters, which consists of points, lines, splines, etc. The topologies, positions, interfaces and the surfaces are modelled by using simple features. Figure 5 shows the development process of any shapes respectively surfaces of entities by using design parameter in an abstract way.

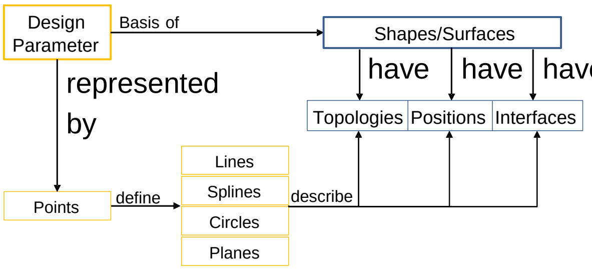


Figure 5 - Design methodology for entities (shapes)

This design procedure is based on design parameters and one variant of the general solid model representation – the Boundary Representation (BRep). The BRep approach forms the basis for the design methodology which will be implemented in DC43. Figure 7 contains an illustration of the individual geometry modeling steps of the synthesis process.

While the design space as well as the initial design parameters (e. g. strut profile) will be manually specified at the beginning of the design process, all required and further design parameter and the resulting surfaces become generated in a fully automatic manner.

The engineering task starts with a package model of the bonnet, which is derived from the concept package model of the car. From this package model the design space can be derived. All functional elements and manufacturing requirements have to be considered and are all placed in the total design space of the whole assembly. As can be seen in Figure 6, all geometrical information is abstractly described using design parameters (simple elements) such as points, lines, etc. By using mathematical logics and algorithms all necessary further design parameters are calculated automatically. That leads to a skeleton model, which contains all positions, geometrical interfaces and geometrical information of each functional element and manufacturing requirement. Based on the skeleton model, the individual surfaces may be created and transformed to a volume model.

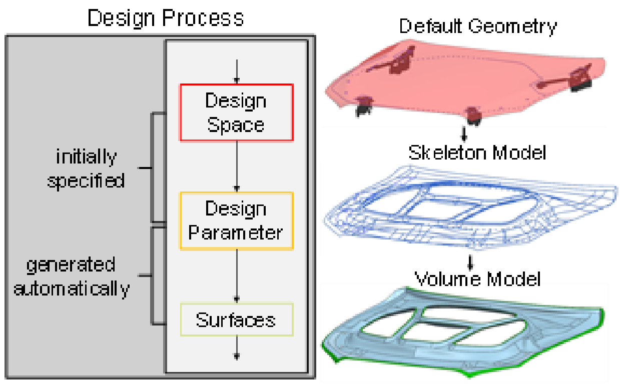


Figure 6 - Modelling Steps

In this project the design process is implemented in DC43. Therefore the product architecture of the bonnet has been modelled in UML. Classes represent the product components such as “Inner Panel” or “Hood Look” (compare Figure 7).

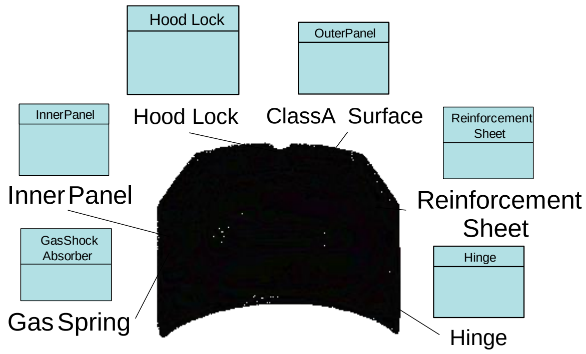


Figure 7 - UML Classes as Representatives of the Product Components

In the next step the classes are instantiated to specific objects of product components by the design compiler – this step leads to the design graph (Figure 8).

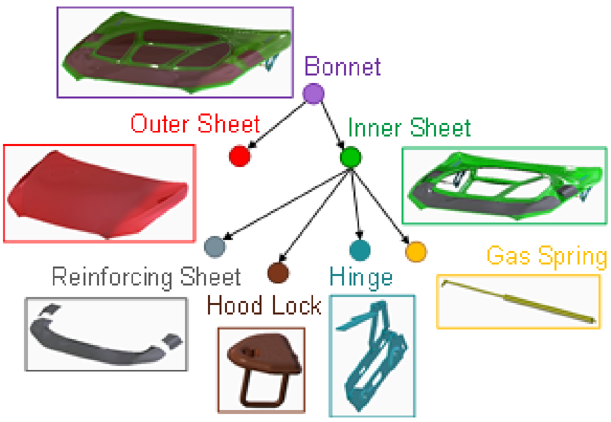


Figure 8 - Design Graph of one specific design variant of bonnet and its geometrical expression in CAD system

This design graph represents the central data model of one specific design variant (one specific automoitve bonnet). Through model to model transformation the design graph may be geometrically expressed in domain specific CAD systems or formats. This graph can also contain information from other phases of the product life cycle (e. g. assembly information) or other domains (e. g. physics).

3.3 Generation of Reinforcement Variations

One important part of a bonnet assembly is the reinforcment structure (inner sheet, ...) which is located below the outer surface (class A design surface). While the outer surface is mainly driven by design, the reinforcement structure is mainly driven by functional considerations. One very important aspect is the pedestrian protection. Today, the Head Injury Criterion (HIC) is used in order to asses the danger of injury if the head of a pedestrian would hit the bonnet in the case of an accident. However, engineers need to consider many other aspects. For instance, the global stiffness requirements are in direct conflict with the need for compliance in the event of a head impact accident.

The automatic generation of the reinforcement geometry within the given design space can be based on two totally different topological variants: on a system of struts and on a sheet with a number of round holes. The engineer can choose certain parameters of the topological variants and the system generates possible solutions (Figure 9).

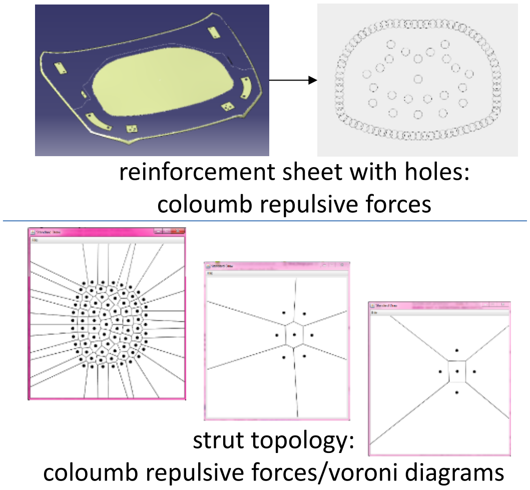


Figure 9 - Determination of geometry within the design space

For the sheets with a number of round holes the coloumb repulsive forces can be applied, while for strut topologies additionally voronoi diagrams are used (Figure 11).

3.4 Holistic Optimzation of Reinforcement Variations

In the developed system an automatic evaluation of the HIC values for a large number of points of the bonnet could be realized. This results can be used in order to optimize the geometry of the bonnet. One example may be the possibility of changing the topological structure of the inner panel in order to influence local stiffness jumps and the resulting high local HIC-values. This effectively reduces HIC levels and minimizes the risk of injury for pedestrians. The optimization starts with an identification of gird points with high HIC-values. In the optimization, the structure of the underlying geometry in the given design space is adapted in order to achieve more compliant behavior at this point of the structure. As a result the structure meets the head impact requirement.

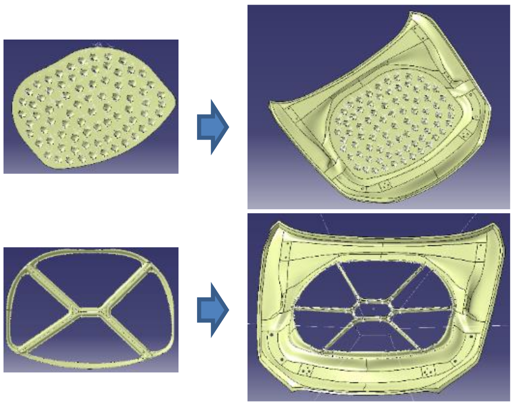


Figure 10 - Possible reinforcement variants

4 CONCLUSIONS

This paper explained possibilities for design space exploration using a systematic optimization framework based on graph-based design languages. In contrast to other research works, the notion “design space” is literally a space within which product components can be realized. Very often, engineers need to realize a certain functionality – a **mission** – within a given design space and need to find the best possible solution with regards to several requirements – they need to perform a holistic **optimization**. On the basis of an automotive engine hood, the presented research was able to demonstrate, that graph-based design languages using UML can allow a executable digital process which leads from a mission over the investigation of the design space to an optimized solution with regard to several requirements, for example safety requirements (e.g. Head Injury criterion HIC).

5 ACKNOWLEDGMENTS

The project „digital product life-cycle (ZaFH)“ is supported by a grant from the European Regional Development Fund and the Ministry of Science, Research and the Arts of Baden-Württemberg, Germany (information under: www.rwb-efre.baden-wuerttemberg.de).

6 REFERENCES

[1] Kang E., Jackson E. and Schulte W.: An Approach for Effective Design Space Exploration. In Calinescu, R. and Jackson, E. (Eds.): Foundations of computer software, pp 33–54, Springer, Berlin, 2011.

[2] Haveman S.P. and Bonnema G.M.: Requirements for High-Level Models Supporting

Design Space Exploration in Model-based Systems Engineering, Procedia Computer Science, Band 16:293–302, 2013.

[3] Palm H., Holzmann J., Gerling D., Schneider S.A., Kögeler H.M. und Pfister, Felix, Schick, Bernhard: Die Zukunft im Fahrzeugentwurf: Systems Engineering basierte Optimierung, in VPC - Virtual Powertrain Creation, Springer Fachmedien, Mainz, 2013.

[4] Vaillant, Moritz: Design Space Exploration zur multikriteriellen Optimierung elektrischer Sportwagenantriebsstränge. Variation von Topologie und Komponenteneigenschaften zur Steigerung von Fahrleistungen und Tank-to-Wheel Wirkungsgrad. Dissertation, Karlsruher Institut für Technologie (KIT), 2015.

[5] Tomiyama T., D'Amelio V., Urbanic J. und Eimaraghy W.: Complexity of Multi-Disciplinary Design, CIRP Annals - Manufacturing Technology, Band 56(1):185–188, 2007.

[6] Ozisikyilmaz B.; Memik, G.; Choudhary, A.: Efficient System Design Space Exploration Using Machine Learning Techniques. In: Proceeding DAC '08 Proceedings of the 45th annual Design Automation Conference, pp 966-969, 2008.

[7] Ipek E., McKee S. A., deSupinski B. R., Schultz M. and Caruana R. Efficiently Exploring Architectural Design Spaces via Predictive Modeling. In Proc. of the ASPLOS, Oct. 2006, San Jose, CA.

[8] Lee B. C. and Brooks D. M. Accurate and Efficient Regression Modeling for Microarchitectural Performance and Power Prediction. In Proc. of the ASPLOS, 2006.

[9] Eyerman S., Eeckhout L. and Bosschere K. D. The Shape of the Processor Design Space and its Implications for Early Stage Explorations. In Proc. of Int. Conf. on ACMOS. 2005, Prague.

[10] Pesce T.: Ein Werkzeug zur Spezifikation von effizienten Antriebstopologien für Elektrofahrzeuge: Dissertation, Fahrzeugtechnik, Verlag Dr. Hut, München, 2014.

[11] Eghtessad M.: Optimale Antriebsstrangkonfigurationen für Elektrofahrzeuge: Dissertation, Band 35 von Schriftenreihe des Instituts für Fahrzeugtechnik, TU Braunschweig, Shaker, Aachen, ISBN 978-3-8440-2782-2, 2014.

[12] Felden M., Butterling P., Jeck P., Eckstein L. and Hameyer K.: Electric vehicle drive trains: From the specification sheet to the drive-train concept, in 2010 14th International Power Electronics and Motion Control Conference (EPE/PEMC 2010), pp. 9–16.

[13] Moses S.: Optimierungsstrategien für die Auslegung und Bewertung energieoptimaler Fahrzeugkonzepte: Dissertation, Band 62 von AutoUni - Schriftenreihe, Logos Berlin, Berlin, ISBN 978-3-8325-3718-0, 2014.

[14] Bertram C. and Herzog H.G.: Optimierung der Antriebsstrangtopologie von Elektrofahrzeugen, in Internationaler ETG Kongress 2013, Band 139 von ETG-Fachbericht, VDE-Verlag, Berlin, ISBN 978-3-8007-3550-1, 2013.

[15] Goncalves, F.S., Raffo, G.V., Becker, L. B.: Managing CPS Complexity: Design Method for Unmanned Aerial Vehicles. IFAC PapersOnLine, Volume 49, Issue 32, 2006, pp 141-146.

[16] Pahl, G.; Beitz, W.; Feldhusen, J.; Grote K.H.: Engineering design: a systematic approach. Springer, 2007.

[17] Ehrlenspiel, K.; Meerkamm, H.: Integrierte Produktentwicklung. Denkabläufe, Methodeneinsatz, Zusammenarbeit. Hanser: 2013.

[18] Eisenbart, B., Gericke, K., Blessing, T. M., McAloone, T.C.: A DSM-based framework for integrated function modelling: concept, application and evaluation. In: Research in Engineering Design (2016).

[19] Ramsaier, M.; Holder, K.; Stetter, R.; Rudolph, S.; Till, M.: Digital representation of product functions in multicopter design. In: Proceedings of ICED17, 2017.

[20] Zech, A.; Stetter, R.; Holder, K.; Rudolph, S.; Till, M.: Novel approach for a holistic and completely digital represented product development process by using graph-based design languages. Proceedings of the 12th CIRP ICME, 18-20 July 2018, Italy.

[21] Rudolph, S. (2002), Übertragung von Ähnlichkeitsbegriffen., Habilitationsschrift, Universität Stuttgart, Stuttgart.

7 BIOGRAPHY

Andreas Zech obtained his M.Sc. degree in Mechanical Engineering at the University of Erlangen-Nuremberg. He is currently academic employee at the University of Applied Sciences Ravensburg-Weingarten.

Ralf Stetter obtained his PhD degree in Mechanical Engineering from the Technical University of Munich. In 2004 he was appointed Professor in Design and Development at the University of Applied Sciences Ravensburg-Weingarten.

Markus Till obtained his PhD degree in Aeronautical Engineering from the University of Stuttgart. He is currently Professor in Simulation at the University of Applied Sciences Ravensburg-Weingarten.

Integrated Optimization based on Balanced Scorecard and Graph-based Languages

R. Stetter[1], J. Kiefer[2], M. Witczak[3], M. Till[1]
[1]Hochschule Ravensburg-Weingarten, Weingarten, Germany.
[2]Hochschule Ulm, Ulm, Germany
[3]University of Zielona Góra

Abstract
In recent years graph-based languages were developed into a framework which allows the digital development of products, manufacturing systems and assembly systems and the holistic optimization of the product life cycle. Currently, the connection between the company's strategy and these processes as well as the digital realization of this connection were not in the focus of research. One prominent tool for supporting the strategic development of a company is the balanced scorecard, which is essentially a performance management tool which combines financial and non-financial measures such as employee satisfaction. Obviously, the innovations, products and production systems and process can be decisive for the success of a company and therefore require an inclusion in a concise strategy. However, this inclusion and especially the continuous tracking of the company's progress in these aspects could frequently not be realized in conventional processes. Graph-based languages allow the digital representation of many aspects of innovations, products and production systems and processes in a holistic framework. Furthermore, they foster optimization cycles which extend beyond the product development phase into manufacturing, assembly and usage; several aspects of the company are influenced by these products, components and processes. Additionally, graph-based languages allow not only parametrical detail design optimization but also topological concept optimization; such optimization process require weighted objectives which also need to be derived from the strategic objectives of the company. This kind of strategies, for instance concerning the quality objectives, can be developed on the basis of a balanced scorecard process and can be divided into operative goals for processes and products. Additionally, the balanced scorecard combined with techniques such as aggregation, triangulation, consensus building, expert panels and Delphi method can be sensibly applied for assessing the innovation performance of a company. An integration framework of the balanced scorecard with graph-based languages is the focus of the paper.

Keywords
Integrated development, Optimization, Balanced Scorecard

1 INTRODUCTION

The main contribution of this paper is a framework for a combination of the balanced scorecard with graph-based languages for the sake of continuous digital process and integrated optimization. Large studies underline the importance of continuous digital processes: "the goal must be that a holistic cross-disciplinary design of a complex system is possible in the course of further concretization in the established development methods and the corresponding tool environments of the affected domains such as mechanic, electrical engineering, software engineering and plant and process technology" [1]. Graph-based languages are promising tools to address this issue and allow the synthesis of products and technical processes [2]. In 2013, the Balanced Scorecard (BSC) was the most widely used management tool in Germany and Europe, worldwide it was ranked 5th [3]. Industrial examples show that the BSC can develop an integrating and coordinating effect in in the field of control of innovation areas and can bridge the gap between strategy and operational business [4]. Consequently, the framework explained in this paper is aiming at a continuous digital combination of these two approaches. For this explanation, the next section describes graph-based languages. The application of the balanced scorecard in innovation processes is explained in section 3. The framework for the combination is presented on this basis in section 4.

2 GRAPH-BASED LANGUAGES

Graph-based languages are a novel way to create and model product designs in a holistic digital manner. In such languages, the individual terms (i.e., the "vocabulary") represent reusable and relatively freely (re) combinable language building blocks. The building knowledge about the individual vocabulary (i.e., the "rules") are mapped as language operations. Graph-based, in the context of graph-based languages, means that in the digital representation of the vocabulary of a design language, the nodes of a graph serve as abstract placeholders (i.e. models) for real objects, processes, or states, and thus represent the product design. With the graph-based design languages, a very powerful and digital engineering framework was generated based on the work of Rudolph [5], which enables the automated processing and reuse of design and manufacturing knowledge [6].

The digital data model of software-derived Unified Modeling Language (UML), which is used to represent graph-based languages, enables to store the entire design knowledge required for product and production development throughout the product lifecycle. This knowledge includes all relevant product requirements, design parameters and interfaces in the development and production of selected components. With UML it can be represented completely independent of a specific, program-specific or proprietary data formats in a free, open source and internationally standardized language standard. Only in the last step is the holistic data model in UML transformed into domain-specific language representations (DSLs). This achieves a clear separation between knowledge representation in the form of an object-oriented language in UML and possible further processing in the diverse process chains and data formats of the various manufacturers [7]. The basic structure is shown in Figure 1.

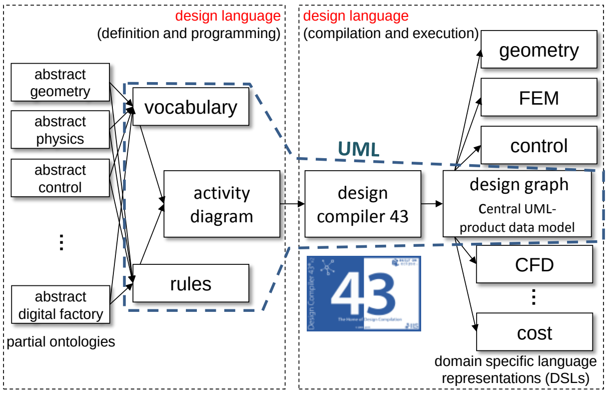


Figure 1 - Basic structure: graph-based languages with UML

Because of this deliberately chosen abstract representation of knowledge, abstractly formulated vocabulary and rules must naturally be found, written down, and processed in a language compiler for all terms and their interactions occurring along the product life cycle. The required abstract representation of a concrete CAD product geometry in the form of a so-called "abstract geometry" (in UML) and for further multi-physical product properties (e.g. in the form of MKS, FEM and CFD models for multi-body, structural-mechanical and fluid-mechanical simulation) in the form of a so-called "abstract physics" (also in UML) in graph-based design languages allows the information representation in UML independent of a specific data formats in a specific domain or from a specific proprietary or proprietary data format. The final transformation from UML into one or more target languages (Domain-Specific Languages, DSLs) then takes place and through another machine processing step in the language compiler or its plug-ins.

The concrete application of a graph-based design language by means of UML diagrams is portrayed in Figure 2.

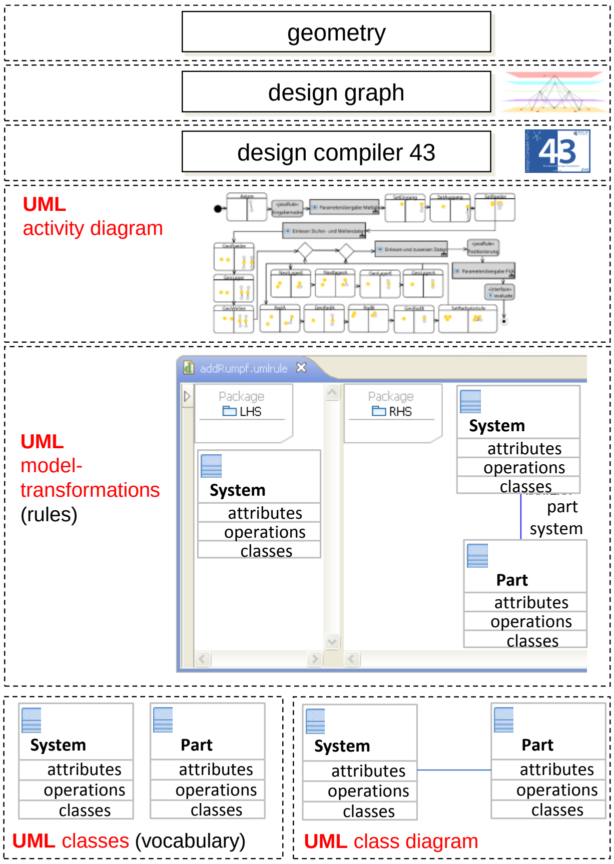


Figure 2 - Basic structure: graph-based languages with UML

The process flow in Figure 2 is from the bottom to the top. In this graph-based language, the engineering objects (e.g., product components, processes) are represented by UML classes (vocabulary of a design language). One example of such a UML class could be systems such as a car with its attributes, operations and sub-classes (components). The various interconnections of the engineering objects (spatial, physical, logical, etc.) as well as the hierarchical structure of the overall system are represented by the UML class diagram. One exemplary interconnection is that a system consists of different parts. The model transformations are represented by rules. It can be seen in Figure 2 that these rules consist of a left-hand side (LHS) containing the structure before the instantiation of the rule and a right-hand side (RHS) describing the structure after the instantiation process (result after considering concrete requirements, e.g. the maximum fuel consumption of the car). In an activity diagram, the sequence of rules is specified by the project engineer. By executing this sequence, the individual vocabulary (UML classes) and their interconnections are instantiated. The UML class diagram can be seen as a blueprint of all possible products within the specific design language.

With the methodology of graph-based languages in UML, a consistent, holistic and very powerful IT modelling of the entire product lifecycle is achieved. Design automation achievable through the rule-based presentation of design and manufacturing knowledge enables an automatic generation of a large number of new or alternative product designs along with their virtual manufacturing simulation. The projection of the features of the topologically and parametrically different product variants into specially selected subspaces allows a greatly improved understanding of the technical and economic interactions (in the sense of the knowledge of hidden system couplings) between the individual components, subsystems or sub-disciplines.

3 BALANCED SCORECARD

This section describes the fundamentals of the balanced scorecard, the application in the area of product and production innovation and the development of strategic innovation objectives and related key figures.

3.1 Fundamentals

The balanced scorecard (BSC) was developed by Robert Kaplan and David Norton in the 1990s and assists companies to clarify their strategy and to communicate the top strategic objectives and priorities. The basic concept of BSC therefore complements the traditional financial key figures through (i) a customer, (ii) an internal process and (iii) a learning and development perspective (Figure 3).

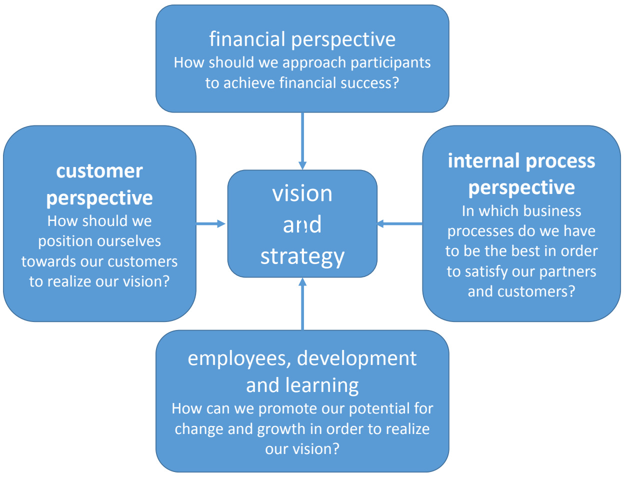


Figure 3 - Four perspectives in the balanced scorecard

In a BSC leading indicators are combined with performance key figures [4], [6], [8]. The addition of the three other perspectives to the financial perspective means that the traditional key performance indicators are extended by leading indicators, such as throughput times or quality indicators. These performance drivers reflect the actual competitive advantages of the company. Every single key figure, should be part of a cause-and-effect chain that ends in a financial goal, which in turn reflects the company's strategy. The BSC it is not a recent collection of isolated measures; it rather specifies how improvements are made in operational areas, to affect financial performance, e. g. by higher sales, or lower costs [4].

3.2 Balanced Scorecard for Innovations

In recent year several approach to apply the balanced scorecard in the area of innovation management have been reported. Vinkemeier [4] describes the connection of a so-called “balanced innovation card” (BIC) with a technology roadmapping. He underlines that the BSC can have a positive effect in key area for the long term success of a company, such as R & D and Innovation. Beeck [9] reports about applications of the balanced innovation card in medium sized enterprises of the automotive supplier industry. Nestle [10] uses the term “innovation balanced scorecard” and presents the interactions between balanced scorecard and innovation balanced scorecard (Figure 4).

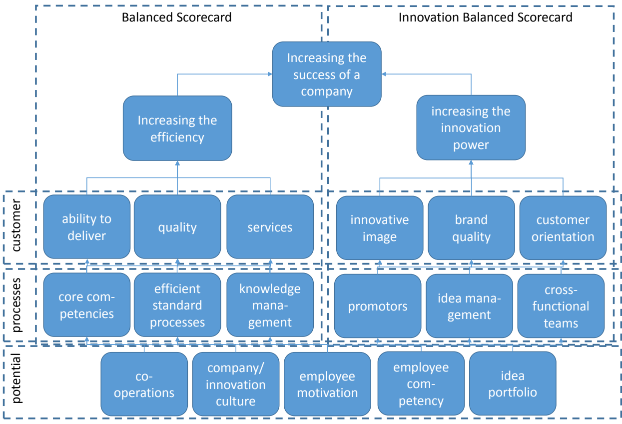


Figure 4 - Interactions between balanced scorecard and innovation balanced scorecard (compare [10])

3.3 Development of strategic innovation objectives and key figures

Beeck [9] distinguishes certain perspectives on the innovative power of a company, i. e. all aspects which influence the innovation management. The three aspects innovativeness (with a cultural and a resource oriented perspective), innovative activities (through the process perspective) and innovative success (through the output perspective) represent the innovative power in a holistic way (Figure 5).

This paper will focus on the two later aspects innovative activities and innovative success, as the cultural and resource oriented perspectives are beyond the scope of the research.

For the aspect innovative activities, Beeck [9] proposes four strategic innovation objectives: “early customer orientation”, “reduction of innovation duration", "optimization of the market entry timing" and "optimization of the evaluation and selection processes". Each of the strategic innovation objective is assigned possible key figures.

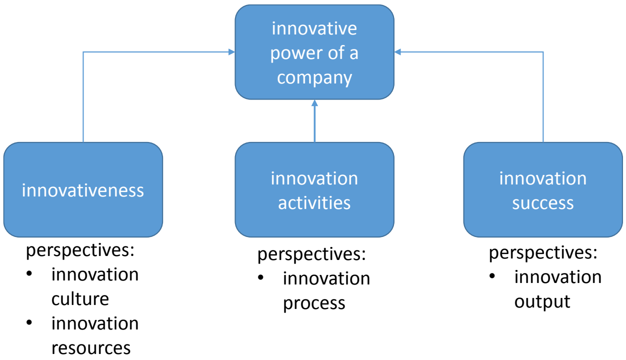


Figure 5 - Perspectives on the innovative power (compare [9])

"Early customer orientation" can be a crucial factor for supplier companies; a possible key figure can be formulated as "the ratio between number of customer contacts before concept development and number of customer contacts before market entry". Also a "reduction of innovation duration" can be very important, because only companies that are able to timely to have finished the innovation processes, can exist on the market [11]; a possible key figure is consequently the "average project duration". Closely related to this is the objective of optimizing the "market entry timing". In supplier industry, the compliance with costumer's market entry strategies is for customer satisfaction of elementary importance. A formulated key figure could be "the ratio of the number of timely finished work packages and the number of work packages before market entry". An optimal use of resources requires at different times decisions about the continuation or termination of innovation projects. This necessitates an "optimization of the evaluation and selection processes". A possible key figure could be "the ratio between the number of terminated ideas before concept realization and the number of terminated ideas before market entry".

For the aspect innovative success, it is sensible to propose five strategic innovation objectives: "enhancement of customer satisfaction", "enhancement of production processes", "enhancement of innovation efficiency", "enhancement of innovation rate" and "enhancement of innovation success" (compare [9]).

An "enhancement of customer satisfaction" can lead to a strategic advantage of a company; a key figure can result from customer surveys and can be labelled "average customer satisfaction". Additionally, due to the enormous importance it can be sensible to distinguish the customer satisfaction into "functional customer satisfaction", "efficiency user satisfaction" "design customer satisfaction", "quality user satisfaction". Many innovations due not change the products of a company but allow the company to produce their products more efficiently with lower cost; this can be captured in the strategic objective "enhancement of production processes"; a possible key figure can be "reduction of production cost". Measuring the achievement of the objective "enhancement of innovation efficiency" allows to estimate whether innovations actually contribute to economic success; a key figure can be formulated as "ratio between present value of revenues with new product in year n and present value of payments for innovation projects in year n". The objective "enhancement of innovation rate" also focuses on a successful strategy implementation; a key figure may be: "the ratio between profit with new products in year n and overall profit in year n". An "enhancement of innovation success" is only possible if the four objectives of perspective innovation process were achieved. While the other three objectives have a direct connection to business success, this objective refers exclusively on the quantitative number of successful innovations; a key figure can be "ratio between number of successful product innovations and overall number of product innovations". The resulting concept is shown in Figure 6.

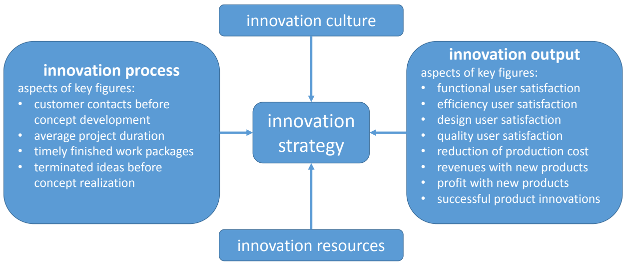


Figure 6 - Concept with aspects of key figures (compare [9])

4 FRAMEWORK FOR THE COMBINATION OF BALANCED SCORECARD AND GRAPH-BASED LANGUAGES

This section seeks to investigate the links between this innovation perspective and digital development processes on the basis of a simplified model of a product and production development process (Figure 7 – compare [12]).

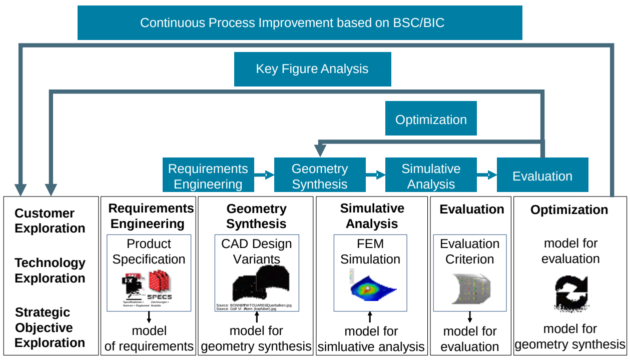


Figure 7 - Model of a product and production development process with BSC/BIC aspects – framework for the combination

The main part of this process are the core elements of digital engineering: requirements engineering, geometry synthesis, simulative analysis, evaluation and optimization. The specific elements for integrating the strategic considerations of the BSC can be found on the left and upper side; this will be explained in the remaining part of this section.

Central key figures of the innovation output (compare Section 3) concern the customer satisfaction. In product and production development processes requirements are an absolutely central element [13]; however today the connection between the fulfilment of certain requirements and an enhanced customer satisfaction is frequently unclear. Consequently, it is a key element between strategy considerations and operative digital processes that the link between user satisfaction and requirements is closed. A considerable body of research concerns the integration of the voice of the customer into the requirement models of digital processes. For this integration a number of models, methods and tools were analysed such as the Kano model [14] and virtual customer environments [15]. Salado and Nilchiani [16] were able to establish categorization models of requirements which are based on a well-known model of human needs. These research results form a solid basis for a connection between user satisfaction key figures from the BSC with the requirements driven development of digital processes with graph-based languages; thus allowing customer exploration (see Figure 7). It is important to note that the relationship between customer satisfaction and requirements can be complex and changing over time; therefore a continuous surveillance, reflection and optimization of the relationship is necessary. Considerable efforts are necessary for capturing the influence of requirements on user experience (UX); however these efforts can be a key factor for successful products, processes and innovations and a cornerstone of the system engineering processes [16], [17]. For such tasks, amongst others, certain elements of the theory of inventive problem solving (TRIZ) and digital product modelling techniques can be used [18].

A profound assessment of the future customer satisfaction frequently requires detail knowledge about the characteristics of future products. Conventional process do not allow the generation of this knowledge in the early phases of innovation selection. Graph-based languages allow a digital representation of many aspects of future products and far-reaching investigations of numerous parametrical and topological realizations. On the basis of this investigations a profound assessment of possible future characteristics and the consequential customer satisfaction is possible; thus allowing a key figure analysis (Figure 7). In the same manner, digital process also allow to investigate more possible technologies due to the efficient and fast evaluation processes; thus allowing a technology exploration (Figure 7). This exploration can also consider the internal processes of the company; a prominent example are more efficient production processes based on technological innovations which will allow to produce cheaper, faster or more flexible.

Key figures which capture the innovation processes are essentially process quality indicators. Digital methods reduce the amount of undocumented tasks and enable a holistic evaluation of the process quality; thus allowing a strategic objective evaluation (Figure 7).

A core element of the application of graph-based languages are digital processes in product and production development as well as adjacent processes. Such digital processes ease and enhance monitoring possibilities of these company processes and the generation of certain process quality indicators can be automated. This kind of quality indicators such as number of iterations, average time for certain tasks or investigated solution variants can lead to a continuous process improvement based on BSC and BIC.

5 CONCLUSIONS

This paper explained a framework to bridge the gap between the strategic considerations which the application of the balanced scorecard especially in the field of innovation management can foster and graph-based languages, which enable effective and efficient digital process in product and production development. Key aspects are user satisfaction, the exploration of a large number of innovation variants and improved processes. Further research needs to focus on a conscious implementation of the framework in company processes.

6 ACKNOWLEDGMENTS

The project „digital product life-cycle (ZaFH)“ is supported by a grant from the European Regional Development Fund and the Ministry of Science, Research and the Arts of Baden-Württemberg, Germany (information under: www.rwb-efre.baden-wuerttemberg.de).

7 REFERENCES

[1] BITKOM e.V. et al. (Hg.) (2015): Umsetzungsstrategie Industrie 4.0. Ergebnisbericht der Plattform Industrie 4.0. Online accessible under https://www.plattform-i40.de/I40/Redaktion/DE/Downloads/Publikation/umsetzungsstrategie-2015.html, verified at 26.02.2018.

[2] Königseder, C.: A Methodology for Supporting Design Grammar Development and Application in Computational Design Synthesis. ETH Zürich, Ph.D.-Thesis, 2015.

[3] Rigby, D., & Bilodeau, B. (2013). Management Tools & Trends 2013. New York: Bain & Company.

[4] Vinkemeier, R.: Gesamtkonzept zur langfristigen Steuerung von Innovationen – Balanced

Innovation Card im Zusammenspiel mit Roadmaps. In: Möhrle, M.; Isenmann, R. (Eds.): Technologie-Roadmapping. Zukunftsstrategien für Technologieunternehmen. Springer: 2017.

[5] Rudolph S. Aufbau und Einsatz von Entwurfssprachen für den Ingenieurentwurf. Forum Knowledge Based Engineering, CAT-PRO. 2003.

[6] Rudolph, Stephan (2006): Know-How Reuse in the Conceptual Design Phase of Complex Engineering Products. Or: 'Are you still constructing manually or do you already generate automatically'? In: Serge Tichkiewitch und Michael Tollenaere (Hg.): Proceedings of the 6th International Conference on Integrated Design and Manufacturing in Mechanical Engineering, IDMME 2006, Alpexpo-Grenoble, 17.-19. May 2006. Grenoble: AIP-PRIMECA, S. 23–39.

[7] Ramsaier, Manuel; Holder, Kevin; Zech, Andreas; Stetter, Ralf; Rudolph, Stephan; Till, Markus (2017a): Digital representation of product functions in multicopter design. In: Proceedings of the 21st International Con-ference on Engineering Design (ICED17) 2017 (Vol. 1: Resource-Sensitive Design), S. 369–378.

[8] Barthélemy, F., Knöll, H., Salfeld, A., Schulz-Sacharow, C., & Vögele, D. (2011). Balanced Scorecard – Erfolgreiche IT-Auswahl. Einführung und Anwendung. Wiesbaden: Springer Verlag.

[9] Beek, C.: Balanced Innovation Card: Instrument des strategischen Innovationsmanagements für mittelständische Automobilzulieferer. In: von Ahsen, A. (Ed.): Bewertung von Innovationen im Mittelstand. Springer: Heidelberg, 2010 (pp. 123 – 139).

[10] Nestle, V.: Innovation Balanced Scorecard. In: ControllingWiki; accessible under https://www.controlling-wiki.com/de/index.php/Innovation_Balanced_Scorecard; verified at 28.06.2018.

[11] Roth S (2007) Wie innovationsfähig ist die Automobilzulieferindustrie? REFA-Nachrichten (4):32–37.

[12] Zech, A.; Stetter, R.; Holder, K.; Rudolph, S.; Till, M.: Novel approach for a holistic and completely digital represented product development process by using graph-based design languages. Proceedings of the 12th CIRP Conference on Intelligent Computation in Manufacturing Engineering, 18-20 July 2018, Gulf of Naples, Italy.

[13] Holder, K.; Zech, A.; Ramsaier, M.; Stetter, R.; Niedermeier, H.-P.; Rudolph, S.; Till, M.: Model-Based Requirements Management in Gear Systems Design Based On Graph-Based Design Languages: In. Applied Sciences. 2017, 7, 1112; doi:10.3390.

[14] Sharif Ullah, A. M. M.; Tamaki, J.: Analysis of Kano-Model-Based Customer Needs for Product Development. In: Systems Engineering, Vol. 14, No. 2, pp. 154-172, 2011.

[15] Nambisan, S.: Designing Virtual Customer Environments for New Product Development: Toward a Theory. In Academy of Management Review, July 1, 2002, vol. 27, no.3 pp 392-413.

[16] Salado, A.; Nilchiani, R.: A Categorization Model of Requirements Based on Max-Neef's Model of Human Needs. In: Systems Engineering, vol. 17, 2014.

[17] Batra, M.; Bhatnagar, A.: A Comparative Study of Requirements Engineering Process Model. International Journal of Advanced Research in Computer Science; Vol. 8, No. 3, March 2017.

[18] Sharif Ullah, A.M.M.; Sato, M.; Watanabe, M.; Mamunur Rashid, M.: Integrating CAD, TRIZ, Customer Needs. International Journal of Automation Technology, Vol. 10, No. 2, 2016.

8 BIOGRAPHY

Ralf Stetter obtained his PhD degree in Mechanical Engineering from the Technical University of Munich. In 2004 he was appointed Professor for Design and Development at the University of Applied Sciences Ravensburg-Weingarten.

Jens Kiefer obtained his PhD degree in Mechanical Engineering from the Technical University of Saarbrücken. In 2013 he was appointed Professor for Quality Management at the University of Applied Sciences Ulm.

Marcin Witczak obtained his PhD degree in automatic control and robotics and his D.Sc. degree in electrical engineering. In 2015 he received a full professorial title and is Professor of automatic control and robotics at the University of Zielona Góra.

Markus Till obtained his PhD degree in Aeronautical Engineering from the University of Stuttgart. He is currently Professor for Simulation at the University of Applied Sciences Ravensburg-Weingarten.

Layout Design for Industrial Production Systems using Graph-Based Design Languages

C. Schlueter[1], T. Breckle[1], J. Kiefer[1], R. Stetter[2]
[1]University of Applied Sciences Ulm, Germany
[2]University of Applied Sciences Ravensburg-Weingarten, Germany

Abstract
Contemporary product design processes are becoming increasingly fast-moving and volatile, thus more agile planning concepts are needed. Modern planning tasks for layout design purposes are performed by trained planning personal and supported by software tools. It is of interest to examine how a planning approach based on graph-based design languages can be implemented in a concept planning environment. The main idea was to establish an autonomous design language, which is able to create production lines as well as shop floor manufacturing systems for larger production scales. This paper proposes a method that supports the design process for manufacturing and assembly layouts in an automated way. Structured data input (from product, process, resource and organizational levels) is used to generate material flow-based manufacturing layouts with the help of graph-based design languages. Within the mapped process, in- and outputs have been set in a way that keeps human interference to a minimum. While the layout optimization process is done manually, layout structuring and dimensioning are done by the prior mentioned design language. Subsequently, the design language evaluates each solution through the comparison of several cost factors (e.g. machine hour rate, fixed and variable costs). This way, an optimized result for layout purposes is achieved. The method is verified by application on a use case in which an assembly layout for a self-balancing two-wheel scouter is planned.

Keywords
Production systems, layout design, design languages

1 INTRODUCTION

Modern planning processes are becoming more volatile and fast living. Through competition, companies are bound to be up-to-date on current product developments as well as on recent production technologies to stay competitive. When disruptive product ideas are entering the market, it is best to have reliable, fast paced and general applicable processes that aid in production design. Those are in turn decreasing the time-to-market for products. There are various challenges and goals regarding the project phase of concept planning such as [1]:

- Reduction of engineering times (e.g. development of production design concepts)
- Reduction of lifecycle costs (e.g. definition of the most economic production setup)
- Highest engineering quality (e.g. quality and consistency of presented solutions)

In order to meet these planning challenges, the process of graph-based digital layout planning represents a promising approach. Therefore, the main research question was to investigate whether and how graph-based design languages can be applied to digital layout planning [1]. As shown in Figure 1, the paper's focus is on the project phase of layout planning, as a part of the concept phase, which describes the early phase of production engineering.

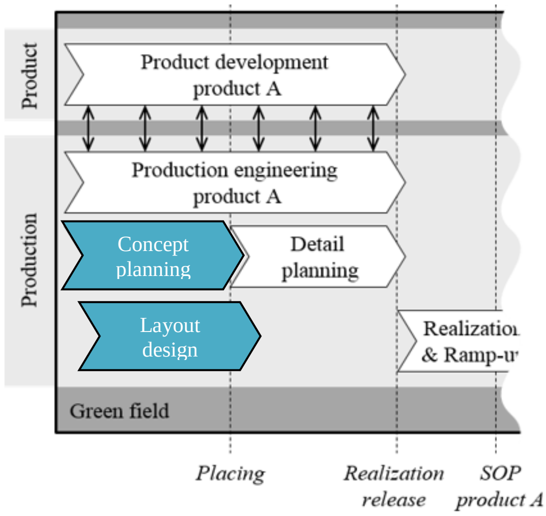


Figure 1 - Layout design process during concept planning phase (adapted from [1]).

The planning engineer's range of tasks are very versatile and can be subdivided into technical, economic and organizational activities. Examples of these tasks are the definition of process sequences (e.g. assembly sequences) and production layouts. In addition to that, the quantity structure of the production equipment is defined and concepts of production plants are clarified. The selection of the most appropriate production concept takes place in two main steps: the planning engineer develops several possible production concepts (e.g. several 3D plant layouts). Afterwards, all these concepts are compared to each other by means of specific criteria (e.g. costs, degree of flexibility). The "best" production concept is finally chosen to be transferred into specifications, which serve as basis for the detailed design of the production facilities [1].

The design language presented in this paper covers the layout planning in the context of production system planning. The layout design process is based on certain input information, which is necessary to generate a feasible layout solution. The target is to generate a functional layout design based on fixed process inputs (provided by earlier planning stages).

For the automatic processing of a large number of planning situations, a consistent approach is needed. Furthermore, it is required to trace process operators and their development throughout the different planning stages. Moreover a layout optimization solution is needed to generate a usable production layout.

2 METHODS AND TOOLS FOR LAYOUT OPTIMIZATION USED IN THIS PAPER

Within the concept of layout design there are numerous different methods and tools available. Optimization tools used within this paper are the transportation matrix and the circular procedure, which are explained in the following section.

2.1 Transportation matrices for layout optimization

Qualitative characteristics (e.g. presence and direction of material flow) and quantitative characteristics (e.g. material flow quantity per period and material flow intensity between areas) can be recognized in material flow matrices. The total number of all elements to be transported is displayed in the directed "From/To-Matrix". Within this matrix, the component elements are combined to load units. This specifies the number of transports that are required to supply a production station with production goods over a certain period of time. The transport matrix is therefore an indicator of the transport effort and takes into account the quantity of transported units [2].

	A	B	C	D	E
A		60	75	150	60
B			0	0	0
C				0	0
D					0
E					

Table 1 - "From/To-Matrix" (directed) for different work areas (schematic).

After all production areas and material flows have been classified, the ideal arrangement of the production departments can be determined through application of the circular procedure.

2.2 Circular procedure application for layout optimization

The circular method for layout design [2] serves as a tool for structuring material flows. The starting model is a circle on which the production areas are applied. Due to the recorded material flow intensities and directions between the individual departments, material flows can now be transferred to the diagram (Figure 2). The arrow thickness and color indicate the intensity (e.g. volume) of the material flow. The arrowhead characterizes the material flow direction. In the next step, the departments that have the largest material flow relationships to each other are arranged in close proximity on the circle. Accordingly, conclusions can be drawn about the ideal layout, which production areas must be close together.

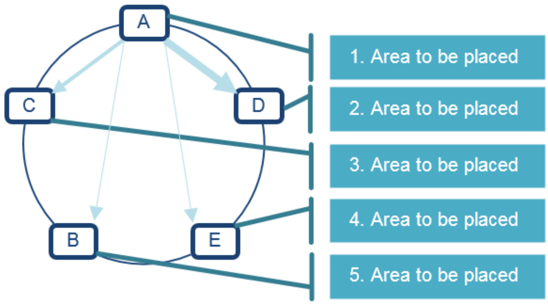


Figure 2 - Circular procedure application.

After the material flow has been completely determined in its direction and quantity, further planning levels (e.g. the degree of automation and the associated factors) can now be considered.

3 GRAPH-BASED DESIGN LANGUAGES

The planning approach presented in this paper is based on the idea of graph-based design languages. Through the integration of these design languages in a digital engineering framework, it is possible to reuse pre-existing models for planning or simulation purposes [1].

In the design process a multiplicity of different domains exist. This causes immense efforts and leads to an interface design and standard for exchanging data between these domains.

The concept of graph-based design languages [3] has evolved into a generic framework for the definition of computerized design processes. Its basic structure is depicted in Figure 3. Partial ontologies (e.g., abstract geometry, abstract digital factory) are expressed as vocabulary in a class diagram and are instantiated by means of rules within this process. In an activity diagram sequences of rules are predefined and processed and compiled in a so-called design compiler. As one example of these design complier, the Design Compiler 43 (DC43) is an eclipse-based software tool to compile design languages, which are formulated in UML (Unified Modelling Language). It is developed by the IILS mbH in cooperation with the University of Stuttgart. The central data model is built during the execution process, which is referenced as “Design Graph”. From this data model, domain-specific models can be developed in the respective software via different interfaces, e.g. geometry or simulation models. By this architecture an improved model consistency can be achieved, which is a central advantage of the described approach.

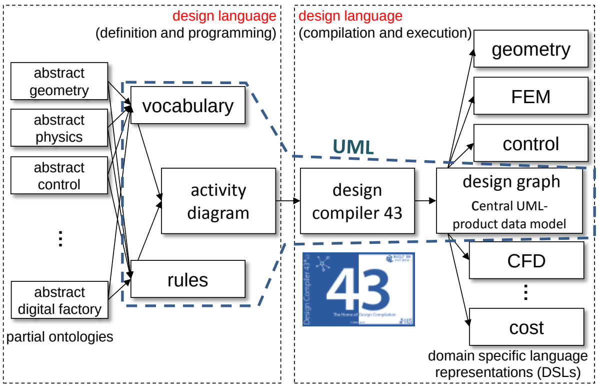


Figure 3 - Concept of graph-based design languages [4].

The process flow for application of a graph-based design language by means of UML diagrams is depicted in Figure 4. The process flow can be seen from top to bottom. The engineering objects (e.g., product components, processes, resources) are represented by UML classes (vocabulary of a design language) in this design language. One example of such an UML class could be a system such as a layout with its attributes, operations and (sub-) classes. By the UML class diagram all interconnections of the engineering objects (spatial, physical, logical, etc.) are represented as well as the hierarchical structure of the overall system. An example for those interconnections is that the layout consists of several parts, such as a sink. The model transformations are generated by rules (e.g. how to build up a production layout). These rules consist of a left side (LHS), which contains the structure before instantiation of the rule, and a right side (RHS), which describes the structure after the instantiation process (result under consideration of concrete requirements, e.g. the degree of automation). The activity diagram, which contains the sequence of the rules, is specified by the planning engineer. The UML class diagram can be seen as a blueprint of all possible layouts within the design language (e.g. all possible configuration of a layout suitable for a specific degree of automation). While executing this activity diagram, the individual vocabulary (UML classes) and their corresponding interconnections are instantiated. The design graph is created automatically taking into account project-specific data entries (e.g. production program). It thus represents a specific system from the blueprint system (e.g. a specific layout for a scooter that meets a specific production program) [1].

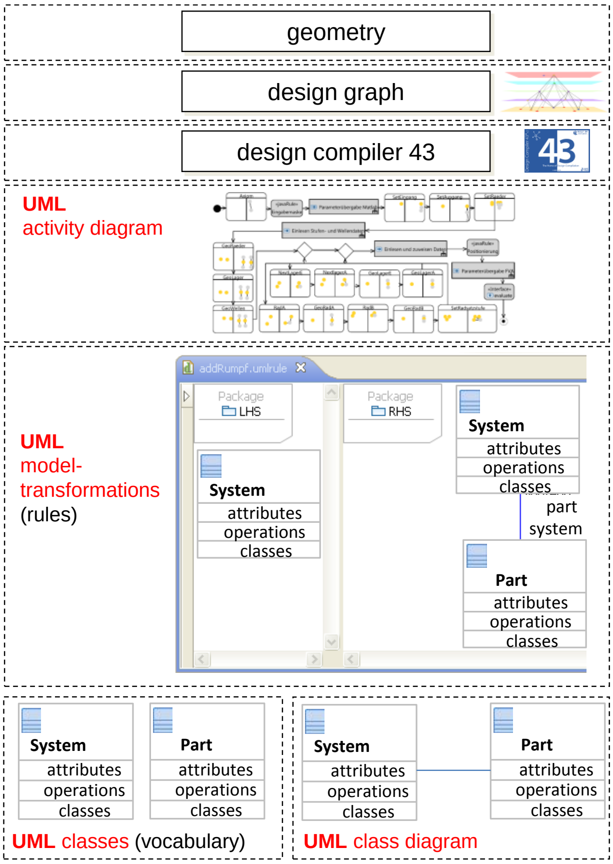


Figure 4 - Basic structure of graph-based design languages with UML framework.

The described digital design process using graph-based design languages, is a consistent and holistic modelling methodology, which achieves a design automation through the rule-based presentation of design and manufacturing knowledge. The digital design process using graph-based design languages. It has already been applied in a number of different use cases. Examples of implemented design languages can be found in [5], [6] or [7].

4 NEW PLANNING APPROACH OF GRAPH-BASED LAYOUT DESIGN

Based on this initial situation (e.g. definition of factors such as component geometry and assemblies) the method used, performs its calculations and in the end generates a possible layout solution. Numerous parameters have to be taken into account during the layout planning phase of a production environment. Influential factors can be identified on product-specific, process-specific, resource-specific and on organizational levels. The following influencing factors are used for further calculations such as the engineering drawing, parts list and operation sheet, production program and further specifications and requirements. The information provided by these data records provides a detailed outlook for capacity planning. The operation sheet is the main factor in the process stage. The production process can be read and entered here. On the resource side, considerations are made, based on a cost perspective. Figure 5 depicts the process steps that the design language performs to generate a layout solution.

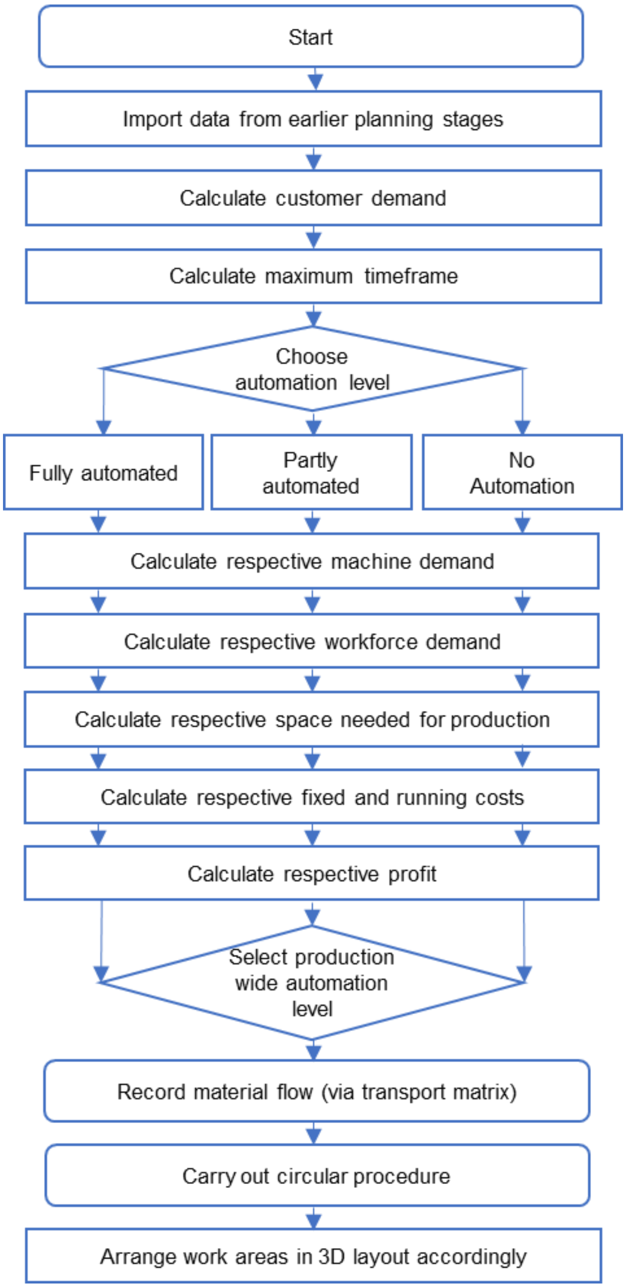


Figure 5 - Sequence of operations processed by the routine of the design language.

Firstly, influences are used for demand calculation purposes (e.g. space, workforce, workstations and their respective costs). This results in a product-specific quantity structure. On the product side, the parts list data (i.e. the necessary components and their respective quantities) are recorded. This has a particular effect on the subsequent material flow analysis. By calculating all production requirements with regard to parts/material flows, personnel and the required workstations, the quantity structure can be used to create the respective layouts for the different production approaches (e.g. different due to varying degree of process automation). The degree of automation has a direct impact on the design of the production system, its output, the investment costs and the operating costs. It is therefore an important decision criterion. A distinction can be made between manual, automated and semi-automated (hybrid) assembly-/manufacturing systems. The degree of automation must be selected for the entire production process environment.

Once the requirements and type of workstations have been determined based on the quantity structure defined in combination with the different automation levels, a final comparison of expenses and income can now be created. From this the profit margin based on the assumption that all customer orders will be fulfilled is calculated. Within the design language, the profit margin serves as the main criterion for selecting the degree of automation and the final layout solution.

After the selection of the automation level, the circular procedure makes it possible to identify the stations with the largest material flow. Likewise, all other areas are arranged in descending order of importance. Through the use of transportation matrices, it is possible to calculate the quantitative material flow via spreadsheet calculation. This enables the calculation of the transport effort of the accumulated material flow. In addition to the costs, material flow quantities are an important indicator for the quality of the layout as a whole.

5 USE CASE APPLICATION AND DISCUSSION

5.1 Assembly design for a self-balancing two-wheel scooter

As one specification of a production system the layout planning for an assembly system is portrayed. This use case deals with a self-balancing two wheel scooter [1], as it is depicted in Figure 6.

Based on the level of automation, respective production costs and demands, the design language selects the best layout variant and displays it to the user.

Figure 6 - Self-balancing two-wheel scooter.

To produce the depicted vehicle five different assembly stations that organized in different factory areas are necessary:

1) Steering Unit

2) Steering Rod
3) Electrical Control Unit
4) Power Electronics
5) Total Assembly

5.2 Assembly scenario

For simulation purposes, an assembly scenario with initial data was created. It is based on a random customer order situation. This includes the variable unit costs, the order quantity, and the revenue per unit (depending on the order quantity). From this the contribution margin per unit is calculated. The requirements calculation for workstations and machines is based on the table values stored in a reference data base. The costing values for different levels of assembly automation are then specified. To determine the operating costs, the totals of the fixed and variable costs are required. In this case, the (notional) fixed costs amount to a composition of premises, maintenance, insurance costs and depreciation. Variable production times and respective costs are based on needed energy, operating resources, media and personnel cost rates that can all be adjusted if desired.

Table 2 shows an example of how certain production parameters (e.g. costs and working times) were defined in the simulation environment.

Working days per Year [d]	200	= 1.500 hours of work per year per worker
Shifts per day	1	
Hours per shift [h]	7,5	

Fixed costs per year in [€]	684.000	= 869,000 €
Variable costs per year in [€]	185.000	

Table 2 - Exemplary values for working time per year as, well as variable and fixed costs.

5.3 Design language

The entities necessary for the layout design process that form the class diagram, are shown in Figure 7. It is possible to generate sources and sinks within the perimeter of the green field layout. Meanwhile the subordinated classes enable the use of a machine data base. Each "FactoryArea" – class has its own workstation data base, which serves as a reference catalogue and thus provides 3D models and table values (e.g. for costs). It is used to generate the necessary machinery/workstations for assembly and/or manufacturing purposes. The "GetData" class functions as a provider of imported data (e.g. from user input, spreadsheets or the production program), thus, the various calculation processes and the generation of the individual functional areas can be carried out on the basis of the selected degree of automation.

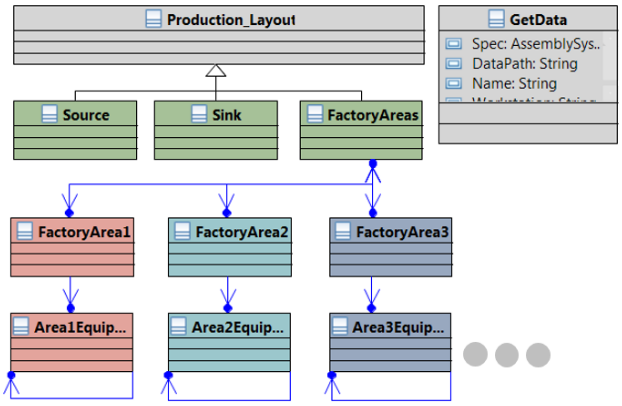


Figure 7 - Class diagram for process-based assembly planning (simplified).

Figure 8 depicts an extract of the activity diagram that contains the procedure (sequence of specified rules) that generates the scooter-specific production concept. Based on user entry (e.g. production program, level of automation) and the base layout, numerous factory areas with specific workstations are created and evaluated by the “Interface” – rule.

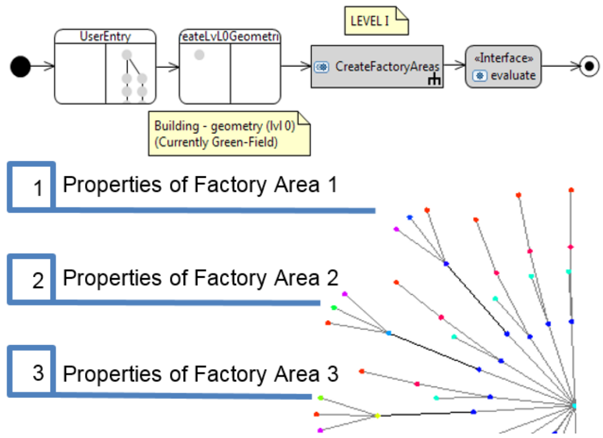


Figure 8 - Activity diagram and design graph for layout design of a self-balancing scooter.

The design graph is a depiction of how the different properties of each factory area (e.g. spatial position and coordinates, space) are combined to create a material flow based layout. It illustrates one solution variant of the scooter-specific assembly system. For this specific solution the branches of the design graph are split up into several generated work areas with their respective properties.

5.4 Resulting layout

This paper presents the method of graph-based digital layout planning during the concept phase as well as its software-technical implementation. The main research question was to investigate whether and how graph-based design languages can be applied for digital layout design planning. By means of a practical example, the application and advantages of the new planning approach are illustrated.

The solution presented by the design language is a block layout (Figure 9), which can also be presented in a 3D environment if desired. Three viable approaches (one for each automation level) have been considered and their respective production areas and the needed workforce have been calculated accordingly. The routine selected the layout with partly automated workstations, due to highest profit margin resulting on the preset circumstances.

Comparison	Full Automation	Hybrid Automation	Manual Assembly
Area [m²]	53,00	270,00	410,00
Required Workforce	15,00	54,00	132,00
Profit in [€]	527.938,80	596.353,86	422.140,66
Transportational effort in [km]	244.018,00	251.934,00	251.934,00

Table 3 - Solution comparison of the generated layouts.

Based on the unit loads that have to be handled, material flow is shown by the arrows, which connect the adjacent factory areas with each other (Figure 9).

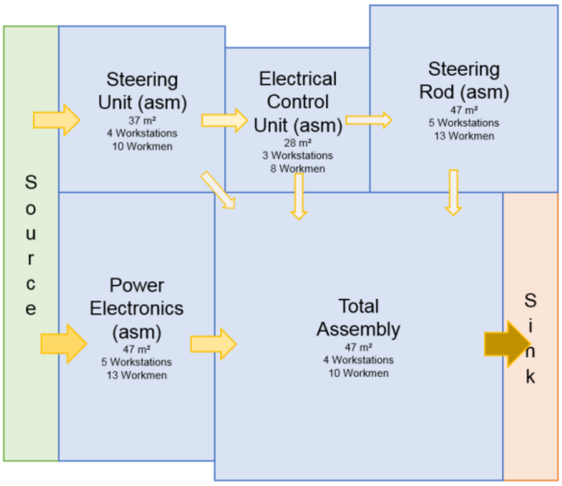


Figure 9 - Resulting block layout according to the rules set within the design language.

6 CONCLUSION AND OUTLOOK

Based from the substance of this investigation the conclusion can be drawn that much progress has been made in the field of graph-based layout design. However, many topics remain to be addressed in further research activities to improve the layout design process with graph-based design languages, such as:

1. Automation of layout algorithm
2. Consideration of uncertain project specific data inputs
3. Closed control loop between production planning and production

Moreover, it is of scientific interest how the product life-cycle can be further digitized using graph-based design languages and how to connect different product life-cycle intersections with each other.

7 ACKNOWLEDGEMENTS

The contents presented in this paper are part of the ZAFH project “Digital product life cycle”. It is supported by the ERDF and the MWK Baden-Württemberg, Germany. For more information please visit: https://dip.reutlingen-university.de/.

8 REFERENCES

[1] Kiefer, J., Breckle, T., Stetter, R., Manns, M., 2018, Digital assembly planning using graph-based design languages, Proceedings of the CIRP 72:802–807.

[2] Grundig, C.-G., 2018, Fabrikplanung: Planungssystematik, Methoden, Anwendungen, Hanser, p. 166.

[3] Rudolph, S., 2003, Aufbau und Einsatz von Entwurfssprachen für den Ingenieurentwurf, 3. Forum Knowledge Based Engineering.

[4] Rudolph, S., 2012, On the problem of multi-disciplinary system design and a solution approach using graph-based design languages, 1st ACCM Workshop on Mechatronic Design.

[5] Ramsaier, M., Holder, K., Zech, A., Stetter, R., Rudolph, S., Till, M., 2018, Digital representation of product functions in multicopter design, Proceedings of the 21st International Conference on Engineering Design (ICED17) pp. 369-378.

[6] Wünsch, F., 2016, Multidisziplinäre digitale Repräsentation des Produktlebenszyklus am Beispiel eines Segways, Master thesis.

[7] Breckle, T., Kiefer, J., Rudolph, S., Manns, M., 2017, Engineering of assembly systems using graph-based design languages, Proceedings of the 21th International Conference on Engineering Design (ICED17).

9 BIOGRAPHY

Christoph Schlueter is a master’s degree candidate from the Ulm University of Applied Sciences. His focus is the layout design process with graph-based design languages

Dipl. Ing. (FH), Theresa Breckle, MBA is a doctoral candidate at the Siegen University. Her specific focus is the concept planning of assembly systems.

Jens Kiefer obtained his PhD degree in Mechanical Engineering from the Technical University of Saarbrücken. In 2013 he was appointed Professor for Quality Management at the University of Applied Sciences Ulm.

Ralf Stetter obtained his PhD degree in Mechanical Engineering from the Technical University of Munich. In 2004 he was appointed Professor for Design and Development at the University of Applied Sciences Ravensburg-Weingarten.

Department of Industrial Engineering
Stellenbosch University
Private Bag X1
Matieland, 7600
Stellenbosch, South Africa
Tel: +27 21 808 4234 | Fax: +27 21 808 4245
E-mail: devonh@sun.ac.za; kvl@sun.ac.za

ISBN: 978-0-7972-1779-9